APPLIED ARTIFICIAL INTELLIGENCE FOR DELAY RISK PREDICTION OF BIM-BASED CONSTRUCTION PROJECTS

CHRISTIAN NNAEMEKA EGWIM

Submitted in partial fulfilment of the requirements of the University of Hertfordshire for the degree of *Doctor of Philosophy*

April 2024

ABSTRACT

Construction projects are complex endeavours requiring intricate coordination among stakeholders, resources, and timelines. However, delays are a pervasive issue, causing financial losses, reputational damage, and compromised outcomes. This research addresses this problem by integrating systematic reviews, expert insights, and cutting-edge artificial intelligence (AI) techniques. The study begins with a systematic review adhering to Systematic Reviews and Meta-Analyses (PRISMA) guidelines to identify common drivers of project delays. Analyzing scholarly articles from various regions and project types, the research develops a conceptual framework categorizing delay risk drivers into nine distinct groups. To validate these findings, an expert survey involving industry professionals is conducted, ensuring the research reflects real-world insights. This results in an empirically validated framework for assessing delay risks. Concurrently, the study reviews AI applications in construction, identifying supervised learning and deep learning as the most impactful technologies for predictive modelling. Using the validated delay risk drivers as features, the research develops hyperparameter-optimized AI predictive models through a process of feature engineering, model development, optimization, and evaluation. Among the evaluated models, the Fully Connected Neural Network (FCNN) demonstrates superior performance. To enhance model interpretability, the study employs SHapley Additive exPlanations (SHAP), providing transparent explanations for model predictions. This transparency fosters trust among stakeholders and enables targeted interventions to address critical delay risk drivers. The development and validation of the FCNN model represent a significant advancement in anticipating and mitigating project delays in construction. The integration of SHAP enhances the model's transparency and interpretability, empowering professionals with a powerful tool for proactive delay risk assessment and mitigation. This research makes substantial contributions to academic knowledge and industry practice by providing a robust predictive model and enhancing model interpretability. Acknowledging limitations such as data scope, industry dynamics, potential biases, and inherent machine learning constraints, the study suggests future research opportunities. These include exploring diverse datasets, incorporating new AI techniques, improving interpretability, integrating decision support systems, and leveraging synergies with emerging technologies.

DEDICATION

In reverence to the king eternal, immortal, and invincible, the sole wise and true God, I humbly dedicate this doctoral research. I express profound gratitude for His guidance, enabling me to accomplish this pinnacle of academic pursuit. I extend heartfelt appreciation to my mother for her unwavering support in all aspects. May the Almighty Jesus grant you longevity to enjoy the fruits of her labour of love in me, in Jesus' name, amen. Words fall short in conveying my sentiments. I also acknowledge my father for his consistent moral encouragement. May the divine presence of Jesus Christ continue to encompass, bless, and preserve him. Furthermore, I express deep gratitude to my cherished wife, Glory Egwim, and my children, Chimamanda, Chikaima, and Chizitere Egwim, who have been pillars of unwavering support and love. Your continuous encouragement is beyond measure, and I cherish and will always cherish you all. May the Lord and Saviour Jesus Christ continue to guide and protect us, granting us longevity, prosperity, and the ultimate blessing of meeting Him face to face on the final day, in Jesus' name, amen.

ACKNOWLEDGEMENT

Foremost, I express gratitude for the grace of Almighty God, without which this work would not have reached fruition. I offer unending thanks, praise, and adoration to Him. I extend heartfelt appreciation to my supervisor, Professor Hafiz Alaka, whom I consider a divine gift. His guidance and direction have been invaluable throughout this project and beyond, serving as a constant source of inspiration in my post-graduate academic journey. I am profoundly grateful for his unwavering support, and if given the opportunity, I would choose him as my supervisor repeatedly. I also wish to acknowledge the significant contributions of my second supervisor, Professor Eren Demir, whose input was immeasurable. Oh, what a man! Furthermore, I am indebted to my family, including my immediate nuclear family, parents, and siblings, for their unwavering moral support. I express appreciation to my colleagues, with whom I formed a formidable team of early career academics under the mentorship of Professor Hafiz Alaka. Your contributions are deeply valued, and you will all be dearly missed. Special recognition goes to Dr Raphael Oseghale and Dr Francesca Gagliardi for their exceptional support. This acknowledgment would be incomplete without mentioning the invaluable assistance provided by the staff at the Doctoral College office, particularly Dr Susan Grey, Lynne Mee, and Charlotte McHarry. Their guidance, timely reminders, and the researcher development programmes (RDP) were instrumental in navigating the programme smoothly. Ultimately, I attribute all praise and glory to my dearest Lord and personal saviour of my soul, Jesus Christ, for guiding me through every step of this journey.

LIST OF RESEARCH PUBLICATIONS

Journal Papers

1. **Egwim, C.N.** et al. (2023) 'Artificial Intelligence in the Construction Industry: A Systematic Review of the Entire Construction Value Chain Lifecycle', Energies 2024, Vol. 17, Page 182, 17(1), p. 182. doi:10.3390/EN17010182.
2. **Egwim, C.N**. et al. (2022) 'Systematic review of critical drivers for delay risk prediction: towards a conceptual framework for BIM-based construction projects', Frontiers in Engineering and Built Environment, 3(1), pp. 16–31. doi:10.1108/febe-05-2022-0017.
3. **Egwim, C.N.** et al. (2021) 'Applied artificial intelligence for predicting construction projects delay', Machine Learning with Applications, 6, p. 100166. doi:10.1016/j.mlwa.2021.100166.
4. **Egwim, C.N.** et al. (2021) 'Extraction of underlying factors causing construction projects delay in Nigeria', Journal of Engineering, Design and Technology, 21(5), pp. 1323–1342. doi:10.1108/JEDT-04-2021-0211.

Conference Papers

1. **Egwim, C.N.**, Alaka, H. and Balogun, H. (2021) 'Effects of Building Information Modelling on Construction Projects Delay: A Systematic Review', in Environmental Design and Management International Conference, Bristol, United Kingdom.
2. **Egwim, C.N**. and Alaka, H. (2021) 'A Comparative Study on Machine Learning Algorithms for Predicting Construction Projects Delay', in Environmental Design and Management International Conference, Bristol, United Kingdom.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

1.0 INTRODUCTION[1]

1.1 Background of Study

The construction industry contributed £117 billion to United Kingdom (UK) economy In 2018, representing 6% of the total economic output and provided 2.4 million jobs, representing 6.6% of all jobs in the UK (Rshodes, 2019). The case is similar in other countries e.g., construction industry represents 3% of the total economic output of Nigeria (Oladinrin, Ogunsemi and Aje, 2012), 4.3% of the total economic output of Germany (European Comission, 2017), 4.1% and 6.8% of the total economic output of the United States of America (USA) and China respectively (Wang, 2018, 2019). On the global scale the construction industry is similarly worth 13% of the global gross domestic product (GDP) with a promising 85% to $15.5 billion globally by the year 2030 (Filipe Barbosa, Jonathan Woetzel, 2017). Furthermore, Woetzel et al. (2017) estimates global infrastructure spending at $3.4 trillion annually from 2013 to 2030, which is roughly 4% of total GDP, solely delivered as large-scale projects. The industry is thus considered a major backbone of any country's economy and a major contributor to the global economy. However, despite its importance the construction industry has continued to underperform. According to Egan (2018) the construction industry is underperforming as evident in its low profitability, capital investment, research and development caused generally by delay of construction projects which results in great dissatisfaction from the industry's clients on its overall performance. Construction project delay has been defined as a project where key dates or milestones have been missed or where the contractual date of completion must be forfeited, (Van *et al*., 2015). Delay has also been described as an occurrence which may result in the loss of income for the client or owner Haseeb, Bibi & Rabbani, (2011). A delay may also be characterized as a somewhat incremental increase in both overheads and labour costs for the contractor and is deeply detested by all parties involved in a construction project.

[1] This chapter is primarily derived from the following journal articles:

Egwim, C.N., Alaka, H., Toriola-Coker, L.O., Balogun, H. and Sunmola, F. (2021) 'Applied artificial intelligence for predicting construction projects delay', Machine Learning with Applications, 6, p. 100166. doi:10.1016/j.mlwa.2021.100166.

Egwim, C.N., Alaka, H., Toriola-Coker, L.O., Balogun, H., Ajayi, S., et al. (2021) 'Extraction of underlying factors causing construction projects delay in Nigeria', Journal of Engineering, Design and Technology, ahead-of-p(ahead-of-print). doi:10.1108/jedt-04-2021-0211.

Randolph et al. (1987) used bid cost to classify project sizes as small (< $50,000), medium ($50,000–$250,000), and large (> $250,000) while Yang, O'Connor & Wang (2006) used total installed cost to characterize project size and categorized it into small (< $5million), medium ($5–50million) and large (> $50million). This research adopts the European Commission (2015) categorization of project size: small (< €10 million), medium (< €50 million) and large (> €50 million), since the research is based in the UK which was a part of the European Union(EU) at the time. Major large-sized projects that have been delayed in the UK for instance as one of the most advanced countries of the world have caused huge financial problems, e.g. the Crossrail project also known as the Elizabeth Line, one of Europe's sizeable construction project is yet to be officially commissioned and is at present £17.6 billion, (18.9%) over the original £14.8 billion budget (Haylen, 2019). The building of the Scottish parliament named Holyrood experienced a 20 months delay which contributed to a cost increase from £195 million in September 2000 to £431 million in February 2004 (White & Sidhu, 2005). Furthermore, the London Jubilee Line Extension project, opened virtually two years late in 1999 and contributed to 80% higher cost than the original budget, and the Channel Tunnel was commissioned a year late in 1994 and had a final cost of £9.5 billion, double the original budget due to delay risk contributions. Construction project delay is not only restricted to large-sized construction project. For example, these medium-sized projects: Almond Bank Flood Protection Scheme project had a 4 months delay that contributed to £17.6 million completion cost rather than the £15 million initial budget and the North Bridge Refurbishment project is currently in progress haven missed its initial Autumn 2019 completion plan and has an increase cost from initial £17.25 million to £22 million (Scape, 2020). Additionally, these small-sized projects: Carnival Pool Multi Storey Car Park of Wokingham Borough Council project experienced a 6 months delay from its 12 months schedule with an increased cost valued at £2.5 million and St Crispin's Community Centre project finished 3 months later than scheduled and resulted in £2.2 million in cost (Scape, 2020).

With reference to Shebob et al.(2012), 60% of the UK's construction projects runs beyond their contract period. In truth, delay on construction projects, and the associated costs is by no means limited to the UK, construction delay is a global phenomenon. Some research publications, for instance, Harris (2013), Flyvbjerg (2014) and Rhodes (2019) indicated that 9 out of 10 global mega projects encounter delay which usually result to excess cost overruns. They argued that delays of seemly 50% in real terms of construction projects are habitual. These papers also stated that delay have been perpetually high over the last 70 years. Investigation by several researchers have shown that delay of construction projects has adverse effect on the reputation of the construction industry's contribution to the global economy. With reference to Abdul-Rahman et al., (2011), the effects of construction delay can be evaluated with respect to its national footprints which with prejudice sway the industry's subsidy to the economy; at an industry level, where delays impact profitability and productivity negatively; and at a project level where delays foster industry client's dissatisfaction on its overall performance, cessation of contracts by the owner, and unprofitability for contractor(s). Furthermore it has been argued (Kumar R, 2016) that delay often lead to project cost overruns, insolvency of organization, loss of opportunity of future projects, dispute among project stakeholders (e.g. clients, contractors, architect, engineers, sub-contractors, suppliers and consultants) and legal actions.

Major delay factors have been identified by several researchers as evident in vast body of international literature (from amongst the oldest articles to the most recent), e.g. bad weather and jurisdictional/contractual disputes in both United States of America (USA) and UK by Baldwin et al., (1971) and Sullivan et al., (1986) respectively; variation orders in both Nigeria and United Arab Emirates (UAE) by Odeyinka, (1997) and Motaleb, (2010) respectively; planning and scheduling deficiencies in Australia, delay in payment certificates in Ghana and poor site management in Malaysia by Shah, (2016); ground problems and inefficient structural connections for prefabricated components in both the UK and India by (Agyekum-Mensah et al., 2017) and Ji et al., (2018) respectively and finally shortage of adequate equipment and poor communication among contracting parties in China by Chen et al., (2019). Over the decades, several research methods and recommendations towards mitigating delay of construction projects have been identified. For instance, Sullivan & Harris, (1986) suggested more teamwork especially at the early stages of project planning. It is the viewpoint of Mansfield, Ugwu & Doran, (1994); Frimpong, Oluwoye & Crawford, (2003) that contractors should buy construction materials at the early stage of work and be more familiar with effective and efficient material procurement systems/software. Also, according to Assaf, Al-Khalil & Al-Hazmi, (1995); Enshassi, AlNajjar & Kumaraswamy, (2009); Owolabi et al., (2014); Alaghbari & Sultan, (2018) clients should adhere to timely payment of progress fee and consider funding levels at the planning stage of project. Furthermore, the survey by Gondia et al., (2020); Yaseen et al., (2020) recommended the use of predictive models to mitigate delay risks and time claim in construction projects. Despite all these delay factors and recommendations towards mitigating delay in construction, delay still strives in the industry. However, though Building Information Modelling (BIM) has brought about a lot of improvement on construction projects as claimed in many studies, (e.g. Ballesty et al., (2007); Azhar, (2011); Azhar, Khalfan & Maqsood, (2012); Jones & Dewberry, (2012); Latiffi et al., (2013); Mohd & Latiffi, (2013); Gibbs et al., (2013); Johansen, (2015)) among others, there has not been a lot of studies on the effect of BIM on construction delay when compared to construction project without BIM (see section 3 for more details).

1.2 BIM and Construction Delay.

Construction delay is the most important factor to consider in the general execution of any construction project as it expands cost overruns (Haq *et al.*, 2017). In the construction industry the term delay is comprehensively used giving rise to vast body of international literature definition of the term (Gibbs et al., 2013). They defined delay as any unexpected extension to the entire scheduled span or the occurrence that lengthen the period of a task without generally influencing the project term (cited in Bramble & Callahan, 2004). It is the viewpoints of Assaf & Al-Hejji (2006) that delay can be defined as the time increase beyond the agreed project delivery planned schedule by stakeholders or beyond a legal contract completion date. Furthermore, delay mean different things to different stakeholders (owner, contractor, consultant etc.) and is oftentimes referred to as time or schedule overruns by various scholars (Abdul-Rahman et al., 2009; Akhund et al., 2017; Al-Hazim et al., 2017; Elawi et al., 2016; Gardezi et al., 2014; Głuszak & Lešniak, 2015; Orangi et al., 2011). For the owner, delay connote loss of revenue or investments at the end of agreed time while to the contractor, a delay can imply an increase in overhead cost(Assaf & Al-Hejji, 2006). The construction process comprises a series of interconnected phases - from building material manufacturing to design, planning, and construction, as well as facilities management each influencing the others directly or indirectly (Egwim, Alaka, Toriola-Coker, Balogun, Ajayi, *et al.*, 2021). Also, several studies have revealed that project delays are more prevalent during the execution stage of construction. For instance, research by Aibinu and Jagboro (2022) found that the construction stage accounts for a significant portion of delays in construction projects. Similarly, the findings of Akintoye et al. (2000) corroborate this, emphasizing that construction-related factors contribute prominently to project delays. Moreover, in contemporary scholarly discourse, a substantial proportion of literature focuses on the phenomenon of construction project delays without BIM implementation referred to this study of non-BIM-based construction project over those with BIM implementation (BIM-Based construction projects) Surprisingly, researchers (Saka and Chan, 2021; Silverio and Suresh, 2021; Tai, Zhang and Li, 2021) have revealed that the delay factors that impact non-BIM-based and BIM-based construction projects are not necessarily the same. This is due to a variety of reasons (see section 3). The hypothesis of BIM was established in 1970 by Professor Charles Eastman at the Georgia Tech School of Architecture as building description systems (BDS) (Eastman et al., 2008). Undeterred by its long-time existence, interest in BIM only took off few years ago. This present-day construction industry is predisposed by the wariness about BIM. Varying concerns around what exactly BIM is, whether BIM is only meant for large projects with complex geometries, how to change from the traditional design process to BIM etc. In 2000, BIM was defined as structured model of data that represents building elements with its usage spanning beyond the pre-construction phase to the post-construction phase in Architecture, Engineering, and Construction industry (Ameziane, 2000). However, facility managers see BIM as a tool used to improve building's performance and manage operations more efficiently throughout a building's life (Abbasnejad & Moud, 2013). The practical adoption of BIM by the Architecture, Engineering, and Construction industry for construction project started around the mid-2000s (Latiffi et al., 2013). BIM was first implemented by the United States of America (USA) with example BIM-based construction projects seen in Sutter Medical Centre, Castro Valley California USA (Davis, 2007). Presently, BIM-based construction projects have been implemented in several countries such as "Sydney Opera House, in Australia; "One Island East Office Tower", in Hong Kong; "Crussel Bridge", in Helsinki, Finland; "National Cancer Institute (NCI)", in Putrajaya, Malaysia; "Barking Riverside Extension and Rail Station", in London, UK etc (Eastman et al., 2008; Latiffi et al., 2013). The rate of adoption of BIM differs from country to another. For instance, according to a report by Bernstein (2010) 50% adoption rate amidst contractors was reached in North America compared to barely 24% of the counterparts in Western Europe meanwhile the Western Europe has more percentage of BIM user rate: 34% against North American BIM user rate: 18%. With reference to House et al. (2007) the major benefit of BIM is its accurate geometrical representation of building parts in an integrated data. Some research (Johansen, 2015; Jones, Young Jr. & Bernstein, 2008; Mohd & Latiffi n.d.; Siddiq, 2018) indicated that BIM is generally used during pre-construction, construction and post construction stages to produce better project design; aid in decision making process; improve collaboration and communication among stakeholders; centralize data administration in a common data environment; reduce changes during construction; reduce conflict during construction; minimize risks in execution period; visualize design solution in 3D; reduce project delay; improve overall project quality and achieve better cost control/predictability. Although the benefits of BIM with regards to construction projects have been enlisted, its application to the real world is seemingly limited to just 3D visualization and clash detection, hence there remains a large gap between its proposed application area and its current implementation (Jin, Tang & Fang, 2015). Consequently, in investigating into whether BIM can assist with construction delay claims, Mohd & Latiffi (2013) reported that the employment of BIM tools by construction players aids the mitigation of delay in construction projects through project scheduling. Examples of such BIM tool includes Autodesk Revit, Autodesk NavisWorks, AutoCAD Civil 3D, Digital Project, Bentley, Vectorworks, Tekla, and Vico (Jones, Young Jr. & Bernstein, 2008). Project scheduling is known to generate an overall project duration by using logic and mathematical calculations to sequence all the activities needed to complete the works (CIOB, 2011). According to Gibbs et al., (2013) a critical path on the schedule is used to represent the major activities necessary to finish a project with the shortest time possible and an overlap on these activities will signify an extension in project time hence resulting to project delay. So, since BIM offers a way of coordinating all project activities throughout its lifecycle stored in a common data environment and linked to a 3D model plus time (4D), these BIM tools are able to offer support for project control against delay(s) (Gibbs et al., 2013; Hartmann et al., 2012). They however, concluded that, understanding cause and effect of change in construction doesn't come easy and suggested the use of multiple dimensions to connect the information generated in the delay analysis to an n-dimension representation of the project. Young et al., (2008) take a similar view, insisting that even though the design capabilities of BIM are long-established, the potential of BIM to offer scheduling functions—also referred to as 4D is still coming to the fore likely due to huge outlays already made in project management software by firms.

1.3 Artificial Intelligence / Machine Learning in Construction.

Artificial Intelligence (AI) represents a powerful contemporary analysis method that has been widely adopted across other industries, but construction industry is slow to adopt (Blanco et al., 2018; Marks, 2017). The adoption of AI and Machine Learning (ML) algorithms in construction is relatively evolving, especially when compared to other industries like healthcare: guiding in the choice of treatment; education: virtual lectures; and transportation: autonomous vehicles, as it currently uses lots of methods that were used in the centuries past (Blanco et al., 2018; Marks, 2017). Furthermore, AI technologies exhibit varying degrees of effectiveness across different stages of construction (Egwim *et al.*, 2024). Certain AI technologies excel primarily during the design stage, while others demonstrate their utility during construction or post-construction stages. Nonetheless, there are instances where AI technologies prove beneficial across multiple stages of the construction process (see section 4). As a common place the industry produces massive amount of data daily on every project, for example data produced from images captured from smart devices, IoT sensors, BIM etc presents a window of opportunity for the industry and its customers to examine and gain profits from insights generated from past construction data through the aid of AI and ML. AI is defined as a collection of state-of-the-art technologies that permit machines or any computer programme to sense, comprehend, act and learn(Goyal, 2019). ML on the other hand is a branch of AI that allows computers to learn by a direct route from examples, data and experience replacing the traditional approaches to programming that relied on hardcoded step by step rules(Royal Society, 2017). This is done by giving the system huge amount of data to learn from as a task leaving it to decide how best to achieve the task in form of a desired output. AI/ML offer transformative potential for addressing delays in construction projects. The unique strengths of AI/ML include several key rationales. Firstly, AI/ML provide data-driven insights. Construction projects generate massive amounts of data daily, ranging from images captured by smart devices and IoT sensors to BIM data. AI/ML can process and analyse these extensive datasets to uncover patterns and insights that are not readily apparent through traditional methods. This data-driven approach helps identify the root causes of delays and provides actionable insights to mitigate them. Secondly, AI/ML exhibit predictive capabilities. ML algorithms excel at learning from historical data to predict future outcomes. By analyzing data from past construction projects, ML models can anticipate potential delays before they occur. This predictive capability allows project managers to take proactive measures, such as adjusting schedules or reallocating resources, to avoid delays. Moreover, AI contributes to optimization and efficiency in construction management. AI can optimize various aspects of construction management, including scheduling, resource allocation, and workflow processes. For instance, AI algorithms can suggest the most efficient sequence of tasks and the optimal distribution of labour and materials, thereby reducing the likelihood of delays caused by inefficiencies. Furthermore, AI provides real-time monitoring and adjustments. AI systems can offer real-time monitoring of construction projects, continuously comparing actual progress with the planned schedule. When deviations are detected, AI can suggest corrective actions to bring the project back on track. This real-time adjustment capability is crucial for minimizing the impact of unforeseen issues and maintaining project timelines. Additionally, AI/ML tools enhance decision-making. These tools support decision-making by providing project managers with data-driven recommendations. They can analyse multiple factors simultaneously, offering insights that lead to more informed and effective decisions. Enhanced decision-making helps in pre-emptively addressing issues that could lead to delays, thus improving overall project management. Several ML algorithms such as Neural Networks, Linear Regression, Logistic Regression, Nearest-Neighbour Mapping, Decision Trees, K-Means Clustering, Random Forests, Support Vector Machines, Principal Component Analysis, Singular Value Decomposition, etc. exist for ML model implementation. Which ML algorithm to use depends on lot of factors e.g. ease of use, accuracy, training time etc. Few researchers have attempted the use of AI and ML algorithms in some aspect of construction. Poh, Ubeynarayana & Goh (2018) used 5 popular ML algorithms to predict accident occurrence and severity of construction sites in Singapore, Zou & Ergan (2019) Leveraged on 3 ML techniques to predict the influence of construction projects on urban quality of life; Arditi & Pulket (2005) and Mahfouz & Kandil (2012) used only 1 and 3 ML algorithms respectively to forecast end results of construction litigation all in the USA. Only a hand full of literature have attempted the adoption of AI or ML to mitigate construction delay. For example, Gondia et al. (2020) used 2 ML models towards expediting precise project delay risk assessments and forecast, Asadi, Alsubaey & Makatsoris (2015) used 2 ML approach (with accuracy value of 79.41% and 73.52% for decision tree and Naive Bayes model respectively) to predict delays in construction logistics in Qatar. Also, Yaseen et al. (2020) developed a hybrid artificial intelligence model (a combination of Random Forest and Genetic Algorithm) and achieved an accuracy value of 91.67% for delay problem prediction. Furthermore, Yaseen et al. (2020) used Artificial Neural Network (ANN) to forecast final budget and duration of highway construction projects in Thailand. Evidently there is no singular literature currently that have attempted to use feature selection and cross validation in the ML algorithms to predict delays of a BIM-based construction project in the world hence the first motivation for this research. Additionally, no research has ever attempted the use of ML algorithms to predict delay risk of BIM-based in comparison to non-BIM based construction projects giving rise to the final motivation of this research. First, this research will highlight the key factors such that when BIM is involved in a project will help to reduce construction delay. Secondly, this research will develop an AI technology model to predict potential delay of BIM-based construction project by using several ML algorithms.

1.4 Problem Statement

In consequent, the problem statements are as follow:

1. There is no clear research evidence of distinguishing factors that affect BIM based construction projects as compared to non-BIM-based ones.
2. There is no clear evidence of a research backed BIM-based framework that can potentially help to mitigate construction delay in the industry.
3. No clear evidence of a research that have taken the advantage of the contemporary analysis method which best explains the factors that can be affecting a phenomenon like delay based on its predictive capabilities.

1.5 Research Questions

A set of research questions that will convey the research aim and objectives are given below:

1. What are the most common factors affecting construction project delays in construction projects, as identified through a systematic review, and how do these factors apply to BIM-based construction projects in terms of their applicability according to expert judgements?
2. What are the most applicable AI technologies across the entire construction value chain lifecycle, as identified through a systematic review of AI in the construction industry?
3. How can the applicable factors identified in the first objective be used as independent features (variables) with the most applicable AI technologies as identified in the second objective to develop hyperparameter optimized AI predictive models?

1.6 Research Aim and Objectives

This project aims to develop AI models for delay risk prediction of BIM-based construction projects. To fulfil the stated aim, the following objectives were formed as follows:

1. Conduct a systematic review toward gathering the most common factors affecting construction project delays and use it to conduct survey of experts to establish the most applicable factors affecting construction project delays in BIM-based construction projects.
2. Conduct a systematic review of AI in the construction industry and use it to establish the most appropriate AI technologies during construction.
3. Utilize the applicable factors in the first objective as independent features (variables) with the most applicable AI technologies as identified in the second objective to develop hyperparameter optimized AI predictive models.

1.7 Research Methodology

This study adopts a positivist research philosophy and deductive approach, aligning with the quantitative research design. The positivist paradigm is chosen as it facilitates hypothesis testing and the establishment of verifiable knowledge through empirical observation and scientific methods. The deductive approach allows the researcher to formulate hypotheses based on existing theories and test them through data collection and analysis. Also, it employs a quantitative research design, which involves the collection and analysis of numerical data. This design is suitable for achieving the research objectives, which include identifying correlations between construction project delays and potential covariates, as well as developing predictive models using machine learning techniques. Quantitative data is essential for machine learning algorithms, which require numerical inputs for prediction and classification tasks. The research strategy adopted is a survey, implemented through an online questionnaire. This approach allows for the collection of data from a large sample of construction industry experts, enabling the identification of the most applicable factors contributing to construction project delays in large-scale, BIM-based infrastructure projects. The survey consists of sections addressing various aspects of project delays, including the frequency of occurrence of delay factors, their relative importance, and the level of detail associated with each factor. The study utilizes a probability sampling technique, specifically stratified sampling, to ensure an appropriate sample size for population subgroups of interest. This method involves dividing the population into homogeneous, mutually exclusive strata and then selecting independent samples from each stratum. The stratified sampling approach ensures adequate representation of different subgroups within the construction industry, such as contractors, quantity surveyors, architects, and engineers. The unit of analysis for this study is construction project delay, falling under the category of "social artifacts." The research focuses on understanding and analyzing delays in construction projects, with construction project delay being the central phenomenon under investigation. The unit of observation, which is the entity on which initial measurements are performed, is also construction project delay, as the measurements and observations are specifically designed to gauge the extent and factors contributing to project delays.

Data collection is conducted through primary sources, utilizing the online survey questionnaire distributed to construction industry experts. Primary data collection is chosen as it allows for the generation of data tailored to the specific research objectives and ensures the relevance and accuracy of the information obtained. Data analysis techniques employed in this study include reliability analysis, exploratory data analysis, data cleaning, outlier detection, feature selection and engineering, hyperparameter tuning, principal component analysis, ensemble methods, and hypothesis testing. These techniques are applied to achieve the research objectives, which involve identifying the most applicable drivers of construction project delays, establishing the most applicable AI technologies for the construction value chain lifecycle, and developing optimized AI models for predicting and mitigating project delays. The systematic review process follows the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines. PRISMA provides a standardized and comprehensive framework for conducting and reporting systematic reviews, ensuring transparency, reproducibility, and scientific rigor. The PRISMA checklist and flow diagram guide the various stages of the systematic review process, from literature search and study selection to data extraction and synthesis.

1.8 Research Contributions

1.8.1 Contribution of Study to Academic Knowledge

This research has made significant contributions to academic knowledge in several ways. Firstly, the systematic review conducted on identifying the most common drivers affecting construction project delays represents a comprehensive and up-to-date synthesis of existing literature on this topic. By rigorously adhering to the PRISMA guidelines, the review process ensured a thorough and unbiased exploration of delay factors across diverse construction project types and geographical regions. The resulting conceptual framework, encompassing nine distinct categories of delay risk drivers, serves as a valuable foundation for future research in this domain, providing a robust starting point for further investigation and validation. Moreover, the integration of industry expertise through the expert survey adds a crucial practical dimension to the study, bridging the gap between academic research and real-world industry practices. This approach not only validates the findings from the systematic review but also ensures that the resulting framework is grounded in the realities and nuances of the construction industry. By incorporating the perspectives of subject matter experts, the study enhances the relevance and applicability of its findings, fostering a stronger connection between academic research and practical implementation. The systematic review on the application of AI technologies in the construction industry represents another significant contribution to academic knowledge. By rigorously analyzing vast body of academic literature, the review provides a comprehensive and up-to-date understanding of the current state of AI adoption in the construction value chain lifecycle. The identification of seven major AI technology types, with supervised learning and deep learning emerging as the most prominent, offers valuable insights for researchers and practitioners alike. Furthermore, the review's exploration of the applicability of these AI technologies across the three major stages of the construction project lifecycle (planning/design, construction/execution, and supply/facility management) provides a roadmap for future research and technological advancements in the industry. The development of hyperparameter-optimized AI predictive models for assessing delay risks in both BIM-based and non-BIM-based construction projects represents a ground-breaking contribution to academic knowledge such that has never been done before in one study. By leveraging the identified critical delay factors as input features and employing the most suitable AI technologies (supervised learning and deep learning), this study has successfully created a robust and accurate predictive model. The Fully Connected Neural Network (FCNN) model, which emerged as the optimal choice, exhibits exceptional performance metrics. This model not only demonstrates the potential of AI in addressing complex challenges in the construction industry but also serves as a blueprint for future research in developing AI-driven solutions for risk assessment and project management. The integration of SHapley Additive exPlanations (SHAP) into the predictive model further enhances the study's contribution to academic knowledge. By providing a transparent and interpretable explanation of the model's predictions, SHAP addresses the long-standing issue of "black box" models in AI, fostering trust and understanding among researchers and construction stakeholders Consequently, this study has made significant strides in advancing academic knowledge in the fields of construction project management, delay risk assessment, and AI applications in the construction industry.

1.8.2 Contribution of Study to Practice

This research makes substantial contributions to the practice of construction project management, particularly in the realm of delay risk assessment and mitigation. By identifying and validating the critical drivers of construction project delays through a systematic review and expert survey, the study provides construction professionals with a comprehensive and industry-relevant framework for understanding and addressing these challenges. The developed hyperparameter-optimized AI predictive model, specifically the FCNN, represents a significant practical contribution to the construction industry. This model offers construction stakeholders a powerful tool for assessing the risk of potential delays in both BIM-based and non-BIM-based construction projects. The model's exceptional performance metrics, including accuracy scores for BIM-based and non-BIM-based projects, respectively, demonstrate its reliability and effectiveness in predicting potential delays. The integration of SHAP into the predictive model further enhances its practical value by providing transparency and interpretability to the model's predictions. Construction professionals can gain valuable insights into the specific factors contributing to potential delays, such as late payment by the owner, inaccurate resource planning, space limitations at the site, reworks due to construction errors, and unskilled labour. This information empowers construction stakeholders to develop targeted strategies and interventions to mitigate these critical delay risk drivers proactively. Also, the predictive model's ability to assess delay risks early in the project lifecycle offers construction professionals a significant advantage in project planning and resource allocation. By identifying potential bottlenecks and high-risk areas before they manifest, construction teams can implement preventive measures, optimize resource allocation, and develop contingency plans to minimize the impact of delays on project timelines and budgets. Furthermore, the study's contribution extends beyond individual construction projects. The validated framework of delay risk drivers and the predictive model's capabilities can be leveraged by construction firms and organizations to enhance their overall project management practices. By integrating these tools into their project management processes, companies can improve their risk management strategies, enhance decision-making processes, and foster a culture of proactive delay mitigation. The practical implications of this research also extend to the broader construction industry ecosystem. Government agencies, regulatory bodies, and policymakers can utilize the findings to develop guidelines, best practices, and industry standards for delay risk assessment and mitigation. By incorporating these insights into industry regulations and certification processes, the construction industry can collectively promote transparency, accountability, and a proactive approach to addressing project delays. Moreover, the study's findings can inform the development of training programs and educational curricula for construction professionals. By incorporating the identified delay risk drivers and the predictive model's capabilities into professional development programs, construction companies can better equip their workforce with the knowledge and tools necessary to tackle project delays effectively.

1.8.3 Research Scope and Limitations

Regarding size, this study focused on large construction projects – projects with a gross budget of €50 million or more, as defined by the European Commission (2015). This threshold ensures that the research captures the complexities and challenges associated with managing large-scale construction endeavours, which often involve substantial financial investments and intricate coordination among various stakeholders. With respect to the type of projects, the study will concentrate on large-scale infrastructure projects. These projects encompass the development and maintenance of critical physical assets, such as roads, telecommunication networks, railways, tunnels, bridges, and other essential infrastructure vital for a nation's economic growth and development (Cullingworth, 2014). The inclusion of infrastructure projects is particularly relevant because most of these large-scale undertakings increasingly incorporate BIM methodologies (Bradley et al., 2016). By focusing on large-scale infrastructure projects that employ BIM, the study aims to address the unique challenges and factors contributing to construction delays within the construction industry. The integration of BIM processes adds an additional layer of complexity, requiring careful consideration of technological adoption, data management, and interdisciplinary collaboration. Investigating these projects thus provides valuable insights into the effective implementation of BIM and its potential impact on mitigating delays in large-scale construction endeavours critical to a nation's infrastructure development. While this study has made significant contributions to both academic knowledge and industry practice, it is essential to acknowledge and address its inherent limitations. One of the primary limitations lies in the scope and representativeness of the data used for model development and validation. The expert survey data, although valuable, may not fully capture the diverse range of construction projects, geographic regions, and industry practices. Consequently, the generalizability of the findings and the predictive model's performance might be limited when applied to contexts outside the scope of the data. Another potential limitation is the dynamic nature of the construction industry itself. The factors influencing project delays are subject to continuous evolution, as new technologies, regulations, and industry practices emerge. This study's findings, while comprehensive at the time of research, may require periodic updates and refinements to remain relevant and accurately reflect the changing landscape of the construction industry. Furthermore, the study's reliance on expert judgements and systematic reviews, although rigorous and systematic, may inadvertently introduce biases or overlook emerging trends or factors that have not yet been widely documented or recognized within the industry. This limitation highlights the need for continuous monitoring and incorporation of new insights as they become available. It is important to note that the FCNN predictive model, despite its impressive performance, is not infallible. Like any machine learning model, it is subject to the limitations of the training data and the assumptions inherent in its architecture and optimization process. Unforeseen edge cases or outliers in real-world scenarios may challenge the model's predictive capabilities, necessitating ongoing refinement and adaptation.

1.9 Thesis Structure.

This section provides the overview of the research study in a tabular form. This comprises of the nine chapters and a brief description of the chapters is provided below:

Table 1.1: Thesis Structure

Chapter Number	Chapter Title	Summary
Chapter 1	Introduction	This chapter provides the overview of this research.
Chapter 2	Theories affecting construction delay.	This chapter reviews the management theories that affect construction delay and how they apply to factors to be extracted for systematic review in the next two chapters.
Chapter 3	Systematic review of construction delay and the effect of BIM on construction delay.	Chapter 3 provides a systematic review of construction delay and the effect of BIM on it.
Chapter 4	Systematic review of artificial intelligence/ machine learning in construction.	This chapter details a systematic review of artificial intelligence/ machine learning in the construction industry.
Chapter 5	Research methodology	This chapter describes relevant aspect of research models in details: Reasoning behind the selection of research approach, technique for data collection etc.

Chapter Number	Chapter Title	Summary
Chapter 6	Research analysis and result	This chapter confers the exhaustive analysis of data (respondent profiles, descriptive, reliability analysis and factor analysis. etc).
Chapter 7	AI model development	This chapter summarises the model development processes including (data preparation, feature engineering, selection and framework validation).
Chapter 8	Discussion	This chapter delivers an evaluation and discussion of the research findings.
Chapter 9	Conclusion and recommendation	This chapter summarises the research findings, contributions, limitations, and future directions.

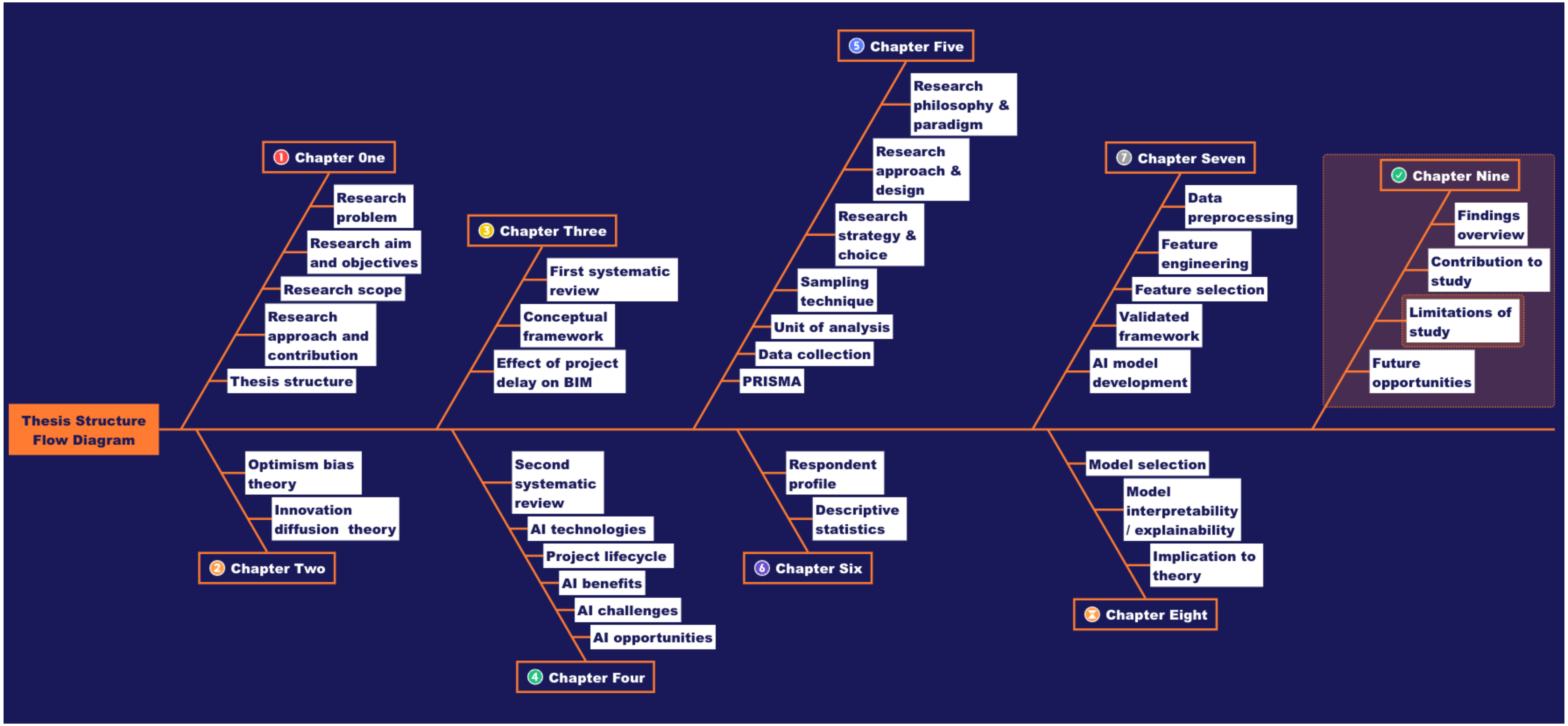

Figure 1.1: Thesis Structure Diagram

1.10 Chapter Summary

To summarise, this chapter introduced my research with a clear explanation to the background of the research problem, the background of the research problem was given in the context of the fact that there is a massive delay on the construction industry which results to efficiency, profitability and sustainability issues which is a great problem to the economy of several countries. BIM has been introduced since 1970, however no clear evidence of predicting models for delay predictions in BIM-based construction projects as established in the BIM and construction delay subsection. This led to this research's aims and objectives, research questions, scope and limitation, research contribution and approach. Key research contributions are improving the quality of decision and risks taken by several stockholders in construction through developing an AI technology that can predict potential delay of their present or future projects. A systematic review of existing literatures and questionnaire of survey were conducted as a research approach for this study. The thesis structure is given. The next chapter will review literatures that are anticipated to be useful and appropriate for undertaking this research study.

CHAPTER 2

2.0 THEORETICAL FOUNDATIONS OF CONSTRUCTION PROJECT DELAYS

2.1 Chapter Overview

This chapter explores the theoretical underpinnings that help explain construction project delays. Two main theories are examined - the Optimism Bias Theory and the Innovation Diffusion Theory. Each offers a unique perspective on the underlying factors that can derail project performance and hinder technology diffusion within the construction sector. The Optimism Bias Theory sheds light on the cognitive biases and risk assessment shortcomings that frequently lead to overly optimistic forecasting of project costs, schedules, and benefits during planning phases. On the other hand, the Innovation Diffusion Theory examines the characteristics of innovations themselves, as well as the social dynamics that influence how rapidly or slowly, they disseminate through a system over time.

2.2 Optimism Bias Theory

According to Macdonald (2002), optimism bias is the propensity to underestimate a project's costs and duration while overestimating its benefits. He defined optimism bias as "a measure of the extent to which actual project costs, and duration (time from business case to benefit delivery and time from contract award to benefit delivery) exceed those estimated". This occurs because humans frequently assume everything will work out as planned, so we make optimistic estimations while understating how long things will take or how much they will actually cost. Optimism bias can be mathematically expressed as:

$$\text{Optimism}_{\text{bias}} = 100 \times \frac{(Actual - Estimated)}{Estimated}\ \%$$

The theory of optimism bias provides a useful account of how it is considered optimistic to estimate project costs or durations when they do not fully account for the possibility of cost and delay or shortfalls. In construction projects, optimism bias can be distinguished into six basic categories: works duration, project duration, capital expenditure, operating expenditure, unitary payments and benefits shortfall (Macdonald, 2002; White, Cunningham and Titchener, 2011). Works duration optimism bias refers to the duration of the execution stage of a project, which includes the design, mobilization, and construction phases. One well-known study that is often cited in research on optimism bias is that of Macdonald (2002), who found that the estimated works duration from the outline business case and contract award are contrasted with the actual works duration such that its measured optimism bias merely indicates how much the works duration has grown and provides no insight into whether the project was completed on schedule. Drawing on the work of a wide range of philosophers, Sharot (2011) advances the notion that the length of time it takes for a project to complete each stage, from planning to execution, is known as the project duration optimism bias. This view is supported by White et al., (2011) who writes that in addition to delays in the actual construction of the project, delays in the project's procurement i.e., before construction really starts are also the cause of project duration overruns. A broadly similar point has also subsequently been made by Flyvbjerg (2004), who argued that as a comparably significant length of time may have elapsed between these phases, the project duration optimism bias is significantly reliant on the life-cycle stage at which the business case information is collected (i.e., strategic outline case, outline business case, or complete business case). In a comprehensive literature review of optimism bias, Caffieri et al. (2018) identified that the capital expenditure optimism bias offers a measurement of the proportional rise in capital expenditure between what was projected in the outline business case as well as at contract award and what was invested in capital. According to Flyvbjerg (2008), variations in the construction cost index and tender price index (pre contact award) frequently contribute to capital expenditure optimism bias (post contract award). Along the same lines, Love et al. (2017) subsequently argued that the project costs—both the expected and actual outlays is usually indexed to a common year in order to eliminate the impact of indexes. According to Macdonald (2002), operating expenditure optimism bias can be defined as the amount by which the business case's projected expenditures are outperformed by the actual operating expenses. Both a capital and an operating component make up unitary payments optimism bias. Raisbeck et al. (2010) made an important point by saying It is vital to identify significant project risk areas that have an influence on the levels of optimism bias for capital and operating expenditures in order to minimise unitary payments optimism bias. As with capital and operational expenditure optimism bias for traditionally purchased projects, controlling these project risk areas would lower the optimism bias for unitary payments. A benefits shortfall optimism bias arises when we compare the benefits delivered with those estimated at the beginning of a business case and at the moment the contract is awarded (Macdonald, 2002). Benefits are sometimes ill-defined; thus, it is necessary to utilise best judgement to identify shortfalls. When a shortfall is identified during project research, the shortfall should be quantified according to the interviewee's viewpoint, the project's decreased capability, or the project's success in achieving its goals. According to a vast body literature (Macdonald, 2002; Flyvbjerg, 2004; Sharot, 2011; Caffieri *et al.*, 2018), the inability to properly identify and manage project risks is the major cause of optimism bias in construction projects. Five typical project risk groups were identified by the Macdonald (2002) research, each of which had a number of project risk areas that were known to lead to cost and schedule overruns as well as benefit deficiencies. In order to comprehend optimism bias better, Flyvbjerg (2004) divided its root causes into four main groups, as indicated in Table 2.1.

Table 2.1: Causes of Optimism Bias

Causes of Optimism Bias	Example
Technical causes	Imperfect information (inadequate data, inaccurate forecast and new or unproven technology)
	Change in project scope or inadequate business case.
	Poor management often reflected in the poor documentation.
Psychological causes	Tendency for humans and organisations to favour optimism appraisal optimism
	Projects promoters are overly optimistic about projects outcomes.
Economic causes	Construction companies and consultants have interest's in advancing projects
Political-Institutional causes	Interests, power, and institutions
	Actors may deliberately lie in order to see their projects/ interest realised

Failure to take into account and actively control optimism bias's causes will lead to cost and delays as well as benefit deficiencies that go beyond what might be accomplished if the reasons were identified and actively handled. However, it is feasible to lessen the optimism bias and increase confidence in project estimates by considering risks while defining the nature and scope of a project and then creating techniques for the efficient management of risks. To lessen the chance of cost and schedule overruns, as well as benefit shortfalls when the project is implemented, it is therefore important to evaluate how well project risks have been recognised and will be handled throughout project assessments. Furthermore, it should be feasible to successfully limit project risks and lessen any potential optimism bias by applying current industry best practises. The phenomenon of optimism bias extends beyond mere miscalculations in project planning and is deeply rooted in psychological and social dynamics. Understanding these underlying factors is crucial for comprehensively addressing optimism bias in construction projects. Optimism bias is not merely a result of poor estimation techniques but is deeply ingrained in human psychology. As Kahneman and Tversky (1979) illustrated through their Prospect Theory, individuals have a tendency to focus on positive outcomes while undervaluing potential risks and negative consequences. This cognitive bias, known as the optimism bias, leads to overly optimistic project timelines and cost estimates. Sharot (2011) further elaborated on this by explaining that the brain's reward system is more activated by positive predictions than negative ones, making optimistic forecasts more appealing and likely to be adopted. Within the organizational context, social dynamics and the pressure to present favourable outcomes can exacerbate optimism bias. Managers and project proponents often face significant pressure to demonstrate the viability and profitability of their projects. This pressure can lead to deliberate downplaying of potential risks and overstatement of benefits, a phenomenon also discussed by Flyvbjerg (2004) under the term "strategic misrepresentation." This strategic behaviour is not always a result of malicious intent but can stem from a genuine belief in the project's success, bolstered by a collective optimism within the organization. The impact of optimism bias on project performance can be profound. Studies by Love et al. (2017) have shown that projects affected by high levels of optimism bias often experience significant cost overruns and delays. This not only affects the financial viability of the project but also its reputation and stakeholder trust. Furthermore, projects with unrealistic timelines and budgets are more likely to encounter stress and resource shortages, leading to compromised quality and increased likelihood of failure. To mitigate optimism bias, several strategies have been proposed and successfully implemented in various projects. One effective approach is the implementation of reference class forecasting, as suggested by Flyvbjerg (2008). This technique involves comparing the current project with a database of similar past projects to provide a more realistic estimate of costs and timelines. Additionally, the use of independent peer reviews can help identify and correct overly optimistic assumptions early in the project planning process. Also, raising awareness about the cognitive and social roots of optimism bias through targeted training programs can also be beneficial. By educating project teams about the common pitfalls of optimism bias and encouraging a culture of critical evaluation, organizations can foster a more realistic approach to project planning and execution. Moreover, advancements in AI and ML offer new tools for combating optimism bias. These technologies can analyse large datasets from previous projects to provide predictive insights and identify patterns that may not be evident through traditional analysis. As discussed earlier, AI/ML can play a crucial role in real-time monitoring and adjustments, further enhancing the accuracy of project estimates and timelines. The adoption and implementation of BIM in construction projects present unique challenges and opportunities. Optimism bias theory is particularly relevant in BIM, as it can significantly influence how BIM is perceived, planned, and executed within projects. One common manifestation of optimism bias in BIM adoption is the overestimation of its potential benefits. While BIM offers numerous advantages, such as improved design accuracy, enhanced collaboration, and better project visualization, stakeholders may exaggerate these benefits without fully understanding the complexities involved. Optimistic bias might predict seamless integration and immediate productivity gains, neglecting the learning curve and the need for substantial upfront investment in training and technology. Optimism bias also leads to underestimation of the costs and timeframes required for BIM implementation. Initial project plans might overlook the extensive resources needed to transition from traditional methods to BIM. These include the costs of software acquisition, training personnel, and restructuring workflows to accommodate BIM processes. Additionally, the time required to fully integrate BIM into existing systems is often underestimated, leading to project delays and cost overruns. BIM relies on the accurate and efficient management of vast amounts of data. Optimism bias may cause project teams to underestimate the challenges associated with data management and interoperability between different software platforms. Inadequate planning for these technical hurdles can result in data silos, errors, and inefficiencies that negate some of the anticipated benefits of BIM. Furthermore, optimism bias can significantly impact stakeholder expectations regarding BIM adoption. Project promoters may present an overly optimistic view of BIM’s capabilities, leading to unrealistic expectations among clients and other stakeholders. When the reality of implementation falls short of these expectations, it can lead to dissatisfaction, loss of trust, and potential conflicts.

2.3 Innovation Diffusion Theory

Diffusion of Innovation (DOI), according to Rodgers (2003), is the process " through which an innovation is disseminated over time among members of the social system through specific channels". Innovation, communication routes, time, and the social structure make up the bulk of the DOI. Since innovation is defined as "an idea, practise, or undertaking that is regarded as novel by an individual or other units of adoption," it need not actually be new, only new to the adopter (Rodgers, 2003). The mechanism through which people exchange information with one another, including interpersonal and media channels, is known as a communication channel. The time pertains to how long it takes to adopt or reject, how long adoption takes compared to being early or late, and how long it takes for a certain number of adopters. While the social structure is made up of interconnected parts that cooperate to accomplish a goal(s). The theory views innovation as occurring at many stages and being impacted by factors including the invention's qualities, those of its adopters, and the local social structure. According to Rodgers (2003), Relative advantage, compatibility, complexity, trialability, and observability are the qualities associated with innovation. Compatibility is defined as "the degree to which an innovation is regarded to be compatible with the existing values, prior experiences, and demands of potential adopters." Relative advantage is "the degree to which an innovation is judged to be superior than the notion it succeeds." The degree to which an innovation is seen as being challenging to understand and use is referred to as its complexity, observability is "the degree to which the consequences of an invention are visible to others," whereas trialability is "the degree to which an innovation may be tested with on a limited basis”. Information is spread through communication channels to educate interested people about the innovation, who then may be convinced (persuaded) to decide whether to accept or reject the innovation. The adoption decision is followed by the implementation, which entails employing the invention for a while before confirmation Rodgers (2003). As a result, different adopters adapt differently based on their innovativeness, or how early or late they adopt in comparison to other adopters in the system as illustrated in figure 2.1 below.

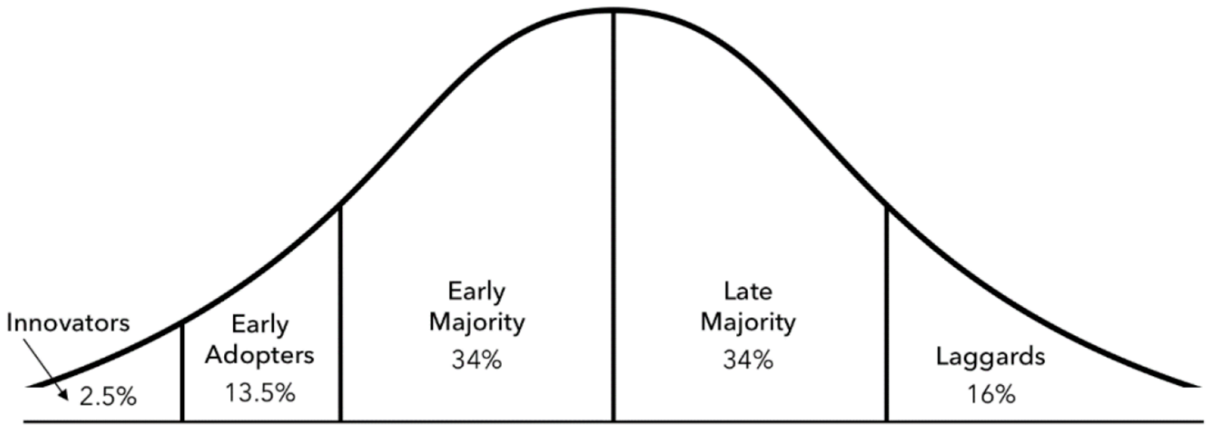

Figure 2.1: Diffusion Model (Rodgers, 2003)

The adopters are divided into laggards, innovators, early majority, early majority, and late majority. It is believed that innovators have access to major mass media and are frequently pioneers and risk-takers who are not afraid to try forth novel concepts. The innovators set the example for the early adopters, who then quickly embrace the invention. In order to conceptualise BIM adoption inside Australian SMEs, Hosseini et al. (2016) applied the innovation diffusion model and stated that the theory is relevant for the BIM adoption process. The theory has drawn criticism, nevertheless, for putting too much emphasis on the technology environment and not enough on the social structure of the system (Ifinedo, 2011; Ishak and Newton, 2016). The oversimplification of the theory and the portrayal of adoption as a binary function that omits the usage were questioned by Bayer & Melone (1989). The research also noted that the classification of adopters based on innovativeness, the absence of adopter representation of innovation discontinuity, the impacts of government mandates on innovation, and the lack of a specification for the interplay between diverse social systems were all unjustified. The Popperian technique was used by Lyytinen & Damsgaard (2001), who compiled the theory's six main conjectures for debunking. The six conjectures, according to Lyytinen & Damsgaard (2001), are as follows: (1) Innovations have distinct characteristics that are readily identifiable by the interested parties; (2) Innovation moves from an independent innovator to other adopters in a discrete manner; and (3) The decision to adopt or reject by an adopter is atomic, isolated, and influenced by pull and push forces. (4) The adopters make decisions based on information made available through communication channels, technical features, and a logical calculus. (5) The adoption rate, which is influenced by the push and pull forces, determines how quickly diffusion occurs. (6) The diffusion process is devoid of feedback; thus, the diffusion history is irrelevant. The study used the example of electronic data interchange to disprove the idea that innovation does not alter over time or within different contexts. It was discovered that complicated innovation would not disseminate in the theory's suggested consecutive phases. Additionally, the choice to adopt or reject may not be based on the knowledge given and may go outside the social structure. Despite the objections, it is commonly known that innovation diffusion theory is valuable for innovative studies; nevertheless, the theory has to be changed to account for the flaws found in sophisticated technologies like BIM. According to Rogers (2003), the innovation-decision process consists of five key steps:

- **Knowledge** - When an individual is first exposed to an innovation and gains some understanding of how it functions.
- **Persuasion** - When the individual forms either a favourable or unfavourable attitude towards the innovation.
- **Decision** - The individual weighs the advantages/disadvantages and decides whether to adopt or reject the innovation.
- **Implementation** - The innovation is put into practice.
- **Confirmation** - The individual finalizes their decision to continue using the innovation.

This multi-stage process highlights that adoption is not a binary event, but rather unfolds over a period influenced by various factors. As Rogers highlights, "The innovation-decision period is the length of time required to pass through the innovation-decision process" (2003, p.172). Rogers identifies several variables that can facilitate or impede adoption, including characteristics of the decision-making unit, characteristics of the innovation itself as mentioned earlier (relative advantage, compatibility, etc.), communication channels, nature of the social system, and role of change agents or opinion leaders. For example, higher socioeconomic status, more years of formal education, and greater cosmopolitanism and interpersonal networking have been found to correlate with higher innovativeness in adopting new ideas (Che Ibrahim et al., 2010). The theory also covers diffusion networks and critical mass - the notion that adoption of interactive innovations like BIM is self-sustaining only after a critical mass of users have adopted (Rogers, 2003). As more interconnected individuals adopt the innovation, network effects increase its value for all, incentivizing further adoption. Furthermore, Rogers makes a key distinction between centralized, hierarchical diffusion systems where adoption choices are relatively optional, versus decentralized diffusion systems with collective agency where choices are more autonomous (Pries & Jansen, 2009). Construction projects often represent temporary, decentralized diffusion systems where adoption patterns are more unpredictable. While impactful, the Innovation Diffusion Theory has seen notable critiques and advancements over time: Attewell (1992) argued the theory oversimplifies by not accounting for knowledge barriers where potential adopters lack understanding of complex innovations like IT systems. He proposed a "Knowledge Barrier Model" incorporating phases of knowledge acquisition before adoption can occur. Another critique is that the theory tends to have a pro-innovation bias that oversimplifies the rational decision-making process (Rogers, 2003). It downplays how adoption is often motivated by institutional pressures rather than rational weighing of costs/benefits (Shi et al., 2008). Swanson and Ramiller (1997, 2004) expanded diffusion concepts to theorize that broader level "organizing visions" around innovations like BIM emerge to legitimize and motivate adoption across an industry. These shared visions shape discourse around an innovation's value proposition. Finally, some argue that the linear Innovation Diffusion Model is too rigid, and that adoption unfolds through more dynamic, recursive cycles of initiation, adoption, implementation, and institutionalization over time (Elenkov & Mohr, 2018; Zolkafli et al., 2012). The DOI, as articulated by Rogers (2003), offers a comprehensive framework to understand the adoption and dissemination of innovations BIM within the construction industry. According to Rogers, the diffusion process involves the spread of an innovation over time among members of a social system through specific channels. The theory's relevance to BIM adoption is particularly significant, given the complex and multidisciplinary nature of construction projects, which require seamless coordination and communication across various stakeholders. BIM, as an innovative technology, fits well within the DOI framework, which outlines five key attributes influencing the rate of adoption: relative advantage, compatibility, complexity, trialability, and observability. Relative advantage refers to the perceived benefits of BIM over traditional methods, such as enhanced visualization, improved collaboration, and increased efficiency. These benefits can be compelling drivers for adoption, especially in a competitive industry seeking to improve project outcomes. Compatibility addresses how well BIM aligns with the existing values, past experiences, and needs of potential adopters. This is crucial in the construction sector, where entrenched practices and resistance to change can pose significant barriers. Complexity is another critical factor, as BIM involves sophisticated software and requires a certain level of technical expertise. This complexity can hinder adoption, particularly in small and medium-sized enterprises (SMEs) that may lack the necessary resources or skills. The attribute of trialability, or the ability to experiment with BIM on a limited basis, can mitigate some of these concerns by allowing firms to explore its benefits without full-scale commitment. Finally, observability pertains to the visibility of BIM's results, which can encourage adoption by showcasing its practical benefits through successful case studies and projects. The process of BIM adoption also aligns with Rogers' stages of the innovation decision process: knowledge, persuasion, decision, implementation, and confirmation. Initially, potential adopters become aware of BIM and gain an understanding of its functions (knowledge). This is followed by the formation of a positive or negative attitude towards BIM (persuasion), leading to a decision to adopt or reject it. Upon adoption, BIM is implemented and tested, with the final confirmation stage involving the continued use or abandonment of the technology based on its performance and benefits.

2.4 Chapter Summary

This chapter presented an overview of two key theories that have been applied to the conceptualization of research on construction project delays: the Optimism Bias Theory and the Innovation Diffusion Theory. The Optimism Bias Theory, as proposed by Macdonald (2002), addresses the tendency to underestimate project costs and durations while overestimating benefits. This bias is quantified as the percentage difference between actual and estimated values. Optimism bias manifests in various forms, including works duration, project duration, capital expenditure, operating expenditure, unitary payments, and benefits shortfall biases. Alternatively, he Innovation Diffusion Theory, proposed by Rodgers (2003), offers a framework for understanding the process by which innovations are adopted within a social system over time. The theory encompasses four key elements: innovation, communication channels, time, and social structure. Innovations are characterized by relative advantage, compatibility, complexity, trialability, and observability, which influence their adoption rate. While the theory remains valuable for innovation studies, it requires adaptation to account for the intricacies of complex technologies.

CHAPTER 3

3.0 SYSTEMATIC REVIEW OF CRITICAL DRIVERS FOR DELAY RISK PREDICTION: TOWARDS A CONCEPTUAL FRAMEWORK FOR CONSTRUCTION PROJECTS[2]

3.1 Chapter Overview

This chapter presents a systematic review of the critical drivers for delay risk prediction towards developing a conceptual framework for BIM-based construction projects. The study recognizes a significant gap in the existing body of knowledge – the lack of a cohesive conceptual framework for identifying and prioritizing the most critical delay risk drivers specific to BIM-based construction projects. Thus, it aims to identify key delay risk drivers in BIM-based construction projects that have a significant impact on the performance of delay risk predictive modelling techniques. It employs the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guideline to conduct a comprehensive systematic review and synthesize the findings. The chapter discusses the background and motivations behind this research, highlighting the persistent challenges of construction project delays and the potential of BIM in mitigating these issues. It emphasizes the lack of a cohesive conceptual framework for selecting the most critical delay risk drivers specific to BIM-based projects, which has hindered the development of highly effective predictive models tailored to BIM-based construction projects.

3.2 Background of Study

Construction projects are one-of-a-kind and are seen to be inherently risky owing to the involvement of many parties with competing interests. The risks associated with construction projects are many and have the potential to result in negative outcomes. One of such risks is delay risk as construction industry rarely completes projects in time due to its complex nature - vagaries of project type, scope, location and size (Egwim, Alaka, Toriola-Coker, Balogun, Ajayi, *et al.*, 2021). Risk management methods that

[2] This chapter is primarily derived from the following journal articles:

Egwim, C.N. et al. (2023) 'Systematic review of critical drivers for delay risk prediction: towards a conceptual framework for BIM-based construction projects', Frontiers in Engineering and Built Environment, 3(1), pp. 16–31. doi:10.1108/febe-05-2022-0017.

are systematic and realistic are required to handle and control delay risks so that project success may be assured. As a result of the detrimental impact of such construction projects delay on the economy and society in general, researchers have recently used delay risk predictive modelling as a major risk management method for developing construction project delay risk predictive models. For instance, Owolabi (Owolabi *et al.*, 2018) developed a big data analytics-based predictive model for estimating delays in public-private partnership projects between 1992 to 2015 across Europe. Also, Yaseen (Yaseen *et al.*, 2020c) developed a hybrid artificial intelligence predictive model to mitigate construction projects delay in Diyala city, Iraq. Furthermore, Egwim (Egwim, Alaka, Toriola-Coker, Balogun and Sunmola, 2021a) developed a multilayer high-performance ensemble of ensembles predictive model utilising hyperparameter optimised ensemble machine learning techniques for construction projects in Nigeria among many others. Prior to the use of delay risk predictive models to mitigate delay, numerous investigations on causes of construction project delay by several researchers exists as evident in vast body of international literature, e.g. planning and scheduling deficiencies in Australia, delay in payment certificates in Ghana and poor site management in Malaysia by Shah (Shah, 2016); ground problems and inefficient structural connections for prefabricated components in both the UK and India by Agyekum-Mensah (Agyekum-Mensah and Knight, 2017) and Ji (Ji *et al.*, 2018) respectively and finally shortage of adequate equipment and poor communication among contracting parties in China by Chen (G.-X. Chen *et al.*, 2019) etc. Sequel to these studies, a few literature-reviews / systematic-reviews studies (Derakhshanfar *et al.*, 2019)(Sanni-Anibire, Mohamad Zin and Olatunji, 2020)(TAFESSE, 2021) were conducted to identify key drivers that causes construction project delay in the construction industry. These drivers put together forms the variables that inform construction project delay. However, a structured model of data that represents building elements with its usage spanning beyond the pre-construction phase to the post-construction phase known as BIM in Architecture, Engineering, and Construction industry has been introduced for a while now (Ameziane, 2000). Surprisingly, researchers (Al-Mohammad *et al.*, 2021; Cooney, Oloke and Gyoh, 2021; Evans *et al.*, 2021; Gharouni Jafari, Ghazi Sharyatpanahi and Noorzai, 2021; Saka and Chan, 2021; Silverio and Suresh, 2021; Tai, Zhang and Li, 2021) have revealed that the delay risk drivers that impact non-BIM-based and BIM-based construction projects are not necessarily the same. This is due to a variety of reasons, including the adoption of BIM for accurate geometric model development within the continuum modelling strategy (Kassotakis and Sarhosis, 2021), ability to Integrate BIM with emerging radio frequency identification technologies in structural engineering (Duan and Cao, 2020), use of BIM to control the geometry of arch pylon as it is being built while taking into account seasonal temperature fluctuations (Wang, Zhang and Wang, 2022), or since BIM now allows for the storage of modularisation data from past projects (Tidhar *et al.*, 2021), and as argued by (Johnston *et al.*, 2018) the weights of various structural forms are more readily available as a result of the increased usage of BIM and structural analysis models, and may be used to swiftly compute embodied carbon etc. The hypothesis of BIM was established in 1970 by Professor Charles Eastman at the Georgia Tech School of Architecture as building description systems (BDS) (Young *et al.*, 2008). Undeterred by its long-time existence, interest in BIM only took off few years ago. This present-day construction industry is predisposed by the wariness about BIM. Varying concerns around what exactly BIM is, whether BIM is only meant for large projects with complex geometries, how to change from the traditional design process to BIM among many others. BIM involves the creation and use of a three-dimensional (3D) virtual model that replicates the design, construction, and operation of a building. Also, BIM is perceived by facility managers as a tool used to improve building's performance and manage operations more efficiently throughout a building's life (Abbasnejad and Moud, 2013). The practical adoption of BIM by the Architecture, Engineering, and Construction industry for construction project started around the mid-2000s (Mohd and Latiffi, 2013). BIM was first implemented by the United States of America (USA) with example BIM-based construction projects seen in Sutter Medical Centre, Castro Valley California USA (Davis, 2007). Presently, BIM-based construction projects have been implemented in several countries such as "Sydney Opera House", in Australia; "One Island East Office Tower", in Hong Kong; "Crussel Bridge", in Helsinki, Finland; "National Cancer Institute (NCI)", in Putrajaya, Malaysia; "Barking Riverside Extension and Rail Station", in London, UK etc (Young *et al.*, 2008; Latiffi *et al.*, 2013). The rate of adoption of BIM differs between countries. For instance, according to a report by (Bernstein, 2010) 50% adoption rate amidst contractors was reached in North America compared to 24% of the counterparts in Western Europe, while Western Europe has higher percentage of BIM user rate(approximately 34%) compared to North American counterpart (18%). With reference to House (House *et al.*, 2007) the major benefit of BIM is its accurate geometrical representation of building parts in an integrated data. Some research (Johansen, 2015)(Jones, Young Jr. and Bernstein, 2008)(Mohd and Latiffi, 2013) indicated that BIM is generally used during pre-construction, construction and post construction stages to produce better project design; aid in decision making process; improve collaboration and communication among stakeholders; centralize data administration in a common data environment; reduce changes during construction; reduce conflict during construction; minimize risks in execution period; visualize design solution in 3D; reduce project delay; improve overall project quality and achieve better cost control/predictability. Undeterred by these BIM-based delay analysis studies with corresponding benefits of BIM to construction projects, there is no amalgamating study that has consolidated key drivers that affect BIM-based construction projects. As a result, using delay risk drivers from BIM-based construction projects is crucial and has been actively promoted in Architecture, Engineering, and Construction industry to accomplish early prediction, which is necessary in any robust predictive modelling to provide adequate time for correction (Narlawar, Chaphalkar and Sandbhor, 2019)(Amany, Taghizade and Noorzai, 2020). Consequently, this study aims to develop a comprehensive conceptual framework that will serve as a foundation for identifying the most critical delay risk drivers for BIM-based construction projects. To accomplish this aim, the following objectives will be used:

- To identify key delay risk drivers in BIM-based construction projects that have significant impact on the performance of delay risk predictive modelling via systematic review of literature.

- To examine the systematic review's summary of findings and rank the discovered delay risk drivers to determine which are the most critical.

- To Identify and present BIM tools used to potentially mitigate delay from vast body of literature in first and second objectives.

The contribution of this study is therefore to fill the gap in lack of a conceptual framework for selecting key delay risk drivers for BIM-based construction projects, which has hampered scientific progress toward development of extremely effective delay risk predictive models for BIM-based construction projects. Furthermore, this study will solve the challenge of analysing variables under all existing drivers to find the most critical ones before developing delay risk predictive models for BIM-based construction projects in future research. As such, only variables that fall under the selected categories will be examined, thus significantly enhance the efficiency of delay risk predictive modelling for BIM-based construction projects. The next section details the research methodology, which begins with an explanation of the systematic review methods. The data analysis section follows afterwards (section 3), explaining the systematic review's data analysis step by step. The analysis' findings are then presented discussed in section 4. The discussion part examines how the findings connect to existing theories, while the conclusion summarises the findings in section 5.

3.3 Methodology

Pragmatism is the philosophical paradigm used in this study. This is because it focuses on practical applied research using several viewpoints to aid in data interpretation such that depending on the research question, either observable occurrences or subjective meanings might give acceptable knowledge (Saunders, Lewis and Thornhill, 2019).This systematic review is conducted in line with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) (Moher *et al.*, 2009). PRISMA is a guideline for conducting systematic reviews and meta-analyses that includes a 27-item checklist and a four-phase flow diagram. It was created by a consortium of twenty-nine professors in the medical community with the aim of improving the clarity and consistency of systematic reviews. PRISMA was therefore, chosen over other existing guidelines because of its comprehensiveness, its use in a variety of disciplines around the world outside of medicine, and its ability to improve accuracy through articles (Pahlevan Sharif, Mura and Wijesinghe, 2019a). In order to record the research process and inclusion criteria, a guideline was created in advance. A thorough literature search was conducted to find articles for this study. More precisely, only articles published until 1st of August 2021 in SCOPUS and American Society of Civil Engineers (ASCE) databases were used as its primary source of information for the search. These databases were chosen because the formal is the "largest abstract and citation database of peer-reviewed literature" (Cantú-Ortiz and Fangmeyer, 2017) while the latter is the "world's largest publisher of civil engineering content" (ASCE, 2010). The abstract, title and keywords of publications in these databases were searched using the following search terms: ("BIM" OR "Building Information Model*" OR "Building Information Manage*" AND ("Delay*" OR "Schedule Overrun*")) with no date, language, and article type restrictions. The eligibility criteria are thus, all articles stored in these databases whose title, abstract or keywords matched the search terms. This is because titles, abstracts and keywords serve as cues for article search. A total of three hundred and eighty-eight articles were identified (see Figure 3.1). The author's name, author's affiliation, articles title, articles abstract, authors keywords, publication year and source title were exported to a Comma-

Separated Values (CSV) file. At first, I screened each title and abstract in the exported CSV file independently. Full text of articles from the file that fell within the eligibility criteria (mentioned above) were subsequently accessed and evaluated.

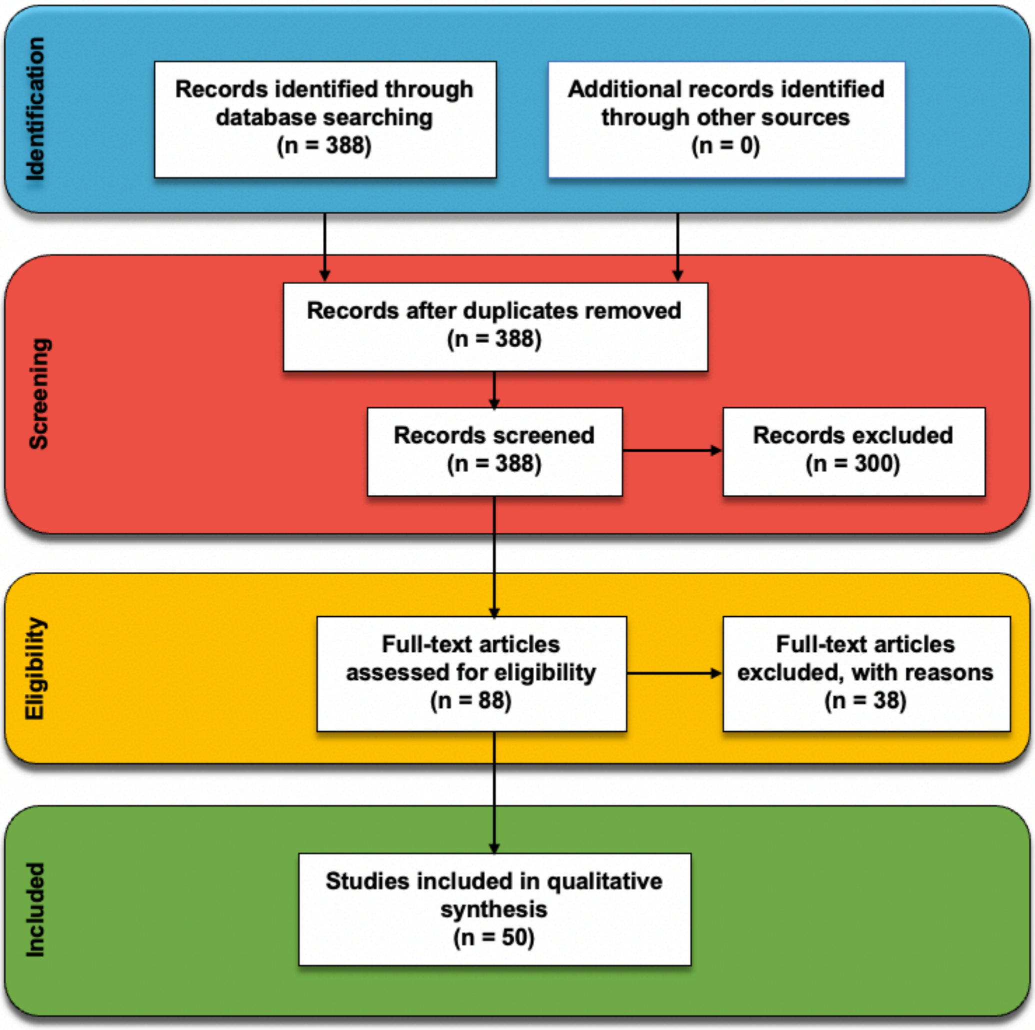

Figure 3.1: Flow diagram of research selection process.

The bibliographic information for the included articles, as well as the necessary elements from the PRISMA checklist (with a few changes) were added to the data management CSV file. Meta-analysis study findings (items 12–16 and 19–23 from the PRISMA checklist) were excluded because they were related to meta-analyses only, outside the scope of this research. In order to optimize the extracted information and to code accordingly, a pilot test of fifteen randomly chosen included articles was carried out. Finally, all articles used were thoroughly examined for data extraction (research aim, project type, country/region, research method(s), tool(s), etc.) and coded as summarized in Figure 3.1. This figure displays the flow diagram of the research selection process. It details the total number of articles identified through database search, total number of articles screened based on eligibility criteria and total number of these articles whose full text were access and finally total number of these articles that were used for analysis in this study. The ranking of delay risk drivers in this study was based on citation frequency. This method involves counting the number of times each delay driver was cited across the reviewed literature. Citation frequency serves as an indicator of the relative importance and impact of each delay driver within the context of BIM-based construction projects. For each of the nine categories, delay risk drivers were ranked according to the number of times they were cited across the reviewed articles. For instance, contractor-related drivers, which include issues such as shortage of resources, financial difficulties, and inaccurate resource planning, were the most frequently cited, thus ranking highest. In contrast, categories with fewer citations, such as consultant-related drivers, were ranked lower. The ranking process is visually represented in Table 3.3 of the study, which details the frequency of citations for each delay driver. This table serves as a quantitative foundation for the subsequent conceptual framework, which identifies the most critical delay risk drivers for BIM-based construction projects. The use of citation frequency as a ranking method is well-supported in academic literature. According to Borrego et al. (2018), citation analysis is a reliable indicator of the impact and relevance of specific topics within a research field. This method ensures that the ranking reflects the collective judgment of the academic community, thereby enhancing the validity of the findings. Furthermore, citation frequency analysis mitigates potential biases that could arise from subjective assessments. As noted by Garfield (2006), citation counts provide an objective measure that is less susceptible to personal or institutional biases. This objectivity is crucial in constructing a robust and credible conceptual framework for identifying key delay risk drivers in BIM-based projects.

3.4 Result and Analysis

Performing the search terms on the electronic databases yielded a total of three hundred and eighty-eight articles (see Figure 3.1). A total of three hundred articles which either despite meeting the eligibility criteria, doesn't relate to the objective of this study or due to my limited access to subscription-based articles were excluded from this research (For example, Yan (Yan *et al.*, 2008), (Studebaker, 2014), (Surendhra Babu and Hayath Babu, 2018), (Moselhi, Bardareh and Zhu, 2020) etc) were not relevant to the effect of BIM on construction projects delay. Thus, remaining a total of eighty-eight articles. Furthermore, I read the full-length of the remaining articles carefully to ensure that they were relevant. Thirty-eight of these were further discarded because they did not satisfy the eligibility criteria. Consequently, a total of fifty articles – thirty-one journals, seventeen conference proceedings, and two books remained and were used for this research. Table 3.1 below summarizes the extracted variables and information as the main characteristics of all the reviewed articles towards the discovery of key delay risk drivers for BIM-based construction projects. Furthermore, exploratory analysis of the data from Table 3.1 revealed seventeen different types of construction projects (see Figure 3.2) from the articles reviewed in this study, demonstrating the validity of the conceptual framework to be built using these articles as being capable of working in a variety of construction projects. In addition, a count of research methods per project revealed that the most used research method, quantitative research, was most prevalent in the general construction project category as shown in Figure 3.2.

Table 3.1: The summary of study characteristics.

SN	Study	Delay risk drivers	Source Type	Country	Project Type	Research Method	BIM Tool Used
1	(Narlawar, Chaphalkar and Sandbhor, 2019)	I. Conflicts between contractor and sub-contractor (s). II. Natural disasters e.g., floods, earthquakes etc.	Journal	India	Residential Building	Quantitative	Autodesk Revit, and Autodesk Navisworks Manage
2	(Narlawar G.S., Chaphalkar N.B., 2017)	I. Changes in specifications. II. Poor contract management.	Journal	Malaysia	General construction	Mixed methods (semi-structured interview and Questionnaire)	None
3	(Ryu *et al.*, 2015)	I. Traffic restrictions. II. Changes in site topology after design.	Journal	Korea	Tunnel	Quantitative	Simulated Annealing (SA) and BIM
4	(Nawi *et al.*, 2014)	I. Ineffective project supervision.	Journal	Malaysia	Industrialised Building System (IBS) Construction Projects	Qualitative	Integrated Project Delivery and BIM
5	(Li *et al.*, 2017)	I. Client's cash flow issues.	Journal	Hong Kong	General construction	Qualitative and Experimental	RFID-enabled BIM
6	(Subramani and Ammai, 2018)	I. Contractor's cash flow issue.	Journal	Multinational	General construction	Qualitative	Primavera

SN	Study	Delay risk drivers	Source Type	Country	Project Type	Research Method	BIM Tool Used
7	(Amany, Taghizade and Noorzai, 2020)	I. Ineffective project plan. II. Shortage of materials on market.	Journal	Multinational	General construction	Quantitative	Last Planner System (LPS) technique with Revit BIM
8	(Husin, 2019)	I. Conflicts between contractor and consultant.	Journal	Indonesia	High-rise building	Quantitative	None
9	(Abdelbary, Edkins and Dorra, 2020)	I. Schedule/program of work.	Journal	Egypt	General construction	Quantitative	None
10	(Shin, Lee and Kim, 2018)	I. Limited space for temporal equipment.	Journal	South Korea	Railway Site	Mixed methods	None
11	(Park and Lee, 2017)	I. Discrepancies in contract documents.	Journal	Korea	hospital building	Quantitative	2D drawings and BIM
12	(Park *et al.*, 2017)	I. Changes in client requirements. II. Reworks due to error in construction.	Journal	Multinational	General construction	Experimental	Web and Database-Based 4D BIM

SN	Study	Delay risk drivers	Source Type	Country	Project Type	Research Method	BIM Tool Used
13	(Sepasgozar *et al.*, 2019)	I. Late delivery of construction materials. II. Inflation or sudden increase in commodities.	Journal	Generic	General construction	Qualitative, Quantitative and Experimental	None
14	(Chiponde *et al.*, 2017)	I. Building permit from authorities. II. Corruption issues. III. Variation in structural design.	Journal	Zambia	General construction	Mixed methods	None
15	(Sami Ur Rehman *et al.*, 2020)	I. Schedule/program of work. II. Labour dispute or strikes.	Journal	Multinational	General construction	Mixed methods	None
16	(Giel and Issa, 2013)	I. Reworks due to error in construction.	Journal	Multinational	General construction	Qualitative	None
17	(Jang and Lee, 2018)	I. Contractors' financial difficulties.	Journal	Korea	General construction	Quantitative	None
18	(Chen and Tang, 2019)	I. Staff's productivity level. II. Shortage of resources (human resources, machinery, equipment etc).	Journal	China	Residential Building	Quantitative	None

SN	Study	Delay risk drivers	Source Type	Country	Project Type	Research Method	BIM Tool Used
19	(Ern, Ooi and Al-Ashmori, 2020)	I. Contractor's poor decision	Journal	Malaysia	General construction	Quantitative	None
20	(Elhusseiny, Nosair and Ezeldin, 2021)	I. Client's cash flow issues. II. Consultant's cash flow issues. III. Conflicts between contractor and consultant	Journal	Egypt	General construction	Quantitative	None
21	(Malacarne *et al.*, 2018)	I. Contractors' financial difficulties.	Journal	Italy	General construction	Qualitative	None
22	(Handayani, Likhitruangsilp and Yabuki, 2019)	I. Change in economic conditions. II. Shortage of resources (human resources, machinery, equipment etc).	Journal	Thailand	Story Buildings	Qualitative	Autodesk Revit
23	(Wong, Zhou and Chan, 2018)	I. Variation in structural design. II. Reworks due to error in construction. III. Poor communication amongst stakeholders	Journal	China	General construction	Quantitative	3D/4D BIM
24	(Tserng, Ho and Jan, 2014)	I. Poor site investigation or management. II. Staff's productivity level.	Journal	Taiwan	General construction	Experimental	Construction BIM-assisted Schedule

SN	Study	Delay risk drivers	Source Type	Country	Project Type	Research Method	BIM Tool Used
							management (ConBIM-SM)
25	(Liu *et al.*, 2020)	I. Ineffective project plan. II. Pandemic	Journal	China	Residential Building	Experimental	None
26	(Vilventhan, Razin and Rajadurai, 2020)	I. Delay in site delivery.	Journal	India	Infrastructure construction project	Qualitative and Experimental	Geographic Information System (GIS)
27	(Delgado *et al.*, 2015)	I. Site accident. II. Political influence.	Journal	Spain	Residential Building	Experimental	Web3D visualizer and BIM
28	(Hossain *et al.*, 2018)	I. Late payment by the client.	Journal	Singapo re	Story Building	Experimental	None
29	(Stegnar and Cerovšek, 2019)	I. Damage storage materials. II. Contractors' financial difficulties. III. Natural disasters e.g., floods, earthquakes etc. IV. Unforeseen site conditions e.g., unanticipated	Journal	Slovenia	Office Building	Quantitative	None

SN	Study	Delay risk drivers	Source Type	Country	Project Type	Research Method	BIM Tool Used
		groundwater, quicksand, mud, rock formation etc.					
30	(Khalesi *et al.*, 2020)	I. Building material's price fluctuation. II. Inflation or sudden increase in commodities. III. Unskilled labourers	Journal	Multinational	Residential Building	Quantitative	None
31	(Tahir *et al.*, 2018)	I. Use of outdated technologies by staffs.	Journal	Multinational	General construction	Qualitative	None
32	(Vahdatikhaki and Mawlana, 2017)	I. Project quality control.	Conference	Canada	General construction	Experimental	None
33	(Tahir Muhammad *et al.*, 2019)	I. Building material's price fluctuation. II. Equipment breakdown.	Conference	Malaysia	General construction	Quantitative	None
34	(Btoush and Harun, 2017)	I. Use of outdated technologies by staffs. II. Change order. III. Damage storage materials. IV. Equipment breakdown.	Conference	Jordan	General construction	Qualitative	3D/4D BIM
35	(Dallasega *et al.*, 2019)	I. Change order. II. Discrepancies in contract documents	Conference	Italy	Hospital construction project	Experimental	3D/4D BIM

SN	Study	Delay risk drivers	Source Type	Country	Project Type	Research Method	BIM Tool Used
36	(Diaz *et al.*, 2019)	I. Site accident. II. Labour dispute or strikes.	Conference	Peru	Road projects	Experimental	Civil 3D, Revit and IfraWorks 360
37	(Ansah *et al.*, 2016)	I. Variation in structural design.	Conference	Malaysia	General construction	Qualitative	3D/4D BIM
38	(Chou and Chen, 2017)	I. Late delivery of construction materials	Conference	Taiwan	General construction	Quantitative	3D/4D BIM
39	(Wang *et al.*, 2015)	I. Inaccurate budgeting	Conference	Australia	Liquefied Natural Gas Construction	Quantitative	Laser Scanning, Mobile Computing and RFID
40	(Kermanshahi *et al.*, 2020)	I. Shortage of materials on market. II. Ineffective government regulations.	Conference	Malaysia	Police Headquarter Building	Quantitative	Autodesk Revit and Autodesk Navisworks
41	(Couto and Ericson, 2017)	I. Ineffective project supervision. II. Building material's price fluctuation	Conference	USA	Airport Building	Experimental	3D/4D BIM and Laser Scanner
42	(Vacanas *et al.*, 2015)	I. Shortage of resources (human resources, machinery, equipment etc).	Conference	Multinational	Infrastructure construction project	Qualitative	3D/4D BIM and Unmanned Aerial Vehicle (UAV)

SN	Study	Delay risk drivers	Source Type	Country	Project Type	Research Method	BIM Tool Used
43	(Suermann and Issa, 2009)	I. Poor site investigation or management.	Conference	USA	Air Force Building	Experimental	3D/4D BIM
44	(Liu and Liu, 2020)	I. Inaccurate resource planning.	Conference	China	Substation Construction	Qualitative and Experimental	Glodon 5D BIM
45	(Ibrahim, Hashim and Ahmad Jamal, 2019)	I. Use of outdated construction methods (e.g., precast twin wall technology, 2D drawing etc.).	Conference	Malaysia	General construction	Quantitative	Traditional 3D/4D BIM
46	(Zhang and Liu, 2021)	I. Conflicts between contractor and sub-contractor(s). II. Level of BIM usage. III. Inaccurate resource planning.	Conference	Multinational	General construction	Qualitative	3D/4D BIM
47	(Ocheoha and Moselhi, 2013)	I. Limited space for temporal equipment.	Conference	Canada	General construction	Quantitative	3D/4D BIM
48	(Sigalov and König, 2018)	I. Inaccurate resource planning.	Conference	Multinational	General construction	Experimental	3D/4D BIM
49	(Teng *et al.*, 2013)	I. Unfavourable weather conditions.	Book	China	General construction	Qualitative	Integrated Project Delivery and BIM

SN	Study	Delay risk drivers	Source Type	Country	Project Type	Research Method	BIM Tool Used
		II. Use of outdated construction methods (e.g., precast twin wall technology, 2D drawing etc.). III. Changes in client requirements.					
50	(Srao, Rai and Mann, 2018)	I. Level of BIM usage.	Book	India	General construction	Qualitative	3D/4D BIM

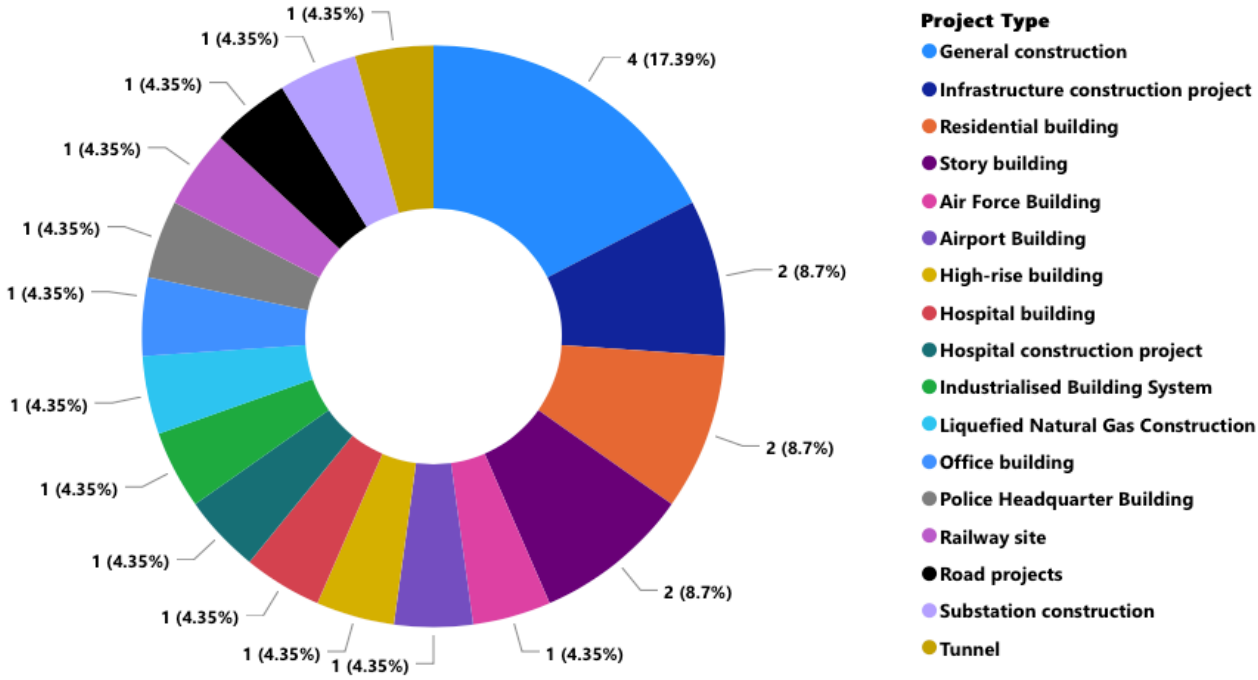

Figure 3.2: A count of research methods by construction project type

By observation (see Table 3.2), it appears that the global construction industry has only paid more attention on the effect of BIM to its project delay in the continent of Asia (in terms of number of publications by researchers and practitioners used in this study) when compared to other continents of the world. This can further suggest the possible increase in the rate of adoption of BIM and a lesser number of delayed projects in that region as was evidently demonstrated in the construction of a 1,000-bed Huoshenshan and 1500-beds Leishenshan Hospitals for COVID-19 patients in Wuhan, China between the 23rd of January and 2nd of February 2020.

Table 3.2: List of research articles used and their respective countries/continents.

SN	Continent	Countries (# of Research Articles)	Total Countries	Total Articles
1	Africa	Zambia (1), Egypt (2),	2	3
2	North America	Canada (2), USA (2),	2	4
	South America	Peru (1),	1	1
3	Asia	India (3), Malaysia (7), Korea (4), China (5), Thailand (1), Taiwan (2), Jordan (1), Singapore (1), Slovenia (1), Indonesia (1), Hong Kong (1)	11	27
4	Europe	Italy (2), Spain (1), Germany (1)	3	4
5	Oceania	Australia (1)	1	1
6	Multicontinental	Multinational (10)	n/a	10

To obtain the key delay risk drivers necessary for development of delay risk predictive models for BIM-based construction projects (as part of the objective of this research) I extracted the top drivers each from the fifty research articles used in this systematic review and grouped them into nine delay categories: *owners related driver, contractor related driver, consultant related driver, design related driver, labour related driver, equipment related driver, project related driver, suppliers related driver, and external related driver.* Furthermore, we ranked them based on these categories (i.e., for each category according to number of times each driver occurred in the fifty articles used) and used the outcome of the ranking to propose a conceptual framework necessary for developing delay risk predictive models for BIM-based construction projects as shown in Table 3.3 and Figure 3.3 below respectively.

Table 3.3: Delay risk drivers Category Ranking

Delay risk drivers	List of Sources	Number of Sources	Delay Category	Total Number of Citations	Category Ranking
Shortage of resources (human resources, machinery, equipment etc)	(Vacanas *et al.*, 2015), (Chen and Tang, 2019), (Handayani, Likhitruangsilp and Yabuki, 2019)	3			
Contractors' financial difficulties	(Jang and Lee, 2018), (Malacarne *et al.*, 2018), (Stegnar and Cerovšek, 2019)	3			
Level of BIM usage	(Srao, Rai and Mann, 2018), (Zhang and Liu, 2021), (Liu *et al.*, 2020)	3			
Inaccurate resource planning	(Zhang and Liu, 2021), (Sigalov and König, 2018), (Kermanshahi *et al.*, 2020)	3			
Schedule/program of work	(Sami Ur Rehman *et al.*, 2020), (Abdelbary, Edkins and Dorra, 2020)	2			
Limited space for temporal equipment	(Shin, Lee and Kim, 2018), (Ocheoha and Moselhi, 2013)	2			
Ineffective project supervision	(Couto and Ericson, 2017), (Nawi *et al.*, 2014)	2			
Poor site investigation or management	(Suermann and Issa, 2009), (Tserng, Ho and Jan, 2014)	2			
Contractor's poor decision	(Ern, Ooi and Al-Ashmori, 2020)	1			

Delay risk drivers	List of Sources	Number of Sources	Delay Category	Total Number of Citations	Category Ranking
Ineffective project plan	(Amany, Taghizade and Noorzai, 2020)	1	Contractor related driver	24	1
Contractor's cash flow issue	(Subramani and Ammai, 2018)	1			
Project quality control	(Vahdatikhaki and Mawlana, 2017)	1			
Traffic restrictions	(Ryu *et al.*, 2015)	1			
Building material's price fluctuation	(Tahir Muhammad *et al.*, 2019), 72], (Couto and Ericson, 2017)	3			
Inflation or sudden increase in commodities	72], (Sepasgozar *et al.*, 2019)	2			
Site accident	(Diaz *et al.*, 2019), (Delgado *et al.*, 2015)	2			
Natural disasters e.g., floods, earthquakes etc.	(Narlawar, Chaphalkar and Sandbhor, 2019), (Stegnar and Cerovšek, 2019)	2			
Pandemic	(Liu and Liu, 2020)	1			
Shortage of materials on market	(Amany, Taghizade and Noorzai, 2020), (Kermanshahi *et al.*, 2020)	2			
Change in economic conditions	(Handayani, Likhitruangsilp and Yabuki, 2019)	1			
Political influence	(Delgado *et al.*, 2015)	1			

Delay risk drivers	List of Sources	Number of Sources	Delay Category	Total Number of Citations	Category Ranking
Ineffective government regulations	(Kermanshahi *et al.*, 2020)	1	External related driver	19	2
Unforeseen site conditions e.g., unanticipated groundwater, quicksand, mud, rock formation etc	(Stegnar and Cerovšek, 2019)	1			
Building permit from authorities	(Chiponde *et al.*, 2017)	1			
Corruption issues	(Chiponde *et al.*, 2017)	1			
Unfavourable weather conditions	(Teng *et al.*, 2013)	1			
Labour dispute or strikes	(Diaz *et al.*, 2019), (Sami Ur Rehman *et al.*, 2020)	2	Labour related driver	10	3
Staff's productivity level	(Tserng, Ho and Jan, 2014), (Chen and Tang, 2019)	2			
Use of outdated construction methods (e.g., precast twin wall technology, 2D drawing etc.).	(Teng *et al.*, 2013), (Ibrahim, Hashim and Ahmad Jamal, 2019)	2			
Use of outdated technologies by staffs	(Btoush and Harun, 2017), (Tahir *et al.*, 2018)	2			
Changes in site topology after design	(Ryu *et al.*, 2015)	1			

Delay risk drivers	List of Sources	Number of Sources	Delay Category	Total Number of Citations	Category Ranking
Unskilled labourers	(Khalesi *et al.*, 2020)	1			
Change order	(Btoush and Harun, 2017), (Dallasega *et al.*, 2019)	2	Owners related driver	9	4
Client's cash flow issues	(Li *et al.*, 2017), (Elhusseiny, Nosair and Ezeldin, 2021)	2			
Changes in client requirements	(Park and Lee, 2017), (Teng *et al.*, 2013)	2			
Delay in site delivery	(Vilventhan, Razin and Rajadurai, 2020)	1			
Changes in specifications	(Narlawar G.S., Chaphalkar N.B., 2017)	1			
Late payment by the client	70]	1			
Variation in structural design	(Wong, Zhou and Chan, 2018), (Chiponde *et al.*, 2017), (Ansah *et al.*, 2016)	3	Design related driver	6	5
Reworks due to error in construction	(Park and Lee, 2017), (Giel and Issa, 2013), (Wong, Zhou and Chan, 2018)	3			
Consultant's cash flow issues	(Elhusseiny, Nosair and Ezeldin, 2021)	1	Consultant related driver		
Inaccurate budgeting	(Wang *et al.*, 2015)	1			
Poor communication amongst stakeholders	(Wong, Zhou and Chan, 2018)	1			

Delay risk drivers	List of Sources	Number of Sources	Delay Category	Total Number of Citations	Category Ranking
Poor contract management	(Narlawar G.S., Chaphalkar N.B., 2017)	1		4	6
Discrepancies in contract documents	(Dallasega *et al.*, 2019), (Park *et al.*, 2017)	2	Project related driver	4	6
Damage storage materials	(Btoush and Harun, 2017), (Stegnar and Cerovšek, 2019)	2			
Conflicts between contractor and sub-contractor (s)	(Narlawar, Chaphalkar and Sandbhor, 2019), (Teng *et al.*, 2013), (Zhang and Liu, 2021)	3	Contractor & sub-contractor related driver	3	7
Late delivery of construction materials	(Sepasgozar *et al.*, 2019), 80]	2	Suppliers related driver	2	8
Conflicts between contractor and consultant	(Elhusseiny, Nosair and Ezeldin, 2021), (Husin, 2019)	2	Contractor & consultant related driver	2	8

Delay risk drivers	List of Sources	Number of Sources	Delay Category	Total Number of Citations	Category Ranking
Equipment breakdown	(Btoush and Harun, 2017), (Tahir Muhammad *et al.*, 2019)	2	Equipment related driver	2	8

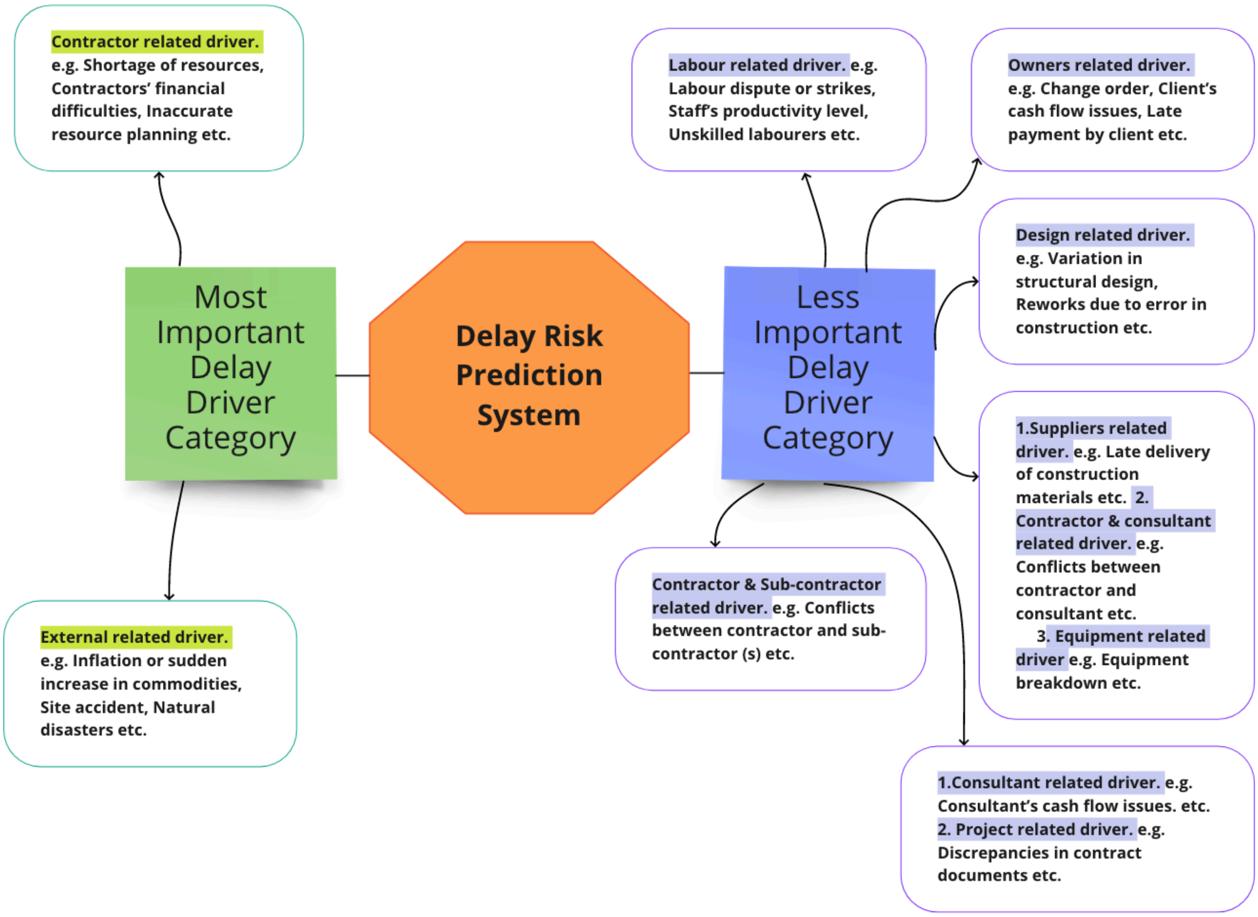

Figure 3.3: Conceptual Framework of delay risk drivers necessary for developing delay risk predictive models for construction projects.

This conceptual framework details contractor related driver and external related driver as the most important delay driver categories (see Figure 3.3) to be considered when developing delay risk predictive models for BIM-based construction projects. This is justified based on citation frequency. More precisely, these categories had the highest number of citations from research articles used in this study. Conversely, labour related driver, owner's related driver; design related driver; suppliers related driver; consultant related driver; contractor & consultant related driver; equipment related driver, project related driver; and contractor & sub-contractor related driver were considered less important delay driver categories based on their respective low citation frequencies. Suppliers related driver, contractor & consultant related driver, and equipment related driver were grouped together because they all had equal number of citation (see Table 3.3). Similarly, because they all received the same number of citations, consultant related driver, and project related driver were grouped together. In order to identify BIM tools used to potentially mitigate delay from vast body of literature, fifteen BIM tools were extracted from the fifty research articles (see Table 3.1) reviewed in this study which includes: *Primavera; Autodesk Revit; Autodesk Navisworks; BIM-based Simulated Annealing (SA); BIM-Integrated Project Delivery (BIM-IPD); Last Planner System (LPS) technique with Revit BIM; BIM-based Laser Scanning; Web and Database-based 4D BIM; Construction BIM-assisted Schedule Management (ConBIM-SM);*

InfraWorks 360; BIM-based Unmanned Aerial Vehicle (UAV); BIM-based Web3D Visualizer; BIM-based Geographic Information System (BIM-GIS); Glodon 5D BIM; RFID-enabled BIM. Each of these tools will be summarised with facts illustrated by using quotations from respective research article used in this study, and discussions are followed up by a larger study of the literature.

Primavera: "Primavera is a project portfolio management software for businesses" (Subramani and Ammai, 2018). It coordinates with various software applications and carries out all types of management, such as mission management, object management, teamwork, and capacity management. More precisely, Primavera can be used to create a standard project's work breakdown structure (WPS) and allocate time, expense, and resources to its activities, allowing Primavera to calculate the project's length and cost.

Autodesk Revit: "Autodesk Revit is a BIM software that comprises three modules: Structure, Architecture, and MEP (Mechanical, Electrical, and Plumbing) used to create 3D models" (Handayani, Likhitruangsilp and Yabuki, 2019). It is downward compatible which implies other dimensional design can be imported into it, e.g. A 2D plan from AutoCAD can be exported and traced by its objects. Furthermore, 3D models developed in Revit can be used to extract quantities such as slab, column footing, door, beam, door etc in a typical construction project as 'Schedule' and 'Material Take-off'.

Autodesk Navisworks: Autodesk Navisworks essentially brings together models (created in any 2D/3D software e.g., REVIT, TEKLA, AUTOCAD, BENTLY etc with their respective proprietary formats), "combines them into a single model and allows them to be interpreted, navigated, measured, and analysed in one environment" (Kermanshahi *et al.*, 2020). Subsequently, a schedule and cost parameters can be added to this single model to develop a 4D and 5D models respectively which can further be used for schedule and cost analysis / simulation.

BIM-based Simulated Annealing (SA): "Simulated Annealing (SA) is an optimization technique for combinatorial and other problems that uses an analogy to the way a metal cools and freezes into a minimal energy crystalline structure (the annealing process) and the search for a minimum in a more general method" (Ryu *et al.*, 2015). When integrated with BIM, SA can allow contractors to assess risks in excavation costs and durations in a typical tunnel construction with full details about ground conditions acquired before construction.

BIM-Integrated Project Delivery (BIM-IPD): IPD is "a project delivery strategy" that blends individuals, procedures, organizational mechanisms, and activities into a mechanism that collaboratively harnesses skills and insights of all project partners to optimize performance, increase value, minimize cost, and maximize productivity during the planning, fabrication, and construction phases. According to reviewed papers, this integration approach enables team members to take advantage of BIM by constructing a graphical model of each phase of the development process (Nawi *et al.*, 2014). Also, BIM-IPD can help by facilitating cooperation between parties in Architecture, Engineering, and Construction industry.

Last Planner System (LPS) technique with Revit BIM: The LPS is one of the theories and management strategies of lean construction, and it fits well with the four stages of project programs, including "master plan (SHOULD), scheduling plan (CAN), forecast plan (WILL), and commitment in planning (DID) based on lean development principles" (Subramani and Ammai, 2018). The primary goal of LPS is to improve the dependability of events planned prior to the implementation level. Hence when integrated with BIM, it can be used to create comprehensive schedule planning according to a developed model.

BIM-based Laser Scanning: "Laser scanning is a measuring technique that produces a point cloud that represents the 3D surface of an object quickly and accurately (Wang *et al.*, 2015). Consequently, it can be used for milestone progress tracking in a construction project by changing the coordinate system of the as-built data to align to the as-planned BIM model.

Web and Database-based 4D BIM (WebD-BIM): "WebD-BIM technologies leverages the power of integrated Web Graphics Library (WebGL) and a central database" (Couto and Ericson, 2017). In this integration, a WebGL can be used to render 3D graphics for sharing daily BIM information in construction via project participants mobile/smart devices while a central database will contain the rules to automatically determine and update status of individual BIM objects.

Construction BIM-assisted Schedule Management (ConBIM-SM) system: "The ConBIM-SM system is designed for all BIM-related participants via a user-friendly portal which serves as a real-time, updated as-built schedule channel for project engineers" (Liu *et al.*, 2020). It is also a solution that gives a single, unified database linked to the as-built models' files with different levels of access determined by user roles. Hence participants can access the BIM model SM information entry and updates, based on their responsibilities in the ConBIM-SM system.

InfraWorks 360: "InfraWorks 360 is a visual 3D design and communication BIM software for civil engineers, designers, and land planners" (Diaz *et al.*, 2019). It can be used to build more realistic models from any database for road, rail, buildings and pipe networks etc.

BIM-based Unmanned Aerial Vehicle (BIM-UAV): "UAV refers to a remotely operated aircraft or helicopter that is fitted with precision sensors such as inertial motion units (IMU) and gyroscopes to recognize the aircraft's alignment and location" (Vacanas *et al.*, 2015). It has a microcomputer that allows for autonomous navigation without the need for any manual intervention from the pilot to capture high-resolution images from a variety of angles in a cost-effective and reliable manner. In terms of construction project time control, BIM-UAV can be used to ensure more reliable data processing and accurate updating of the works plan, allowing for complex development project management.

BIM-based Web3D Visualizer: "A Web3D visualization tool", in general, helps users to explore outdoor and indoor areas, as well as apply labels corresponding to cubic cell boundaries, thereby providing support for activity monitoring in BIM-based construction processes and the creation of a systemic relation that can grow over time (Diaz *et al.*, 2019). As a result, we have an ''augmented" model that includes unique metadata to visualize objects at various level of details (LoDs) of an open standardised data model like City Geographic Markup Language (CityGML).

BIM-based Geographic Information System (BIM-GIS): "GIS is a location-based information framework that allows for data mapping, querying, simulation, and analysis" (Delgado *et al.*, 2015). BIM-GIS can be applied to underground utility infrastructure to visualize emerging underground utilities, identify utility collisions, and manage utility details thus managing conditional information such as condition score, inspection and maintenance records of individual utilities.

Glodon 5D BIM: “Glodon 5D BIM software is a five-dimensional representation of the spatial and functional features of standard Architecture, Engineering, and Construction industry projects focused on computer modelling” (Liu and Liu, 2020). Its main goal is to introduce a precise and efficient cost estimate based on a 3D digital model, which prevents errors created by manual calculations or estimations. 5D BIM, which is focused on 3D modelling, schedule simulation (4D), and cost prediction, is also seen as a potential trend for the application of multi-disciplinary experience and decision making for building design and construction (5D).

RFID-enabled BIM: “RFID is a type of automated identification technology” that collects statistical construction site progress through radio frequency acquisition and transmission of construction site data. It is most often used to display product attributes, material data, and positioning (Li *et al.*, 2017).

3.5 Discussion

One of the most evident findings (see Figure 3.3) of our analysis concerns the key most important delay driver categories (based on citation frequency) necessary for development of delay risk predictive models for BIM-based construction projects which are first contractor related driver and then external related driver respectively. This is very interesting as studies (Ilozor and Kelly, 2012; Piroozfar *et al.*, 2019; Elghaish *et al.*, 2020) have shown that Integrated Project Delivery (IPD) method which includes, Construction Manager at Risk (CMAR) and design-build procurement process, where a contractor is brought together with architect(s) designers(s) and client(s) to help develop a project plan that is buildable and meets the budgeting constraints from the design phase is the most effective project delivery method to be used in BIM-based construction projects due to the coordination of contractors, subcontractors, and fabricators in a design process, allowing the virtual prototype of the building to be fully studied before construction even begins. Thereby providing huge benefits such as cost estimation, quality control, schedule plan, constructability, expedited issue resolution among many others at the design phase with great outcomes or value for money for the client/ owner. As a result, construction projects will potentially be in the anticipated budget range and is buildable in the anticipated timeframe. One thing to keep in mind though is that typically client(s) will expect the contractor to identify these issues at design phase, so that when a final cost is agreed upon, it doesn’t change during construction. This type of project delivery method thus places an additional burden on the contractor as they are expected to be experts. However, solace can be taken from the fact that one of the most important benefits from BIM for the contactor is derived from the close coordination that can be achieved when all major subcontractors use the BIM model for detailing their portion of work thus enabling accurate class detection and correction of clashes before they become a problem onsite. Therefore, contractors are strongly encouraged to include subcontractors and fabricators in their BIM efforts. Furthermore, another important finding of our analysis concerns the variety of different BIM tools used in different construction projects and their unanimous effect on project delay. More specifically, while only one of the studies (Suermann and Issa, 2009) was neutral on the effect of BIM on delay, others concluded BIM's positive effect (use of BIM/BIM tool to achieve research aim/objective(s)) on construction projects delay even though they implemented BIM differently on different construction projects across different regions of the world (see table 3.1). This is good as these BIM tools allows construction stakeholders like BIM coordinators and structural engineers to have in the same place, in one file, all their views, plans, section, elevation and 3D, but also any kind of object. For instance, the Autodesk Revit software identified in this study as one of the BIM tools can be used in design authoring as schedule of quantity take-off, which could be a door schedule or window schedule and could be related to anything from materials to quantity. Also, Autodesk Navisworks BIM tool identified in this study can be used during 3D coordination process as a clash detection software to eliminate system conflict prior to installation thus decreasing construction time (mitigate delay risks) and increase productivity onsite. It is important to note however, that merely using BIM tools on a construction project does not equate to success if the owner's goals for the project are not clearly set, and BIM requirements do not correlate to achieve those goals. In order to use BIM/BIM tools effectively, a construction project should begin with defining the BIM requirements. Owners should establish processes, standards, and deliverables for BIM-based construction projects that can be continually shared and agreed upon by the owners and the rest of the project team.

3.6 Conclusion

Various attempts at mapping and synthesising the current body of information have emerged in the literature since BIM in Architecture, Engineering, and Construction industry (as an area of inquiry) has increasingly grown and supposedly reached intellectual and analytical sophistication in decades. Systematic analyses have emerged as one of the key methods for assessing the key delay risk drivers necessary for development of delay risk predictive models for BIM-based construction projects among the various forms of research undertaken by scholars. Undeterred by these BIM-based delay analysis studies with corresponding benefits of BIM to construction projects, there is no amalgamating study that has consolidated key drivers that affect BIM-based construction projects. As a result, using delay risk drivers from BIM-based construction projects is crucial and has been actively promoted in Architecture, Engineering, and Construction industry to accomplish early prediction, which is necessary in any robust predictive modelling to provide adequate time for correction. Thus, led to the development of a comprehensive conceptual framework that serves as a foundation for identifying the most critical delay risk drivers for BIM-based construction projects. This was accomplished by first identifying key delay risk drivers in BIM-based construction projects that have significant impact on the performance of delay risk predictive modelling via systematic review of literature. Secondly, by examining the systematic review's summary of findings and ranking the discovered delay risk drivers to determine which are the most critical. Finally led to the identification and presentation of BIM tools used to potentially mitigate delay from vast body of literature. Consequently, this study therefore filled the gap in lack of a conceptual framework for selecting key delay risk drivers for BIM-based construction projects, which has hampered scientific progress toward development of extremely effective delay risk predictive models for BIM-based construction projects. Secondly, this study’s analyses further confirmed a positive effect of BIM on construction project delay risk even though the studies used for its analyses implemented BIM differently on different construction projects across different regions of the world. This is good as the novel BIM is said to have the potential to be regarded as a disruptive technology by the Architecture, Engineering, and Construction industry across the globe which this study’s analyses is further confirming. Also, the BIM mandate required from contractors/consultant by policy makers or government agencies across regions will be realised there by reducing conflicts or law orders between project stakeholders as construction projects achieves timely completion of its projects. This systemic review, as an evidence-based methodology, should essentially be able to assist business policymakers in several areas, such as decision-making. More precisely, the information from this systematic review can enable construction owner(s) to better understand the design intent by ensuring that the building can be designed to achieve the most efficient and best overall performance. Finally, when this conceptual framework for BIM-based construction projects is well-planned and executed during predictive modelling will potentially lead to good project coordination across various phases enhancing project delivery schedule and project management and contribute to post-construction asset and facility management, building automation and control, and many other benefits, including increased property resale values of the building as well as leasing revenues. It's crucial to remember that just mandating BIM on a project does not guarantee success if the owner's project goals aren't clearly defined, and BIM requirements aren't linked to achieving those goals. A project should start by identifying the BIM requirements in order to use BIM successfully. For BIM-based projects, owners should develop processes, standards, and deliverables that can be shared and agreed upon by the owners and the rest of the project team on a regular basis. Also, as the efficiency of BIM is tied very closely to the project delivery method, there is no size or complexity a project should reach to be BIM-based. Each project is unique and will need particular attention from issues, such as a house in a snowy region or a hospital in an earthquake zone. Therefore, it is highly recommended that construction stakeholders consider IPD, which has been shown to be the most effective project delivery method, when using BIM in their projects. Future work may consider trying to identify new drivers that forms the factors that informs construction project delay. More so, future work should conduct comprehensive systematic review on other pertinent issues common to the construction industry.

3.7 Chapter Summary

This chapter summarizes the findings of the systematic review conducted to develop a comprehensive conceptual framework for identifying the most critical delay risk drivers for BIM-based construction projects. Through a rigorous analysis of the existing literature, the study identified contractor-related drivers and external-related drivers as the most crucial categories to consider when developing delay risk predictive models specific to BIM-based projects. The chapter presents the proposed conceptual framework, which offers a structured and evidence-based approach to selecting the most relevant delay risk drivers. It discusses the significance of this conceptual framework in enhancing the efficiency and effectiveness of delay risk predictive modelling efforts, enabling project stakeholders to focus their resources and efforts on mitigating the most significant factors contributing to delays. Furthermore, the chapter highlights the study's analyses, which confirmed the positive impact of BIM on mitigating construction project delays, reinforcing the potential of this technology to address long-standing challenges within the industry.

CHAPTER 4

4.0 ARTIFICIAL INTELLIGENCE IN THE CONSTRUCTION INDUSTRY: A SYSTEMATIC REVIEW OF THE ENTIRE CONSTRUCTION VALUE CHAIN LIFECYCLE[3]

4.1 Chapter Overview

This chapter presents a systematic review of the application of artificial intelligence (AI) technologies across the construction value chain lifecycle. The review amalgamates findings from 70 rigorously selected studies exploring AI utilization in various the construction industry. A central focus is the identification and categorization of seven primary AI technologies documented in the literature. Also, the chapter explores the three major stages of the construction project lifecycle where these AI technologies have been applied. Furthermore, the chapter examines benefits and challenges reported in the literature, shedding light on the potential for design expansion, facilitation of big data analytics, improved workplace health and safety, increased productivity, and enhanced risk mitigation. Concurrently, it addresses the hurdles faced, such as low accuracy levels due to data scarcity, data transformation issues, and the lack of real-world applicability.

4.2 Background of Study

On a regional, national, and global scale, construction is considered a big sector with strategic importance (Egwim, Alaka, Toriola-Coker, Balogun and Sunmola, 2021a). It's also an industry that's been plagued by a slew of issues for decades, including low production, slim profit margins, waste, and safety concerns. Its projects are extremely complex, and the danger of inefficiency and risk, which eventually contribute to project costs and delays, grows geometrically with the project's scale (P.S.Kulkarni1, 2017; Egwim, Egunjobi, *et al.*, 2021). In the past, to mitigate these issues the construction industry traditionally concentrated on generating operational benefits by employing

[3] This chapter is primarily derived from the following journal articles:

Egwim, C.N. et al. (2023) 'Artificial Intelligence in the Construction Industry: A Systematic Review of the Entire Construction Value Chain Lifecycle', Energies 2024, Vol. 17, Page 182, 17(1), p. 182. doi:10.3390/EN17010182.

technology to streamline processes and procedures, but the data gathered as a result of this digitization trial is often overlooked (Zhang and Li, 2011; Egwim, Alaka, Toriola-Coker, Balogun, Ajayi, *et al.*, 2021). Surprisingly, this industry is still on the edge of digitization, which is said to disrupt existing traditional procedures while also opening up a slew of new prospects (Bajpai and Misra, 2020). In recent years, there has been a surge in the global digitization of corporate processes and concepts, such as industry 4.0 and digital twins and digital technology development is growing at such a quick pace that the construction industry is struggling to catch up with latest developments. A formidable digital technology, Artificial Intelligence (AI) is now a vital component of the digital shift (partly due to big data revolution), having been widely adopted across different industries such as healthcare: aiding in diagnoses of patients using genetic data (Huang *et al.*, 2021; Malik, Khatana and Kaushik, 2021); manufacturing: use in managing workforces, production process and allowing predictive maintenance (Chen *et al.*, 2021); education: virtual lectures (Bajaj and Sharma, 2018; Harmon *et al.*, 2021); finance: fraud detection (Iong-Zong Chen and Lai, 2021; Bao, Hilary and Ke, 2022) and transportation: self-driving autonomous cars (Manoharan, 2019; Ma *et al.*, 2020) among many others. The definition of AI has evolved throughout time, but at its foundation has always been the goal of creating machines that are capable of thinking like humans. It does this by trying to imitate human intelligence through hardware and software solutions (Wang *et al.*, 2019). With more data being generated every second, AI technologies such as Robotics, Machine learning, Language Processing, Speech Recognition, Expert Systems, Computer Vision etc. have aided the scientific community in harnessing the growth of big data (Wu *et al.*, 2022). On these massive datasets, scientists can extract information that human eyes cannot interpret quickly enough using AI. As a result, it's clear that AI can help the construction industry improve decision-making, drive project success, and deliver projects on time and on budget by proactively unlocking new predictive insights from its ever-growing volume of project data, which was previously only archived for future reference. For instance, data collected from smart device, Internet of Things (IoT) sensors, BIM, and other sources can be analysed by AI technologies to find patterns in the performance and usage of current infrastructure assets and determine what sort of infrastructure is needed in the future and how it should be supplied (Egwim, Alaka, Toriola-Coker, Balogun, Ajayi, *et al.*, 2021). Furthermore, the number of incremental stages necessary to bring infrastructure designs to operational status will most likely be reduced by AI. This will save time and money in the production of construction materials as well as the development and maintenance of our infrastructure networks. In this regard vast body of international literature have investigated the use of AI technologies to tackle concerns related to construction projects. For example, machine learning have been applied to mitigate construction project delay risks (Gondia *et al.*, 2020; Yaseen *et al.*, 2020a; Egwim, Alaka, Toriola-Coker, Balogun and Sunmola, 2021a), occupational health and safety issues in construction (Xie *et al.*, 2019; Lee *et al.*, 2020; Rijo George, Nalluri and Anand, 2021), construction and demolition waste generation to name a few (Xiao *et al.*, 2019; Cha *et al.*, 2020a; Cha, Moon and Kim, 2021a). It is the view point of Gamba, Balaguer and Chu (Gambao, Balaguer and Gebhart, 2000; Balaguer *et al.*, 2002; Chu *et al.*, 2013) that robotics can be used to automate assembly of building elements (e.g., masonry walls, steel structures etc.). Also Bruckmann, and Wu (Bruckmann, Reichert, *et al.*, 2018; Bruckmann, Spengler, *et al.*, 2018; Wu *et al.*, 2018) made an important point by adding that a robotic system with a gripper coupled to a frame by cables can be used for bricklaying. Furthermore, an expert system for calculating fault rates in construction fall incidents has been developed (Hadipriono, 1992; Imriyas, 2009; Talat Birgonul *et al.*, 2016) and natural language processing has been applied to extract and exchange information, as well as a variety of downstream applications to aid management and decision-making in smart construction projects (Jallan *et al.*, 2019; Faraji, Rashidi and Perera, 2021; Wu *et al.*, 2022). More recently, some studies have conducted traditional narrative critical/literature review for a specific AI technology in the construction industry (e.g. computer vision by Xu (Xu *et al.*, 2020), natural language processing by Wu (Wu *et al.*, 2022), robotic system by Davila (Davila Delgado *et al.*, 2019) etc) while a few other studies have conducted traditional narrative critical/literature review for adopting generic AI technologies in the construction industry with a specific goal (e.g. Parveen (Parveen, 2018) focused on AI's legal issues and regulatory challenges, Schia (Schia *et al.*, no date) focused on AI's impact on human behaviour, and Abioye (Abioye *et al.*, 2021) focused on AI's present status, opportunities and future challenges). However, no study to the best of our knowledge has conducted a systematic review of AI in the construction industry, hence the first motivation of this study. In conducting a systematic review, independent researcher(s) design a system, based on specific guidelines and the system then makes the decisions to determine the outcome of the research thus producing a research outcome that is explicit, reproducible and without a prior assumptions (Pahlevan Sharif, Mura and Wijesinghe, 2019b). Meanwhile, in a typical traditional narrative literature review, the identification, selection, inclusion, and extraction of research articles solely (all) depends on the judgement of the author(s) in order to support their model, hypothesis and to identify the research gaps. This poses a great concern of subjectivity, repeatability, and reproducibility of results from such research (Pahlevan-Sharif, Mura and Wijesinghe, 2019). Secondly, and as a final rationale of this research, no study to the best of our knowledge has conducted any kind of AI review towards its applicability to the entire lifecycle of construction value chain. To fill the gap, this study aims to present an exhaustive systematic review of artificial intelligence and its application to the full construction value chain lifecycle - from building material manufacturing to design, planning, and construction, as well as facilities management. The review is guided by the following research questions:

- What AI technologies have been documented in the literature so far?
- What are the different stages of the construction project lifecycle in which those AI technologies are applied?
- What are the potential benefits of implementing the identified AI technologies, as well as the current hurdles and gaps in their adoption in the industry?

The rationale for need to consider the full construction value chain lifecycle is firstly due to the fact that the construction process comprises a series of interconnected phases, each influencing the others directly or indirectly. Delays occurring in one phase can propagate downstream, affecting subsequent stages of the project. For instance, delays in the design phase might impede material procurement or construction activities, leading to cascading delays and ultimately impacting project timelines significantly. Secondly, the complexity of interactions between stakeholders, resources, and processes throughout the construction value chain necessitates a comprehensive understanding. By scrutinizing the entirety of the lifecycle, from building material manufacturing through design, planning, construction, and facilities management, researchers and practitioners can discern intricate relationships and dependencies. This holistic perspective is indispensable for accurately identifying potential delay risks and devising effective mitigation strategies. Moreover, a thorough exploration of the construction value chain facilitates the identification of critical intervention points. By pinpointing stages where delays are most likely to occur or have the greatest impact, stakeholders can prioritize resources and interventions accordingly. For instance, AI technologies in material manufacturing processes can optimize production efficiency, enhance supply chain management, and mitigate delays in material procurement, thereby alleviating downstream project delays. Additionally, adopting a holistic approach to delay risk prediction enables researchers and practitioners to harness the power of AI across the entire lifecycle. AI algorithms can analyse vast datasets from various phases of the construction process, uncovering valuable insights and patterns that inform proactive risk management strategies. By leveraging AI for predictive modelling and scenario analysis, stakeholders can anticipate potential delays, optimize workflows, and allocate resources more effectively, thereby enhancing project efficiency and resilience. This research makes a significant contribution to the body of knowledge by addressing the knowledge gap in the field of AI in construction industry, specifically by addressing several imaginable application cases for AI in various stages of the construction project lifecycle and the potential benefits of implementing AI technologies, as well as the present roadblocks and gaps in their industrial adoption. This will immensely help the Architecture, Engineering, and Construction industry and the entire built environment ecosystem in identifying opportunities for technological advancement.

4.3 Methodology

Pragmatism is the philosophical underpinning used in this study. This is because it focuses on practical applied research using several viewpoints to aid in data interpretation such that depending on the research question, either observable occurrences or subjective meanings might give acceptable knowledge (Saunders, Lewis and Thornhill, 2019). This study employed a systematic review methodology. A systematic review, in contrast to a traditional literature review, employs a defined, thorough, repeatable, and auditable approach for assessing and interpreting all available research relevant to a specific research question, topic, or field of interest (Pahlevan Sharif, Mura and Wijesinghe, 2019b). Furthermore, by looking at the overall picture and merging discrete components to synthesise results in an organised fashion, a systematic review can overcome the inadequacies of a traditional narrative literature review, which is commonly used in vast body of literature. To develop its systematic review guideline this study employed The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines. PRISMA is a guideline for conducting systematic reviews and meta-analyses that includes a twenty-seven-item checklist and a four-phase flow diagram. It was created by a consortium of twenty-nine professors in the medical community with the aim of improving the clarity and consistency of systematic reviews. As such, our topic of interest, search strategy, inclusion, exclusion, eligibility criteria, data extraction, and synthesis procedures were all outlined in this guideline which was chosen because of its comprehensiveness, wide acceptance, and applicability in different field of study, despite the fact that it was originally established for the medical and health domain (Pahlevan-Sharif, Mura and Wijesinghe, 2019). For a start, we broke down the review into four steps: articles identification, articles screening, critical assessment, and data extraction and synthesis. During the articles identification process (step 1), a thorough literature search was undertaken to find articles for this study. More precisely, only articles published until 21st of January 2022 in the Scopus electronic database was utilised as the primary source of information for the literature search. This database was chosen over others like ScienceDirect and Web of Science because it is the “largest abstract and citation database of peer-reviewed literature” (Cantú-Ortiz and Fangmeyer, 2017). Furthermore, Scopus indexes practically the whole ScienceDirect database, and Scopus offers a greater choice of journals than Web of Science, as well as quicker citation analysis and coverage of more articles (Falagas *et al.*, 2008). The abstract, title, and keywords of publications in this database was searched using the following search terms: ("artificial intelligence" OR "machine learning" OR "deep learning" OR "reinforcement learning" OR "automation" OR "robotics" OR "expert system" OR "natural language processing" OR "optimisation" AND ("construction industry" OR "building industry" OR "built environment " OR "Architecture Construction and Engineering")) with no date, language, and article type restrictions. This search term is divided into two major parts separated by “AND” operator namely AI technologies and the construction industry. The search terms also contained several interesting synonyms, word variations, and exact phrase searching symbols, such as the usage of double quotation marks in “machine learning”, "building industry" among many others. At the article screening process (step 2), the abstracts of 2716 articles were reviewed to see whether they were related to the research questions and to make sure there were no duplicates. As such, led to the removal of 2,306 items, leaving 410. At the critical assessment (step 3), three inclusion and exclusion criteria were employed. Articles were included first and foremost if their focus was on the application of any artificial intelligence technology in a construction project and excluded otherwise. Secondly, only articles with first-hand research data were considered, whereas review articles were not. Finally, each article's relevance was determined using a previously developed rating scale by (Pahlevan Sharif, Mura and Wijesinghe, 2019b; Pahlevan-Sharif, Mura and Wijesinghe, 2019; Gharbia *et al.*, 2020). The scale was adapted based on practical results of how artificial intelligence technology is used in a construction project, with "1" indicating low relevance, "2" medium relevance, and "3" indicating high relevance. Consequently, the full text of all articles having information relating to genuine case studies of AI technology implementation in construction projects or AI technology application proven in a laboratory environment were extracted, exported to a Comma-Separated Values (CSV) file, given a "3" rating and included in the evaluation during the data extraction and synthesis process (step 4). Finally, all articles used were thoroughly examined for data extraction (research aim, project type, country/region, research method(s), AI technology, etc.) and coded as summarized in Fig 4.1. This figure displays the flow diagram of the research articles selection process. It details the total number of articles identified through the database search, total number of articles screened based on eligibility criteria and the total number of these articles fully accessed and finally the total number of these articles that were used for analysis in this study.

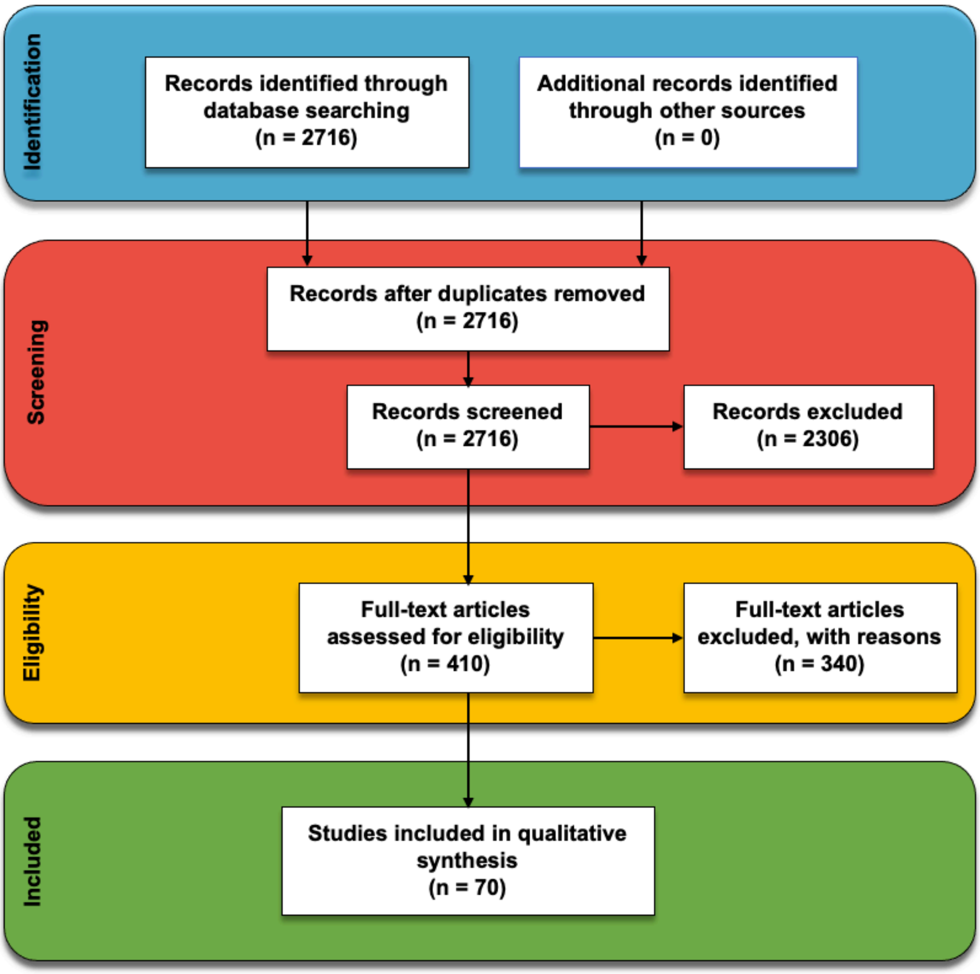

Fig. 4.1: Flow diagram of the systematic review process.

4.4 Results

4.4.1 Summary of selected articles

A summary of the selected articles organised by their respective publication source is provided. It was discovered that the journals "Automation in Construction" and "Journal Construction Engineering and Management," as well as conference proceedings from the "International Conference on Computer-Aided Architectural Design Research, CAADRIA," have the most significant number of articles, accounting for 20 percent of the total number of papers selected. In general, 75.71 percent of the papers (53 out of 70) were published in peer-reviewed journals, whereas 16 articles (22.86 %) were presented at conferences, and just one piece was part of a book series. For two decades between 1993 and 2013, there were eight articles, each with a different year of publication (See Fig. 4.2). There is a notable constant increase in the quantity of research publications throughout the AI in construction research community. More specifically, between 2017 and 2021, there was a constant increase in the number of research publications published in the research community, with a total of 56 articles, indicating a rising interest in the application of artificial intelligence technology to the construction industry. The number of publications according to the first author's institute's location is shown in Fig. 4.3. Predominantly, researchers from China (12 articles), the United States of America (USA) (8 papers), Republic of Korea (6 papers), Italy, the United Kingdom of Great Britain (GBR), and Australia (4 papers each) published most of the papers relevant to the research topic. When comparing the geographical distribution of papers related to artificial intelligence research applied in the construction/ execution lifecycle of the construction value chain, we discovered that researchers in Republic of Korea and China appear to lead research in this area (4 articles each), followed by researchers in GBR and Australia (3 articles each). Interestingly, researchers in China tend to have devoted the most attention to AI applications in the supply/facility management lifecycle of the construction value chain, with seven publications dedicated to it, and five articles dedicated to the planning/design lifecycle of the construction value chain as well.

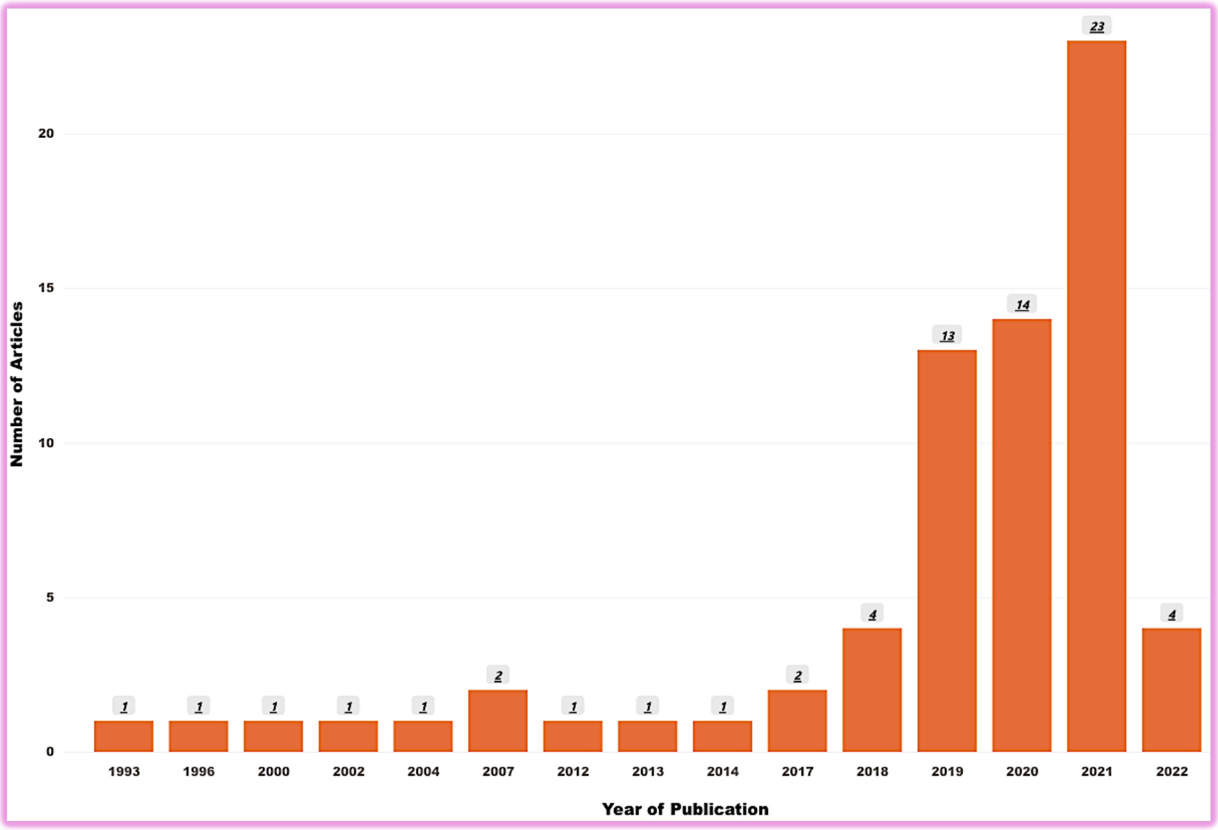

Fig. 4.2: Sequential distribution of articles (total number of articles is 70).

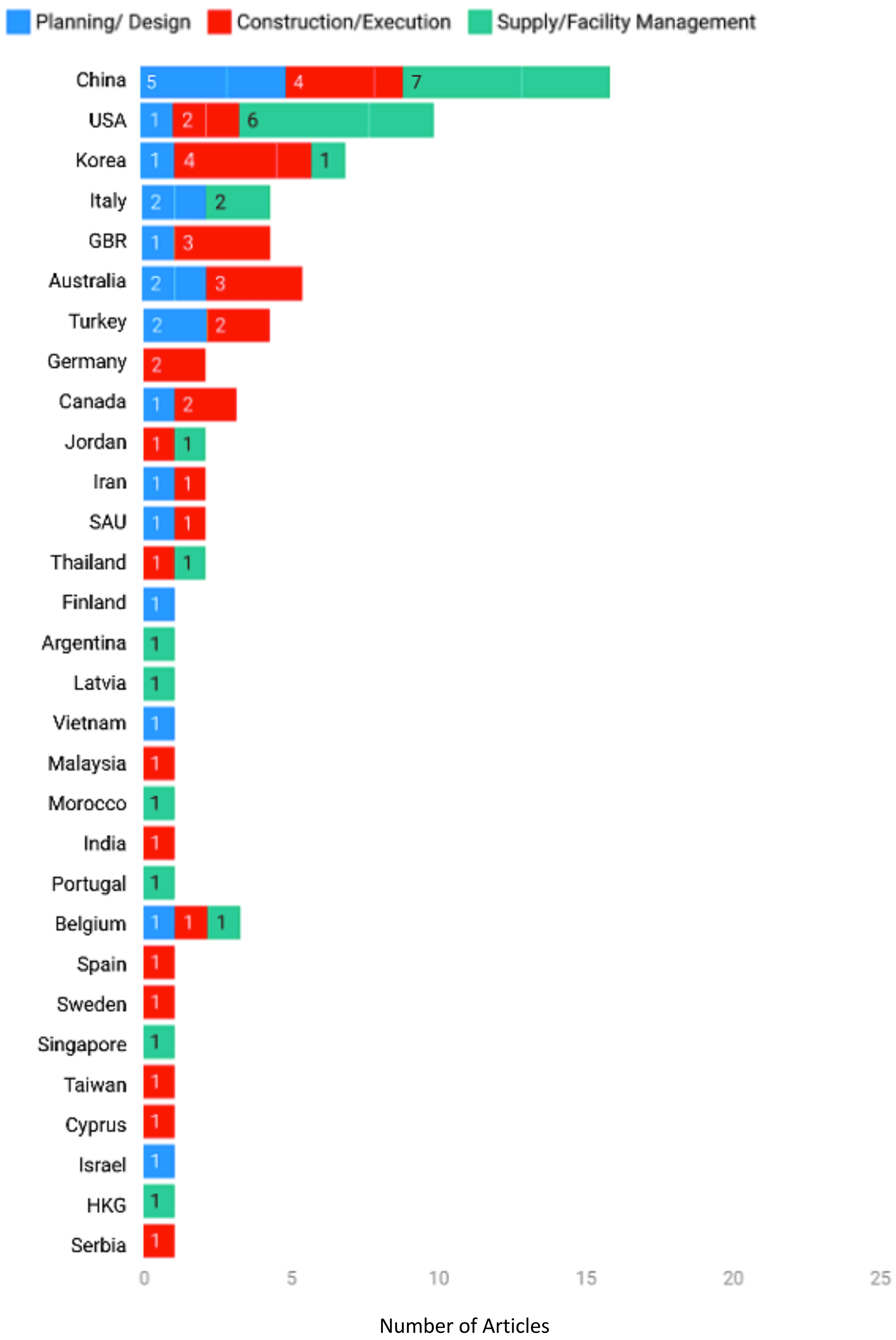

Fig. 4.3: Geographical distribution of articles based on project life cycle.

4.4.2 Types of AI technologies and categorization

Table 4.1 shows the different types of AI technology and their distribution across different application areas. In general, the seventy reviewed studies referenced seven AI technologies for use in the construction industry, with supervised learning, deep learning, knowledge-based systems, robotics, and natural language processing being the most often mentioned. On the other hand, AI technologies such as optimisation and reinforcement learning, garnered less attention. In terms of the distribution of AI technologies to their application areas in health and safety management, supervised learning was the most researched (4 articles), followed by deep learning and knowledge-based systems (3 articles each). Deep learning and supervised learning have shown to be effective time and cost management technologies (2 articles each). Also, robotics was the most often mentioned technology for prefabrication (2 articles). Furthermore, the most promising technologies for heating, ventilation, and air conditioning (HVAC) optimum control were identified to be optimisation and deep learning technologies (2 articles each) while most papers highlighted the application of natural language processing in relation to sustainable concrete and regenerative sustainability (1 article). Quite notably (see Table 4.2), the articles included in this study highlighted the emergence of diverse subtypes within each AI technology. For instance, the most cited subtypes of supervised learning AI technology were support vector machine and artificial neural network (6 articles each), followed by the connected neural network subtype of deep learning (6 articles), and the expert system subtype of knowledge-based systems (5 articles). Also as shown in Table 4.2, most of the papers included cited adaptive manufacturing (4 articles) and Q-learning (4 articles) as the most widely used subtype of robotics and reinforcement learning AI technology respectively. In addition, genetic algorithm appeared to be the most favourable subtype of optimisation technique for the researchers (2 articles).

Table 4.1: Types of AI technology and their area of application

AI Technology	Description	Subtype	Application Area	Reference
Supervised Learning	A type of machine learning in which a computer algorithm is trained on labelled input data for a certain output.	Support Vector Machine (6), Artificial Neural Network (6), Backpropagation(3), Decision Tree (5), Random Forest (4), k-Nearest Neighbors (3), Gradient Boost Machine (2), Adaptive Boosting (2), Naïve Bayes (2), extreme Gradient Boosting (2), Logistic Regression (2), Ensemble Method (2), Light Gradient Boosting Machine (1), Extra Trees (1),	Health & Safety management (4), Time & cost management (2), Building structures (2), structural reliability (1), Sustainable concrete (1), demolition waste management (1), Constructability analysis (1), Construction monitoring (1), Construction equipment (1), Occupant behaviour (1), Site layout (1), Cementitious composite (1), Energy savings & demand response (1), Project Selection (1), Construction negotiation & conflict resolution (1)	(Zhang, Liu and Coble, 2002; Barai and Nair, 2004; Vahdani *et al.*, 2014; Pereira, Ramos and Simões, 2019; Ayhan and Tokdemir, 2019; Bagheri, Nazari and Sanjayan, 2019; F. Zhang *et al.*, 2019; Gondia *et al.*, 2019; Lee, Scarpiniti and Uncini, 2020; Pham *et al.*, 2020; You and Feng, 2020; Cha *et al.*, 2020b; Amin *et al.*, 2021; Koc, Ekmekcioğlu and Gurgun, 2021; Mahjoubi *et al.*, 2021; Milošević, Kovačević and Petronijević, 2021; Sanni-Anibire, Zin and Olatunji, 2021; Shehadeh *et al.*, 2021; Varouqa, 2021; Wang and Li, 2021; Zhang, 2021; Zhou and Chang, 2021; Ayhan, Dikmen and Birgonul, 2021; Cha, Moon and Kim, 2021b; Amini Toosi *et al.*, 2022; Xue and Yao, 2022)

AI Technology	Description	Subtype	Application Area	Reference
Deep Learning	A type of machine learning that trains computers to accomplish things that humans do instinctively.	Convolutional Neural Network (6), Deep Neural Network (4), Autoencoder (1), Long Short-Term Memory (1)	Health & Safety management (3), Time & cost management (2), HVAC optimal control (1), Construction monitoring (1), Intelligent building design (1), 3D datasets (1), Building information modelling (1), monument recognition (1), surface defect detection (1), building recognition (1), Filing architectural drawings (1) Parametric design (1)	(Šatrevičs, no date; Charoenkwan and Homkong, 2017; Lomio *et al.*, 2018; Yu *et al.*, 2018; Ayadi *et al.*, 2019; Mei Yee *et al.*, 2019; Palma, 2019; Fisher-Gewirtzman and Polak, 2019; Ajayi *et al.*, 2020; Lee, Scarpiniti and Uncini, 2020; Norrdine and Motzko, 2020; Bassier and Vergauwen, 2020; Hu *et al.*, 2020; Keshavarzi *et al.*, 2020; Spallone and Palma, 2021; Su *et al.*, 2021; Davila Delgado and Oyedele, 2021; Elsheikh *et al.*, 2021; Kim *et al.*, 2021)
Knowledge-Based System	A computerised system designed to capture and imitate human intellect in symbolic form, often through a series of if–then rules.	Expert system (5), Case-based reasoning (2), Fuzzy logic (1)	Health & Safety management (3), Building automation (1), Productivity estimation (1), Site layout (1), Building diagnosis & repairs (1), Performance evaluation (1), Construction negotiation & conflict	(Koo and Tiong, 1993; Li, 1996; Zhang, Liu and Coble, 2002; Ko, Cheng and Wu, 2007; Sierra *et al.*, 2007; Muqeem *et al.*, 2012; Amiri, Ardeshir and Fazel Zarandi, 2017; Ayhan and Tokdemir, 2019; Kruachottikul *et al.*, 2021; Yangxuan

AI Technology	Description	Subtype	Application Area	Reference
			resolution (1), Occupant behaviour (1) Architectural Innovation (1)	and Zhaoqianjing, 2021; Zhang and Yuen, 2022)
Robotics	A technology that deals with the creation, design, construction, and operation of programmable machines.	Additive Manufacturing (4), Robotic beam assembly (1), Soft robotics (1), Robotic bricklaying (1), Mobile robot (1)	Robotic prefabrication (2), Digital fabrication (1), Collaborative Robotics (1), Block assembly (1), Intelligent Hoisting (1), Health & Safety management (1), Environmental impact analysis (1)	(Gambao, Balaguer and Gebhart, 2000; Jung, Chu and Hong, 2013; García de Soto *et al.*, 2018; Krieg and Lang, 2019; Firth *et al.*, 2020; Hu *et al.*, 2020; Li, Luo and Skitmore, 2020; Wagner *et al.*, 2020; Kontovourkis and Konatzii, 2021)
Natural Language Processing	An artificial intelligence technology that utilises computers to comprehend, generate, and analyse human languages is known as natural language processing.	Text clustering (1), Word segmentation (1), Information Retrieval and Extraction (1), Text analysis (1)	Health & Safety management (2), HVAC optimal control (1), Sustainable concrete (1), Regenerative sustainability (1)	(Sonetti, Naboni and Brown, 2018; Zhang *et al.*, 2018; F. Zhang *et al.*, 2019; Kim and Chi, 2019; Lee, Yi and Son, 2019; Hong *et al.*, 2021)
Optimisation	A technique that seeks to alter an existing process to enhance the occurrence of good results and decrease the occurrence of bad outcomes	Genetic Algorithm (2), Grey wolf optimization algorithm (1), Genetic Algorithm, Stochastic Gradient Descent (1), Genetic programming (1)	HVAC optimal control (2), Health & Safety management (2), Sustainable concrete (1), Building structures (1)	(Ayadi *et al.*, 2019; Bagheri, Nazari and Sanjayan, 2019; Z. Zhang *et al.*, 2019; Koc, Ekmekcioğlu and Gurgun, 2021)

AI Technology	Description	Subtype	Application Area	Reference
Reinforcement Learning	A type of machine learning that allows an agent to learn through trial and error while receiving feedback from its actions.	Q-learning (4)	Energy savings and demand response (1), HVAC optimal control (1), Health & safety management (1), Look-Ahead Schedule (1)	(Vázquez-Canteli *et al.*, 2019; Z. Zhang *et al.*, 2019; Aghalari *et al.*, 2021; Soman and Molina-Solana, 2022)

4.4.3 Construction project types and their lifecycle application area

In terms of the types of construction projects in which AI technologies were used, Fig. 4.4 suggests that a majority of the scholarly articles included (58.60%) were related to built environment and residential building (28 articles in built environment and 13 articles in residential building). Following that were papers regarding high-rise and commercial buildings, which made for 18.60 percent of the articles chosen (7 articles in high-rise building and 6 articles in commercial building). Bridge/ highway project and office construction project were discovered in 7.10 percent and 4.30 percent of the total of 70 articles, respectively. Moreover, power plant, timber construction, and architectural heritage projects all had the same number of articles (2 each), with only one article (1.40%) suggesting the use of AI technology in retrofit building and water treatment plants. Furthermore, Fig. 4.5 depicts the distribution of articles by construction project type and life cycle stage. The majority of articles on built environment were focused on the construction/execution stage (15 articles). However, there were only a few articles that focused on the planning/design stage of the built environment (3 article). Likewise, the planning/design and supply/facility management stages of the construction lifecycle for residential buildings received the greatest attention (6 articles each), with a smaller number of articles devoted to the construction/execution stage (1 articles). Besides, papers related to high-rise buildings had a focus towards the construction/execution stage (3 articles). All the included papers on commercial buildings (6 articles) and bridge/roadway projects (5 articles) focused on the three stages of the construction life cycle in near equal measure. Interestingly, articles pertaining to power plant projects (2 articles) and retrofit building (1 article) exhibit a distinct emphasis solely on construction/execution stage within the construction value chain lifecycle. Conversely, water treatment projects primarily concentrated on the planning/design stage of the construction value chain lifecycle. Additionally, projects involving timber construction, and architectural heritage shared equal number of articles across their respective stages in the construction value chain lifecycle.

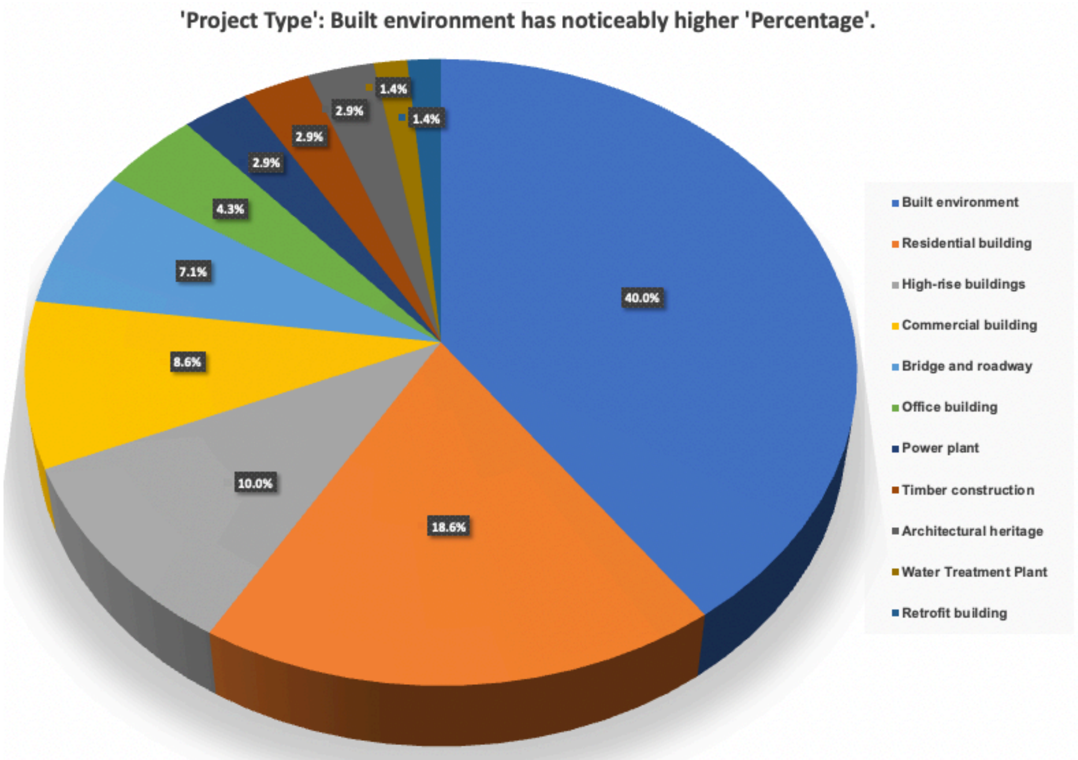

Fig. 4.4: Distribution of articles based on project type.

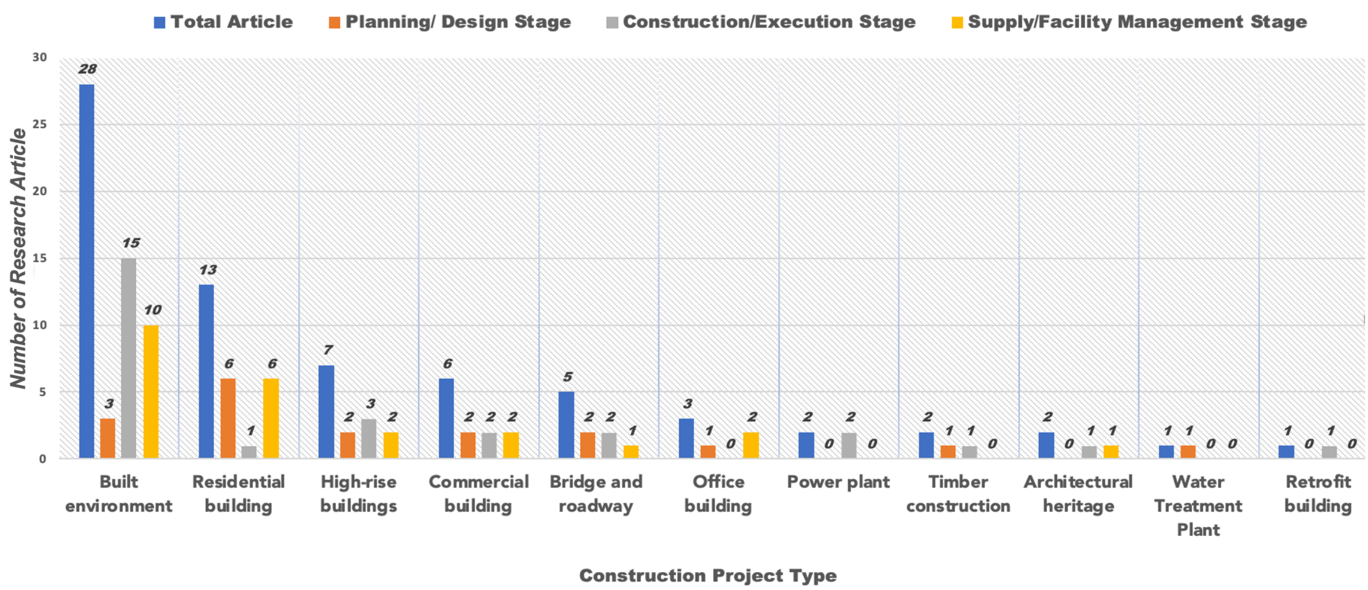

Fig. 4.5: Distribution of articles based on project type and life cycle.

4.4.4 Benefits, challenges, and opportunities for technological advancement

Benefits

This subsection covers a wide range of benefits that can be gained from artificial intelligence technologies' growing popularity in the construction industry. Table 4.3 summarises the primary benefits of implementing AI technology in the construction industry, as described in twenty-six of the seventy selected articles.

Potential for Design Expansion

Intelligent room layouts for better natural ventilation are one example of how AI technologies can lead to novel design aspects. As mentioned by Sonetti (Sonetti, Naboni and Brown, 2018) who developed AI solutions for human-centered regenerative design, AI technologies are strong supporters of human-centric regenerative design when it comes to developing technologies that improve interactions between buildings and their occupants. It is the viewpoint of Lamio (Lomio *et al.*, 2018) that, the application of AI technologies to automate building design process demonstrates that an image taken from a virtual model can accurately distinguish the building type. They developed an AI tool using classical and modern machine learning techniques to categorize images of building designs into three classes. This is especially important considering the large number of BIM structures with missing information or incorrect labelling.

Possibility for Big Data Analytics

AI technologies are exposed to an unending quantity of data to learn from and improve on every day at a time when vast amounts of data are being produced in the industry (Egwim *et al.*, 2022a; Egwim, Alaka, Pan, *et al.*, 2023a). For instance, the research led by Palma' (Palma, 2019) explored the integration of Convolutional Neural Networks (CNNs), a subset of deep learning methods, into the realm of architectural heritage by developing a mobile app aimed at monument recognition, pioneered the use of AI in this domain. Palma' (Palma, 2019) argument is compelling, especially in terms of his pointing out that the output of their adoption of AI technology resulted in the creation of open datasets for testing and evaluating AI applications in the field of architecture and architectural heritage. In addition, Keshavarzi (Keshavarzi *et al.*, 2020) who developed a generative system that addresses the challenge of limited 3D datasets for deep learning methodologies in the built environment stated that their AI technology has the potential to facilitate data augmentation of parametric 3D scan datasets by taking an extant 3D scan as input and generating alternative iterations of the architectural configuration, encompassing walls, doors, and furnishings, accompanied by corresponding textures. This process extends the prevailing 3D geometry datasets, which are conventionally constrained in their scope.

Workplace Health and Safety

AI technology can provide a project with precise job-site safety best practices based on learnt knowledge. As one of the most hazardous industries to work for, this surveillance keeps people safe and accidents to a limit. In the pursuit of minimizing accident occurrences within construction sites, Zhang (F. Zhang *et al.*, 2019) employed a diverse set of AI technologies. Specifically, an ensemble model was devised, integrating text mining, natural language processing (NLP), and machine learning methodologies for the comprehensive analysis of construction accident records. The objective was to discern and extract salient elements associated with accidents, ultimately mitigating potential hazards. With reference to Yu (Yu *et al.*, 2018) AI technologies can be used as non-invasive tool for workload monitoring and thorough ergonomic assessment for various construction tasks, such as assessing risk factors for work-related musculoskeletal disorders by developing an AI tool that employs a smartphone camera with advanced deep learning algorithms to extract construction workers' skeleton data, complemented by smart insoles to quantify plantar pressures during various construction activities. More so, Su (Su *et al.*, 2021) adopted an AI technology to predict the smoke motion and the available safe egress time in a typical atrium.

Increase in Productivity

Some AI technologies can complete repetitive tasks swiftly and precisely while being fatigue-free. For instance, Li (Li, Luo and Skitmore, 2020) detailed the creation of a vision-based intelligent mobile robot hoisting system to improve the hoisting process, including the process of hooks identifying the hoist points and autonomously releasing the components without the need of on-site construction employees. According to García de Soto (García de Soto *et al.*, 2018) by offering a process for evaluating productivity based on total cost and time per unit installed, it is conceivable to get considerable economic advantage from the use of additive digital fabrication to create complicated structures through the development of AI-driven robotic construction method as part of digital fabrication in the construction industry. Furthermore, investigation by several researchers (Jung, Chu and Hong, 2013; Krieg and Lang, 2019; Firth *et al.*, 2020; Hu *et al.*, 2020; Li, Luo and Skitmore, 2020; Wagner *et al.*, 2020) have shown the possibility of replacing risky and difficult manual construction work with automated robots.

Enhanced Risk Mitigation

All construction project has a few risks, which can take numerous forms, including quality, timeliness, and cost. A particular strength of Hong' (Hong *et al.*, 2021) argument is that AI technologies can assist in assigning time and cost contingency to completing a construction project through the development natural language-related AI technologies including clustering methods, including latent semantic analysis (LSA), latent Dirichlet allocation (LDA), and word2vec, amongst many others for quantitative analysis in construction scheduling. Varouqa (Varouqa, 2021) concurred and went on to say that AI technologies can be employed as optimization strategies in prefabricated construction projects to save time and money. Furthermore, Lee (Lee, Yi and Son, 2019) adopted AI technology to perform a preemptive contract-risk evaluation, which can offer stakeholders with contractual positions and rights based on contract facts, minimizing the number of claims and conflict cases between participating parties during construction.

Challenges

There several challenges described in the seventy selected articles about the implementation of AI technology in the construction industry (see Table 4.2). In general, low accuracy level due to scarcity of available data was found to be the most frequently cited challenge (41.43 percent) during the adoption of supervised learning AI technology (15 occurrences), followed by data transformation techniques not transferable to data from other regions (12.86 percent) during the implementation of the same AI technology (5 occurrences), lack of real-world applicability (11.43 percent) when using deep reinforcement learning AI technology (3 occurrences), and incorrect image classification of structures (4.29 percent) during deep learning AI technology adoption (2 occurrences). However, 2.86 percent of the articles considered the combination of industrial robot size and weight limits and high demand for sophisticated algorithms and computing power owing to massive volumes of data to be equally troublesome when adopting deep learning and supervised learning AI technology in the construction industry. Other notable challenges include difficulty in model calibration and excessive modelling errors for heating demand prediction, long installation time for robots, difficulties in developing inference rules for expert systems, misclustering of some project milestones into building works for natural language processing among many others.

4.4.5 Practical implications of AI applications for the construction industry

The integration of Artificial Intelligence (AI) into the construction industry heralds significant advancements in various operational domains, promising enhanced efficiency, safety, and productivity. The practical implications of AI in construction can be leveraged by practitioners in the following ways. Foremostly, AI technologies facilitate the generation of multiple design alternatives based on pre-existing data, thereby augmenting the design process. Engineers and architects can input design objectives and parameters into AI systems, which then explore all possible permutations, providing innovative design solutions that might not be otherwise considered. This capability not only boosts creativity but also ensures that designs meet predefined requirements more effectively. Secondly, AI's ability to analyse and predict potential safety hazards significantly improves workplace safety. For instance, the integration of natural language processing (NLP) and machine learning can help analyse construction accident records, identifying and mitigating potential risks. Moreover, AI tools employing deep learning algorithms can monitor workers' ergonomics, helping to prevent musculoskeletal disorders by analyzing posture and pressure points during various construction tasks. Moreso, automating repetitive and labour-intensive tasks through AI technologies such as intelligent robotics can significantly increase productivity. AI-driven robotic systems can handle tasks like hoisting materials or precision-based construction activities without fatigue, reducing the need for manual labour and minimizing human error. This leads to faster project completion times and reduced costs. Furthermore, the construction industry generates vast amounts of data that, when analysed using AI, can provide actionable insights. AI technologies, such as convolutional neural networks (CNNs), can process large datasets to optimize various aspects of construction projects, from material selection to project scheduling. This data-driven approach helps in making informed decisions that enhance overall project efficiency. Additionally, AI applications often face challenges due to the scarcity and imbalance of labelled training data. Data augmentation techniques, such as those employing undercomplete sparse deep and variational autoencoders, can generate synthetic data to overcome these challenges. This improves the robustness and accuracy of AI models, enabling their application even in data-constrained environments.

4.4.6 Comparative analysis of AI applications across different regions and construction project types

The adoption and effectiveness of AI technologies in the construction industry vary significantly across different regions and types of construction projects. In examining the global landscape of AI adoption in the construction sector, it becomes evident that certain regions are at the forefront of integrating these technologies into their practices. Notably, China and the Republic of Korea are leading the way, particularly in the construction/execution and planning stages of the construction lifecycle. This leadership is not coincidental but rather a result of substantial funding and robust governmental support, which have spurred a significant volume of both academic publications and practical implementations of AI technologies. These regions are setting the pace in exploring how AI can streamline and enhance construction processes from inception to completion. Meanwhile, in Europe and North America, there is also a notable momentum in the adoption of AI, albeit with a slightly different focus. Here, the emphasis tends to be on enhancing existing construction practices through AI-driven safety measures and productivity tools. European projects, for instance, often prioritize sustainable construction practices, leveraging AI to optimize energy use and ensure material sustainability. This reflects a broader regional commitment to sustainability and efficiency, which AI technologies are well-positioned to support. Also, the versatility of AI technologies is further illustrated when examining their applications across different types of construction projects. In the built environment and residential buildings, AI applications are predominantly concentrated on stages such as design, planning, and facility management. These stages benefit greatly from AI's capabilities in design automation, safety management, and energy optimization, leading to more efficient and effective project outcomes. For instance, AI can generate multiple design alternatives rapidly, allowing architects and engineers to explore a wider range of options and optimize building performance from the outset. In the context of high-rise and commercial buildings, AI technologies are extensively used during the construction/execution stage. These technologies play a crucial role in project management, safety monitoring, and optimizing construction processes. AI-driven systems can analyze vast amounts of data in real-time, providing insights that help manage complex construction activities more efficiently. This leads to improved safety standards and streamlined project timelines, which are critical in the fast-paced environment of high-rise and commercial construction. Specialized construction projects, such as power plants, timber construction, and architectural heritage conservation, also present unique opportunities for AI application. In architectural heritage projects, for example, AI focuses on preserving and managing historical data, ensuring that restoration efforts maintain the integrity of historical sites. In timber construction, AI can optimize material usage and ensure structural integrity, addressing specific challenges associated with this type of building material. These tailored applications demonstrate AI's adaptability and potential to address a wide array of construction needs.

Table 4.2. Cited challenges for integrating AI technologies and opportunities for technological advancement in the construction industry.

Challenges	Opportunities	AI Technology	No. of Articles	Reference	%
Low accuracy level due to scarcity of available data	Data augmentation	Robotics, DL, Expert System, Optimisation, SL (15), NLP	29	(Hu *et al.*, 2020), (Muqeem *et al.*, 2012), (Koc, Ekmekcioğlu and Gurgun, 2021), [67], (Lee, Scarpiniti and Uncini, 2020), (Shehadeh *et al.*, 2021), (Bassier and Vergauwen, 2020), (Zhang, 2021), (Sanni-Anibire, Zin and Olatunji, 2021), (Vahdani *et al.*, 2014), (Varouqa, 2021), (Ajayi *et al.*, 2020), (Kim *et al.*, 2021), (Yu *et al.*, 2018), (Kontovourkis and Konatzii, 2021), (Yangxuan and Zhaoqianjing, 2021), (Amin *et al.*, 2021), (Charoenkwan and Homkong, 2017), 99], (Kruachottikul *et al.*, 2021), (Amini Toosi *et al.*, 2022), (Milošević, Kovačević and Petronijević, 2021), (Cha, Moon and Kim, 2021b), (Ayhan, Dikmen and Birgonul, 2021), (Pham *et al.*, 2020), (Šatrevičs, 2006), (Ayhan and	41.43

Challenges	Opportunities	AI Technology	No. of Articles	Reference	%
				Tokdemir, 2019), (Lee, Yi and Son, 2019), (Ko, Cheng and Wu, 2007)	
Data transformation not universally applicable	Model generalisability and Transferability	NLP,SL,Optimisation, Expert System, DL	9	(Zhang *et al.*, 2018), (Pereira, Ramos and Simões, 2019), (Mahjoubi *et al.*, 2021), (Amiri, Ardeshir and Fazel Zarandi, 2017), (Li, 1996), (Xue and Yao, 2022), (Gondia *et al.*, 2019), (Zhang, Liu and Coble, 2002), (Davila Delgado and Oyedele, 2021),	12.86
Only experimental with lack of real-world applicability	Real-world applicability	SL, Optimisation, DL, RL	8	(Wang and Li, 2021), (Ayadi *et al.*, 2019), (Bagheri, Nazari and Sanjayan, 2019), (Spallone and Palma, 2021), (Fisher-Gewirtzman and Polak, 2019), (Aghalari *et al.*, 2021), (Su *et al.*, 2021), (Soman and Molina-Solana, 2022)	11.43
Incorrect image classification of structures	Computer vision and AR	DL, SL	3	(Palma, 2019), (Mei Yee *et al.*, 2019), (Norrdine and Motzko, 2020),	4.29
Size and weight restrictions of industrial robots	Design optimisation	Robot	2	(Krieg and Lang, 2019), (Jung, Chu and Hong, 2013),	2.86

Challenges	Opportunities	AI Technology	No. of Articles	Reference	%
High data demands, require advanced algorithms.	Cloud computing infrastructure	DL, SL	2	(Elsheikh *et al.*, 2021), (You and Feng, 2020)	2.86
Limited Computing Speed	Compiling Keras /TensorFlow	RL	1	(Vázquez-Canteli *et al.*, 2019)	1.43
Model calibration challenges, errors in predictions.	Multi-objective RL	RL, Optimisation	1	(Z. Zhang *et al.*, 2019)	1.43
Long installation time	Programming safety and recalibration of peripherals	Robotics	1	(Wagner *et al.*, 2020)	1.43
Saturated neural network accuracy, optimization challenges.	Enhanced random neural network structure generation.	SL, DL	1	(Lomio *et al.*, 2018)	1.43
Difficulties in developing inference rules	High Optimisation techniques	Expert System	1	(Sierra *et al.*, 2007)	1.43
Misclustering of some project milestones into building works	Construction schedules analysis	NLP	1	(Hong *et al.*, 2021)	1.43
Inadequate network architecture	Large (big) datasets	SL, Optimisation	1	(Barai and Nair, 2004)	1.43
Omission of planning in robotic design.	Design optimisation	Robot	1	(García de Soto *et al.*, 2018)	1.43
Limited scope in classification of building structures	Computer vision and AR	SL	1	(Zhou and Chang, 2021)	1.43

Challenges	Opportunities	AI Technology	No. of Articles	Reference	%
Threshold limitations on Elasticsearch queried data	Load balancing	NLP	1	(Kim and Chi, 2019)	1.43
Limited customization of textures for walls	3D modelling and AR	DL	1	(Keshavarzi *et al.*, 2020)	1.43
Control approach for large hydraulic robots	New control methods	Robot	1	(Gambao, Balaguer and Gebhart, 2000)	1.43
Robot response hindered by image quality	Hgh-definition camera	Robot	1	(Li, Luo and Skitmore, 2020)	1.43
Natural language exhibits diverse expressions	Advanced optimisation	NLP, SL	1	(F. Zhang *et al.*, 2019)	1.43
Inability to repair walls and columns	Data-driven machine learning	Expert system	1	(Koo and Tiong, 1993)	1.43
Heterogeneous hardware and software integration	AI knowledge experts	SL	1	(You and Feng, 2020)	1.43
Social barriers to the adoption of AI	AI education and trainings	NLP, SL, DL	1	(Sonetti, Naboni and Brown, 2018)	1.43

Opportunities for technological advancement

More-so, Table 4.2 emphasises how AI technologies opens a slew of opportunities for technological advancement in the construction industry, giving it a competitive edge by improving efficiency across the whole value chain from building materials manufacturing to design/planning, construction/execution, and supply/facility management.

Data Augmentation

In concrete, it shows that 41.43% of the selected papers who experienced low accuracy level due to scarcity of available data suggested the need for future studies to augment datasets for the development of more robust AI technologies in the industry. For instance, Hu's (Hu *et al.*, 2020) automatic robotic disinfection framework was unable to investigate the relationship between adequate UV light exposure and the effects of pathogen eradication due to low accuracy in segmenting the areas of potential contamination on small objects such as doorknobs and cabinet handles in adverse conditions. Furthermore, Davila Delgado (Davila Delgado and Oyedele, 2021) successfully demonstrated the application of undercomplete, sparse, deep and variational autoencoders as novel techniques for data augmentation and generation of synthetic data in construction management which can provide useful insights regarding the underlying non-linear relationships among variables in the datasets amongst many other selected studies.

Model Generalisability/ Transferability

The opportunity for AI model generalizability and transferability became eminent as the unique data transformation employed in 12.86% of the selected articles are not transferable to the data from other regions as typical of any data driven model. Zhang's (Z. Zhang *et al.*, 2019) argument is persuasive in this aspect, particularly when they pointed out that their building energy AI model has better generalizability because it is based on fundamental scientific laws. A building energy model, for example, can properly estimate the energy performance of a new unseen control technique while a data-driven model may not. This is the case since the data-driven model is built using a training dataset that contains no information about the unseen control technique. In addition, Koc (Koc, Ekmekcioğlu and Gurgun, 2021) raised awareness for future studies to take advantage of model generalizability, since he encountered the difficulty of an unbalanced dataset while using AI technology to assess construction workers' post-accident impairment status.

Real-world Applicability

About 11.43 percent of the selected articles only tend to have simulated the adoption of AI technologies in a controlled environment thus lacking the confidence to validate their methods in a real-world setting. According to Vázquez-Canteli (Vázquez-Canteli *et al.*, 2019), a useful graphical user interface (GUI) that allows users to write machine learning code or set hyper-parameters of the algorithms after the simulation environment is compiled is required for their fast AI-based building energy simulator implemented in an integrated simulation environment to be tested in a physical setting. Furthermore, as illustrated in Hong's (Hong *et al.*, 2021) AI framework for clustering construction schedules in UK based construction projects, Gondia's (Gondia *et al.*, 2019) AI-based model lacks the use of real construction project schedule information for their construction project delay risk prediction implementation in Egypt.

Computer Vision and Augmented Reality

The application of the most recent computer vision techniques and augmented reality functions followed suit, with 4.29 percent of selected articles recommending them as a means of advancing technology in the construction industry. More precisely, In the implementation of automatic image recognition of architectural heritage sites by Palma (Palma, 2019), for example, items in the same cultural site appeared to be extremely similar, or could appear together in the same view, making it impossible to distinguish one part from another. However, these circumstances emphasise the practical value of AI rather than the recognition of landmarks that are far apart. Thus, implies the most up-to-date computer vision algorithms will be key to obtain more detailed information than previously or the use of augmented reality functions to enhance their interaction. Other notable opportunities for technological advancement that received less attention from the selected papers includes, design optimisation and cloud computing infrastructure with 2.86%, followed by multi-objective reinforcement learning, safety programming and re-calibration of peripheral modules enterprise AI knowledge experts, AI education and trainings among many others (1.43% each)

4.5 Discussion

The systematic review results show foremostly the distribution of research articles according to their publication source has received wider coverage of this topic in academic journals which is consistent with the work of (Abioye *et al.*, 2021),(Debrah, Chan and Darko, 2022) that the adoption of artificial intelligence in the construction industry has taken a quantum leap relatively due to the availability of funding in that area. There is also, a notable constant increase in the number of academic research publications that are based on the application of AI technology in the construction industry within the entire research community ecosystem. This arguably shows a significant improvement in the promotion of research and development of trustworthy AI solutions by funding bodies and agencies across the globe. Interestingly this is in line with the findings of Rahkovsky (Rahkovsky *et al.*, 2021) who argued that artificial intelligence research clusters are experiencing extreme growth due to great support from research funding organizations that is currently been led by the National Natural Science Foundation of China (NNSFC). It is no surprise from the results of this systematic review that researchers from China are dominating the research space of the application of AI technologies in the construction industry since NNSFC is the largest funder of AI technology research over other large funding bodies such as National Institutes of Health and National Science Foundation from the USA, European

Commission and European Research Council from Europe, and Japan Society for the Promotion of Science from Japan among others. Secondly the analysis of this study further provided answers to all the research questions stated in the first section. More specifically, although seven major AI technology types were found in the literature, supervised learning emerged as the most influential AI technology of choice for most researchers especially towards its applicability in health and safety management. Supervised learning is branch of machine learning in which computer algorithms are trained on labelled input dataset for a certain output. From a labelled training dataset (i.e., a dataset that already has a known value for each record's output variable) supervised machine learning algorithms can find insights, patterns, and correlations. When proper answers for a given task during training is provided, the machine learning algorithm can learn how the rest of the characteristics relate to the output, allowing you to unlock insights and make predictions based on past data. This is extremely crucial for the industry and consistent with the findings of (Egwim, Alaka, Toriola-Coker, Balogun, Ajayi, *et al.*, 2021) who argued that the industry can derive key benefits from AI to drive further profitability only when it leverages the amount of data produced from backlog of project schedules, as-built drawings and models, computer-aided designs, costs, invoices, among many other sources. Furthermore, the results showed that although AI technologies can be applied in three major stages of construction project lifecycle, more attention are drawn towards the supply/facility management stage (see section 3). It can be argued that this is owing to the massive quantity of data collected over time (from the design stage all the way through) making it ideal for the adoption of the most significant AI technology (supervised learning). Thus, creating a great opportunity for the industry to capitalise by allowing, for example facility managers to take proactive action. For instance, as argued by (Z. Zhang *et al.*, 2019), AI can recognise portions of buildings that aren't being utilised and automatically turn off the heating, ventilation, and air conditioning, substantially decreasing energy use. Recent studies have continued to validate these trends, emphasizing the transformative impact of AI on construction practices. For instance, Johnson et al. (2023) investigated AI's role in predictive maintenance, revealing significant cost savings and operational efficiency in large-scale construction projects. Their findings support the notion that AI can reduce downtime and enhance the longevity of construction equipment through predictive analytics and real-time monitoring. Similarly, research by Smith et al. (2023) has underscored the importance of AI in optimizing construction logistics. By utilizing machine learning algorithms to analyse supply chain data, their study demonstrated improvements in material procurement processes, reducing delays and costs associated with material shortages and excesses. This aligns with earlier findings on AI's capability to streamline construction processes and enhance project management efficiency. In terms of sustainability, recent work by Wang and Li (2023) has highlighted AI's potential in promoting eco-friendly construction practices. Their research focused on AI-driven models that optimize energy consumption and material usage, leading to a reduction in the carbon footprint of construction projects. This is particularly relevant given the increasing emphasis on sustainable development within the industry. Furthermore, Patel and Desai (2023) have explored the application of AI in enhancing worker safety. Their study introduced AI-based systems for real-time hazard detection and risk assessment on construction sites. The implementation of these systems resulted in a noticeable decrease in workplace accidents, corroborating the findings of previous studies on AI's role in improving construction site safety. Moreover, recent advancements in AI technologies continue to address the socio-technical challenges identified in earlier research. For example, Liu et al. (2024) have developed new data augmentation techniques to mitigate the issue of data scarcity, enhancing the robustness of AI models in construction applications. This innovation is crucial for ensuring that AI systems can operate effectively even with limited training data, a common challenge in the industry. Lastly, a study by Kim et al. (2024) has demonstrated the successful integration of AI with BIM to enhance project visualization and collaboration among stakeholders. Their research showed that AI-enhanced BIM systems could predict potential project delays and budget overruns with high accuracy, enabling more proactive project management and decision-making. Moreover, AI technologies in the construction industry were found to hold promise for applications in many types of construction projects and their respective lifecycle application area with about twenty percent of the literature reporting their implementation in any structures and systems that are part of the built environment (urban area, pedestrian walkways, parks etc.) projects. Kılkış' (Kılkış, 2021) argument is about this is compelling, especially when they mentioned that the built environment impacts all parts of our life, including the buildings we live in, the distribution systems that provided us with water and energy, and the roads, bridges, and transportation systems we use to move about. Additionally, most articles acknowledged that artificial intelligence technologies' growing popularity in the construction industry would offer a wide range of benefits with potential for design expansion as a key benefit according to most of the selected literature. We argue that as engineers and architects spend a lot of time working on the design of buildings as designers and with access to a database of many previously built building plans, an AI technology system can produce design alternatives based on the information it acquires from the designs in the database. As a result, designers can simply enter design objectives and parameters into the system, and the system will be able to investigate all conceivable permutations of a solution, creating design alternatives that fulfil all the previously defined requirements, learning what is a better design option with each iteration, and making it a stronger tool with each new project. Beyond these possible design benefits, generative design has the potential to boost creativity. Take for example, it can enable architects to discover previously unimaginable methods of designing forms and curves or lead them to design solutions that they would not have explored otherwise. However, most of the articles reported that it is challenging to apply AI technologies in the construction industry because of low accuracy level due to scarcity of available data. Data scarcity arises when there is a paucity of labelled training data or when there is none at all. It might be a shortage of data for a particular label as compared to the other labels (known as data imbalance). It was discovered from some of the selected literature (Bassier and Vergauwen, 2020), (Zhang, 2021), (Sanni-Anibire, Zin and Olatunji, 2021), (Vahdani *et al.*, 2014), (Varouqa, 2021) which is also in line with research (Egwim, Alaka, Toriola-Coker, Balogun and Sunmola, 2021a) that mega infrastructure projects often have access to a lot of data, however they may have data imbalances, whereas small sized projects typically have a limited amount of labelled training data. As a result, resolving this issue cannot be overstated, reported articles (41.43%) universally agreed, citing "Data Augmentation" as one of the quickest prospects for technical improvement in this area. For instance, the research (Davila Delgado and Oyedele, 2021) successfully demonstrated the application of undercomplete, sparse, deep and variational autoencoders as novel techniques for data augmentation and generation of synthetic data in construction management which can provide useful insights regarding the underlying non-linear relationships among variables in the datasets amongst many other selected studies. Notwithstanding, practical obstacles beyond just data accuracy persist around cultural readiness, ethical risks, skill shortages and flaws in security posture for many construction projects exploring AI solutions. As the researches (Egwim and Alaka, 2021) and (Egwim *et al.*, 2023) assessed, construction has often lagged significantly in digital transformation and technology assimilation compared to other industries. Coupled with an aging workforce leaning on legacy methods, this exacerbates reluctance and barriers to AI change management. For instance, PwC (PwC, 2021) notes generational shifts may gradually improve receptiveness, like modelling shows younger workers are 67% more open to retraining on AI tools relative to senior staff. But broad culture change inevitably remains long-term. Customized change management programs fitting construction realties are hence vital to align teams behind AI via strategic internal communications campaigns and leadership vision as exemplified by firms such as Bechtel. Additionally, the opaque decision-making of AI systems poses ethical dilemmas around accountability as flagged by Parveen (Parveen, 2018). Lack of explainable outcomes or audit trails can impede transparency and responsible oversight of automated systems. There's also dearth of standardized governance principles as highlighted in Egwim (Egwim, Alaka, Toriola-Coker, Balogun and Sunmola, 2021a)'s delay risk assessment. The specialist expertise needed is another capacity challenge evidenced by widening talent gaps globally per Johnson (Johnson *et al.*, 2021)'s labor market analysis. Most construction firms aren't staffed with multidisciplinary data scientists or algorithm auditors. As such, the shortage of such AI and analytics roles may worsen for small and mid-size construction companies lacking resources to reskill staff or attract experts. Additionally, many construction industry jobs also require on-site client coordination, hence diminishing flexibility that technology candidates expect. Therefore, targeted training programs are crucial to developing well-rounded internal capabilities. Finally, the vulnerability of connected tools or data handling processes to malicious threats leaves unprepared adopters exposed to crippling breaches as studied across industries (Saka *et al.*, 2023, 2024). Also, lack of transparency around data rights or algorithmic decision-making processes also introduces major ethical risks. Furthermore, antiquated security postures coupled with failures to implement robust and resilient protections can negate any assumed productivity gains. Thus, addressing these open socio-technical problems demand coordinated efforts across construction stakeholders to formulate frameworks, standards and cultural shifts guided by construction specific nuances.

4.6 Conclusions

The systematic review study covers 70 studies that were judged to be rigorous, credible, and relevant in their application of AI technology in the construction sector. The research content of these 70 publications demonstrated that artificial intelligence research in the construction sector has taken a quantum leap, with increased interest in academic journals, particularly in the last few years, owing to the availability of funding in that area. Most articles pertinent to the research topic in general were published by Chinese researchers. More precisely, scholars from the Republic of Korea and China contributed the most publications to the construction/ execution lifecycle stage of the construction value chain. Furthermore, China also published most of the related articles concerning AI applications in the planning and facility management lifecycle stages of the construction value chain. Construction AI technology was discovered to be a growing application field, with supervised learning, deep learning, knowledge-based systems, robotics, natural language processing, optimisation, and reinforcement learning AI technologies all appearing to have more potential to influence the development of AI research for increased efficiency and productivity. Regarding the Construction AI technology categories given above, the bulk of the featured publications used the supervised learning approach. A substantial number of the articles were connected to the built environment and residential building in terms of the construction project types in which such AI technologies are used. The papers on high-rise and commercial buildings came after that. A few studies advocated the application of AI in building retrofits and water treatment plants. Most publications on the built environment concentrated on the construction/execution stage. Similarly, the planning and facility management lifecycle stages of the residential building garnered the most attention. According to the findings, the most significant number of studies across all building construction disciplines focused on the possibility for regenerative design expansion. However, there are various obstacles to implementing AI technology in the construction industry, with low accuracy owing to a lack of relevant data being the most mentioned issue. It's also worth noting that, despite being the most prevalent AI construction method, supervised learning has been the technology of choice for the most difficult challenge to be solve in the industry. And as sparse, deep, and variational autoencoder approaches show promise in providing meaningful insights into the underlying non-linear correlations among variables in datasets, data augmentation was identified as one of the most promising areas for technical advancement. This study presents an all-inclusive systematic review of vast body of knowledge on artificial intelligence in the construction industry. The findings of this study present a comprehensive assessment of the many types and categories of AI technologies, as well as their application areas and the advantages of using them at the three lifecycle stages of the construction value chain. This knowledge will assist construction organisations across the world in recognising the efficiency and productivity advantages that AI technologies can provide while helping them make smarter technology investment decisions. It will point construction organisations in the right direction in terms of imagining the construction problems that AI technology could solve. In addition, it is possible to integrate evidence from the sorts of construction projects where AI technologies were used to address technological difficulties and see what new AI technologies can accomplish in the future. Evidently, the findings of this study are based on a systematic review methodology. Given that the research article keywords were domain specific, the principal drawback of this study approach might be bias in publication selection. As a result, it's possible that some important papers were overlooked throughout the search. Additionally, the PRISMA guideline mandated the use of predetermined inclusion and exclusion criteria for article selection, implying that important publications that did not meet these criteria may have been overlooked as well. Furthermore, any breakthroughs in the field of AI technology in construction are pushed by experts who are unable to publish in book series, conference proceedings, or academic journals. Consequently, there's a chance that any important research from the experts or somewhere else were overlooked throughout the search. Although the study explored a variety of AI technologies for various construction projects, further research is needed to figure out how to simplify these complicated systems and processes to establish an integrated AI system for the construction sector. Therefore, an implementation framework is crucial to soften the introduction of a system and bridge the adoption gap by addressing low accuracy due to a scarcity of available data, model generalisability, incorrect image classification of structures, high requirements for sophisticated algorithms, and limited computing speed across the existing construction value chain.

4.7 Chapter Summary

This chapter presents a systematic review of the application of artificial intelligence (AI) technologies in the construction industry, synthesizing findings from a rigorous examination of seventy relevant research articles. The review follows the PRISMA guidelines, ensuring a structured and reproducible methodology. The analysis reveals a surge in research publications related to AI in construction, particularly in recent years, reflecting the growing interest and investment in this domain. Chinese researchers emerge as prominent contributors, alongside scholars from the Republic of Korea and the United States, indicating the global significance of this field. Seven primary AI technologies are identified and categorized: supervised learning, deep learning, knowledge-based systems, robotics, natural language processing, optimization, and reinforcement learning. The review explores into the specific subtypes and applications of each technology within the construction context, providing a detailed understanding of the technological landscape. The chapter explores the three major stages of the construction project lifecycle where AI technologies have been implemented: planning/design, construction/execution, and supply/facility management. Notably, the review synthesizes the potential benefits of AI adoption in construction, including the potential for design expansion, facilitation of big data analytics, improved workplace health and safety, increased productivity, and enhanced risk mitigation. Concurrently, the chapter critically examines the challenges encountered, such as low accuracy due to data scarcity, data transformation issues, and the lack of real-world applicability in certain studies. Recognizing the importance of technological advancement, the chapter identifies opportunities for further progress, including data augmentation techniques, model generalizability and transferability, real-world application testing, and the integration of advanced computer vision and augmented reality technologies.

CHAPTER 5

5.0 RESEARCH METHODOLOGY[4]

5.1 Chapter Overview

The previous chapter provided a systematic review of the application of AI technologies in the construction industry, synthesizing findings from a rigorous examination of seventy relevant research articles. This chapter now describes the specific methodology that will be used to conduct the research study. The purpose of this chapter is to explain and justify the philosophical assumptions, approaches, designs, strategies, and techniques that will guide how data is collected, analysed, and interpreted. It lays out the overarching research process and methods that will be employed. At the end it explains and rationalizes the methodology adopted to effectively carry out this research project on construction delays.

5.2 Research Philosophy

A research philosophy is a collection of assumptions about how the world under investigation works (Bryman, 2012). It is the fundamental explanation of what knowledge is. Research philosophies might also differ with regard to the objectives of the study and the most effective strategy for achieving these objectives (Goddard and Melville, 2005). These are not always distinct, but the research project's chosen research philosophy is determined by the kind of information being examined (May, 2011). According to Mark & Lyles, (2005) and Rose et al., (2014), knowing research philosophy is important because it helps researchers to clarify their research designs, detect the effects of various research

[4] This chapter is primarily derived from the following journal articles:

Egwim, C.N., Alaka, H., Toriola-Coker, L.O., Balogun, H. and Sunmola, F. (2021) 'Applied artificial intelligence for predicting construction projects delay', Machine Learning with Applications, 6, p. 100166. doi:10.1016/j.mlwa.2021.100166.

Egwim, C.N., Alaka, H., Toriola-Coker, L.O., Balogun, H., Ajayi, S., et al. (2021) 'Extraction of underlying factors causing construction projects delay in Nigeria', Journal of Engineering, Design and Technology, ahead-of-p(ahead-of-print). doi:10.1108/jedt-04-2021-0211.

Egwim, C.N. et al. (2023) 'Artificial Intelligence in the Construction Industry: A Systematic Review of the Entire Construction Value Chain Lifecycle', Energies 2024, Vol. 17, Page 182, 17(1), p. 182. doi:10.3390/EN17010182.

designs, and create new research designs. According to Rouse, (1997), the focus of any research will determine its course. It is impossible to overstate the benefits of comprehending philosophical concerns in research. For instance, it helps in defining the type of evidence required as a component of the study strategy to address the research questions. Also, it aids in choosing the ideal research approach to achieve the goals and aid in the modification of the study design on the limits of many fields of knowledge (Mark and Lyles, 2005). The research paradigm or philosophy is highly important and critical since it comprises the fundamental beliefs about how knowledge develops, which will influence the research methods and research strategies. There is no one philosophical perspective that fits all research problems, and none is superior to the others, according to Saunders et al. (2009). Ontology, epistemology, and axiology make up the three aspects of philosophical beliefs that are crucial to the research process.

5.2.1 Ontology

Essentially, ontology is the study of reality. It explains the nature of reality, what thoughts occur to mind when doing research, and what effects it has on people and the environment (Ritchie and Lewis, 2003). More precisely, ontology examines whether entities should be viewed as originating from the perceptions and acts of social actors or from sources outside of these actors (Crotty, 1998). The distinction between reality and how we see it is made obvious by ontology. Additionally, it teaches us about how people's conduct is affected by nature. Researcher confidence in the nature and presence of the objects they study is increased thanks to ontology (Ritchie, 2003). What "truth claims" about reality may a researcher, for instance, make? Who judges what is "genuine" and legitimate? How do researchers handle contrasting and incompatible views of reality? In addition, Bryman (2012) introduces the idea of "social ontology," which he describes as a philosophical question in research that addresses the nature of social entities, i.e., whether these social entities are or can be objective entities that exist independently from social actors or rather they are social constructions in themselves created from the perceptions, actions, and interpretations of the individuals in society. Likewise, Snape & Spencer (2003) claim that ontology is concerned with the question of "whether or not there is a shared social reality or only multiple, context-specific ones and, closely related to this, whether there is a social reality that exists independently from human conceptions and interpretations. For instance, a real world exists independently of human experience according to realist ontology; this reality may be studied, understood, and experienced as a "truth." Relativist ontology, on the other hand, is founded on the idea that reality is created by the human mind and that there is no such thing as a "true" reality. Instead, reality is "relative" to how people see it in any moment and location". The ontological worldview primarily encompasses two philosophical perspectives. They are subjectivism and objectivism. This study adopts objectivism as its ontological standpoint foremostly because it aims to develop AI technologies for delay risk prediction, which inherently involves analyzing empirical data related to construction projects. Objectivism aligns with this aim by emphasizing the importance of gathering and analyzing data objectively to uncover underlying patterns and relationships. Secondly, as the study objectives emphasize systematic reviews and expert surveys to gather empirical evidence regarding factors affecting construction project delays and the applicability of AI technologies in the industry, objectivism provides a suitable ontological framework for conducting such research activities, as it emphasizes the importance of unbiased observation and systematic analysis of data to generate reliable and generalizable findings. Furthermore, the utilization of gathered data to develop hyperparameter optimized AI predictive models relies on the assumption that there are underlying patterns and relationships within the data that can be objectively identified and utilized for predictive purposes. As such, objectivism supports this assumption by asserting the existence of objective reality that can be measured and analysed to uncover predictive insights.

5.2.2 Epistemology

In general, epistemology refers to the presumptions we have on the type or form of knowledge or how one might learn about the outside world (Richards, 2003; Snape and Spencer, 2003). Epistemology is frequently utilised in scientific study, and this is because it aids in the discovery of knowledge that can be proven beyond a reasonable question; in other words, it looks for generally accepted knowledge and addresses the facts accordingly. As such, one must specify what level of expertise is appropriate in the area of their study and provide details on the findings of in-depth testing. According to Crotty (1998), epistemology is a means of understanding the world and how to look at it. It incorporates knowledge, and thus, it embodies a certain perception of what that knowledge means. The 'nature' of knowledge, its possibility (what knowledge is conceivable and may be pursued and what is not), its extent, and its validity are all topics encompassed by epistemology, according to him. Cohen et al. (2002) asserted that epistemology is concerned with one's presumptions about "the basic roots of knowledge - its nature and form, how it may be gained, and how it is conveyed to other human beings" in order to further clarify what epistemology is all about. The authors also emphasise how our epistemological presumptions about knowing have a significant impact on how we approach learning about social behaviour. They are referring to the choices the researcher will have to make on the type(s) of methods they will employ in their research in light of their epistemological presumptions. In other words, if knowledge is seen as something that is hard, objective, and tangible, then the researcher must take on the position of an observer and adhere to natural scientific procedures like testing, measuring, etc. The researcher is forced to reject the techniques of natural science and become more involved with their topics if knowledge, on the other hand, is seen as being personal, subjective, and unique. The philosophical stances that fall under the epistemological worldview include idealism, and realism. This study adopts realism epistemological standpoint because in construction projects, there are concrete and measurable phenomena, such as project timelines, resource allocation, and environmental factors, which exist independently of individual perceptions. By adopting a realist epistemological standpoint, the study acknowledges the importance of uncovering these objective realities through systematic observation and analysis. Secondly, realism acknowledges the role of causal mechanisms and structures that underlie observable phenomena. Therefore, in the quest to implement delay risk prediction in construction projects, there are underlying causal factors and relationships that contribute to project delays. Realism will obviously help to identify and understand these causal mechanisms, rather than merely describing observable patterns. This approach allows for a deeper understanding of the underlying dynamics driving delay risks, thereby enabling more effective predictive modelling and risk mitigation strategies.

5.2.3 Axiology

In the philosophy subfield of axiology, value judgments are examined (Saunders, Lewis and Thornhill, 2019). The word "axiology" is Greek in origin and means "value" or "worth." Axiology evaluates the impact of the researcher's own value at every stage of the research process (Yan, 2016). It is a relatively new contribution to the field of philosophy study. The term "axiology" generally relates to the research's "aims." This area of research philosophy tries to make it clear if your goal is only to comprehend the world, or whether you're also looking to explain or forecast it (Lee and Lings, 2008). In other words, axiology focuses on what you value in your research. This is significant because it determines what you value in your study and how you perform your research. Axiology teaches us how views and values affect the gathering and evaluation of one’s study data. According to Silverman (2000), it helps one comprehend the influence that public opinion has on the gathering and analysis of research. It aids in your comprehension of how important it is to include others' opinions when conducting research. Thus, the name "axiology" refers to an endeavour to bring together and critically assess a wide range of previously existing and overlapping problems about what constitutes goodness, moral behaviour, worth, and responsibility (Given, 2012). In other words, axiology deals with issues around what is valued and seen as desirable or "good" for people and society. This study adopts a pragmatic axiological standpoint as pragmatism aligns with the practical orientation of the study, which aims to address real-world challenges in the construction industry by developing predictive models that can improve project management and efficiency. By prioritizing practical outcomes and focusing on solutions that work in practice, this pragmatic approach helps to ensure that the research findings are relevant and applicable to industry stakeholders. Additionally, pragmatism allows for flexibility in research methods and approaches, enabling researchers to adapt their strategies based on the specific needs and constraints of the construction industry.

5.3 Research Paradigm

Different research paradigms, including positivism, realism, interpretivism, pragmatism among many others, are derived from these three philosophical elements. We learn about social events and the many meanings that individuals give them through objectivism as shown in Figure 5.1 (Saunders et al., 2009). It distinguishes between the effects of various social occurrences on various persons. Contrary to objectivism, constructivism holds that individuals are the ones who produce social phenomena. According to Creswell (2014), pragmatism paradigm places more emphasis on research issues and questions than procedures. It derives from results and actions rather than antecedents' circumstances. As a result, the philosophy applies to the mixed method where the researcher has the freedom to make their own decisions and does not follow a particular philosophy or reality. It should "focus on practical applied research, combining multiple viewpoints to assist interpret the results," according to Saunders et al. (2009). Theories are employed by pragmatism to find a solution to a particular problem. It is an alternative to others and is relatively new in comparison to others.

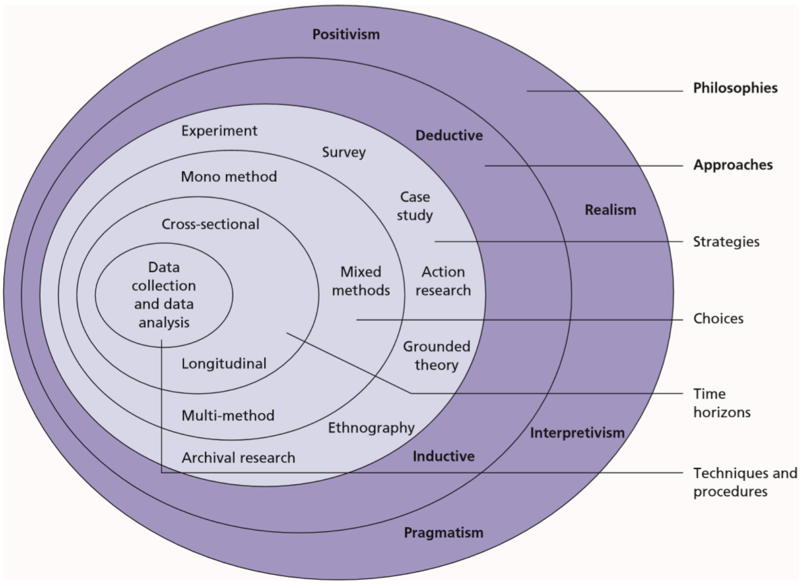

Figure 5.1: Research Onion (Saunders et al., 2009)

Positivism is often known as scientific method, empirical science, or positivist research. This paradigm is founded on the concepts of philosophers such as August Comte, Newton, and Locke and embraces a deterministic worldview in which causes determine effects (Creswell, 2014). It is based on observation, experimentation, and objective reality. According to Saunders et al. (2009), what defines knowledge in this paradigm is observable phenomena with a focus on causality law, generalisation, and a reductionistic approach. The researcher is supposed to remain impartial and value-free when doing the research. As a result, the data are gathered from highly structured, huge samples and are primarily quantitative. Johnson and Onwuegbuzie (2004), on the other hand, pointed out that the paradigm is not as objective and value-free as its proponents claimed because the researcher would determine what to study, develop the instrument, select a specific test, interpret the results, and determine which results are practically significant and subjective. **Realism** is a platonic philosophy that maintains that reality exists irrespective of human brains and discernment. It is, however, perceived by societal conditioning (Saunders et al., 2009). It is based on observable occurrences and is contextual in nature. As a result, unlike positivism, it is value-laden, and the researcher is likely to be influenced by world views, upbringing, and cultural experience (Saunders et al., 2009). According to the **interpretivism** paradigm, knowledge may be acquired through a variety of different methods and that reality is socially created, it entails gathering data through observation and giving it arbitrary interpretations. As a result, the researcher is involved and is inextricably linked to the research (Saunders et al., 2009). According to Johnson and Onwuegbuzie (2004), there are issues with the relativism or constructivism approach, it is challenging to determine which claim is reliable, and the concept of different realities raises issues.

5.3.1 Research Paradigm Adopted

Positivism research is the most appropriate paradigm to be used for this study since one of its objectives is set out to establish the true (most) applicable factor causing construction projects delay. The premise of positivism is that through observing social life, scientists may get accurate and consistent information about how society functions. Scientists frequently search for connections, or "correlations," between two or more variables while conducting positivist research. The comparative technique is what is used in this. In order to gain a comprehensive picture of society and identify social patterns, such as the link between educational attainment and social class, the positivist tradition emphasises the value of doing quantitative research such as large-scale surveys. Positivists advocate quantitative approaches with high reliability and representativeness, such as social surveys, structured questionnaires, and government statistics. Bryman (2008) proposed four key elements of positivism: only knowledge proven by the sciences is actual knowledge; theory creates hypotheses that may be examined for verifiable 'laws' according to deductive paradigm; science must be value-free in order to be objective; and inductive paradigm holds that knowledge is obtained through accumulating facts that serve as the foundation for laws. Furthermore, positivism paradigm has been widely adopted in related research. For instance, Egwim et al. (2021) adopted the positivism approach to extract the underlying factors causing construction project delay in Nigeria from the most applicable. Similarly, Egwim, Alaka, Toriola-Coker, Balogun, & Sunmola (2021) found the positivism paradigm suitable for developing a multilayer high-performing ensemble of ensembles predictive model utilising hyperparameter optimised ensemble machine learning techniques for construction projects delay risk mitigation.

5.4 Research Approach

The second layer of the research onion comprises the phrases deductive and inductive approaches. Knowing the purpose of the research and its constraints is crucial since the previous layer of the onion has an impact on this one. The deductive approach creates a hypothesis or set of hypotheses based on an existing theory before formulating a research strategy to test them (Silverman, 2013). The positivism paradigm, which allows for the formation of hypotheses and the statistical testing of anticipated findings to an agreed degree of probability, can be seen of as being particularly well suited to the deductive approach (Russell, 2010). It is defined as the progression from general to specific: first, the general theory and knowledge foundation are developed, and then the particular information obtained via the research process is compared to it (Kothari, 2012). However, a deductive approach may also be employed with qualitative research methods, howbeit in those situations, the expectations established by prior research would be created differently than through hypothesis testing (Saunders, Lewis and Thornhill, 2019). Deductive approach employs questionnaires to develop understanding of observation, allowing you to contrast various interpretations of individuals using actual facts. The acquired information aids in confirming or disproving the question; the procedure can be repeated. In contrast to the deductive method, which requires you to embrace an established theory, the inductive method lets you develop your own theories. This demonstrates how the two strategies differ from one another. The inductive approach is distinguished by a shift from the particular to the universal (Bryman, 2012). Since there is no framework that guides the data gathering in this approach, the study topic can be determined after the data has been gathered (Flick, 2011). Although the analysis of the data may reveal that it fits into an existing theory, it is also true that this may be considered as the moment at which new ideas are produced (Bryman, 2012).

5.4.1 Research Approach Adopted

Deductive approach was adopted for this study because it is consistent with positivist epistemology. By using empirical data, the deductive research approach enables understanding of many perspectives on corporate social responsibility activities. For this study, the deductive research approach will offer the observations required in formulating the conclusions that will answer the study's objectives. This is done by conducting primary research using a structured questionnaire in order to better comprehend the observations and address the research questions. Furthermore, because this study also used objectivism as the ontological philosophical stance, the choice was made to use a deductive research approach in the assessment of the observations. In contrast, an inductive research approach would have depended on subjectivity in the appraisal of observations to support the study findings. Additionally, the deductive approach, which employs questionnaires to develop understanding of observation by allowing researchers to contrast various interpretations of individuals using facts, aligns with one of the objectives of this study, which tends to carry out a survey of experts in the form of a questionnaire on the aggregated factors to establish the most applicable factors of construction delay in BIM-based infrastructure projects. The acquired information aids in confirming or disproving the research questions; and the procedure can be repeated.

5.5 Research Design

According to Johnson and Onwuegbuzie (2004), this is the strategy for going from gathering data to presenting the study findings. It is connected to research paradigms. It includes the research topics, the data gathering techniques, and the data analysis procedures. The research strategy or design might be either quantitative, mixed-method, or qualitative. As illustrated, there is little difference between the quantitative and qualitative techniques. A midway ground between the far-quantitative right's approach and the far-qualitative left's approach is said to be the mixed-method (Creswell, 2014; Johnson & Onwuegbuzie, 2004). The positivist paradigm is frequently associated with the quantitative method, the interpretivist paradigm with the qualitative approach, and the pragmatic paradigm with the mixed approach. The natural sciences are the foundation of quantitative design. This entails employing statistics and quantitative representations of the beliefs, patterns, and attitudes of a sampled population or testing whether a given therapy has an effect on a certain result using control and test groups. In this way, experimental or nonexperimental quantitative design is possible (Creswell, 2014). The generalizability, use of a large sample, creditability to many people in power, testing and validation of already constructed theories are highlighted by Johnson and Onwuegbuzie (2004) as the strengths, while the difficulty in application to a particular situation, lack of contexts, and the focus on theories that may lead to overlooking of relevant phenomena are highlighted as the weaknesses. With roots in the arts, sociology, and anthropology, qualitative design emphasises description and storytelling above quantitative. Examples include case studies, grounded theory, ethnography, phenomenological research, and narrative research (Creswell, 2014). In-depth analysis, description of complex events, context-oriented, and responsiveness to change are highlighted by Johnson and Onwuegbuzie (2004) as positives in contrast to its lack of generality and difficulties in producing a quantitative forecast. The drawbacks of the study include the time-consuming data processing and the researchers' subjective opinions. Mixed Design is described as "the type of research where the researcher mixes or integrates quantitative and qualitative research methodologies, methods, approaches, concepts, or language into a single study" by Johnson and Onwuegbuzie (2004). It seeks to minimise the shortcomings of both quantitative and qualitative design while maximising their benefits. Triangulation, complementarity, commencement, development, and growth can all be reasons for choosing a certain approach. According to Creswell (2014), the application of qualitative and quantitative methodologies throughout a research stage might take the form of convergent parallel, explanatory sequential, and exploratory sequential mixed approaches. Johnson and Onwuegbuzie (2004), on the other hand, argue that it can be a mixed model (combining qualitative and quantitative design within or across phases of the research process) or a mixed method (combining qualitative and quantitative design in the entire research process).

5.5.1 Research Design Adopted

This study adopts the quantitative design approach. Quantitative design was chosen for this study firstly because it helps to achieve the research aim as ML algorithms uses quantitative data (numerical values) for prediction. Also as a positivist research (positivists predict) quantitative data is required (Fuller *et al.*, 2001). Secondly quantitative design helps to achieve the research objectives – use of statistical analysis methods to find correlations between construction delay and potential covariates. Other research design (qualitative / mixed methods) was not selected since this research employs empirical process and empirical accounts – focusing on real world issues rather than ideology. Also qualitative method was not considered as its data cannot be read by most ML algorithms for classification-predictive purposes (Chen et al., 2018). More so, quantitative research method is based on objectivity and is especially useful when it is possible to obtain empirical measures of variables and findings from population samples. For the purpose of gathering numerical data, quantitative research uses formal tools and systematic processes. The numerical data is gathered methodically, objectively and often analysed through statistical tools such as SPSS, Python, R, Stata, or Scikit-Learn among many others.

5.6 Research Strategy

The research questions steer the study in the path that is chosen for it by the research strategy. It also goes by the name "research method" and includes techniques/approaches like "experiment," "survey," "case study," "action research," "grounded theory," “Interviews,” “systematic review” "archival research," and "ethnography" (Creswell, 2014; Saunders et al., 2009). These could be illustrative, descriptive, or exploratory. Experimental research is the strategy of developing a research procedure that compares the findings of an experiment to the predicted results. It can be applied to any field of study and generally requires the evaluation of a small number of variables (Saunders, Lewis and Thornhill, 2019). The research onion's survey strategy is frequently associated with the deductive approach. It is a superb and cost-effective research technique. This strategy allows you to acquire rich and trustworthy data. Surveys are commonly employed in quantitative research initiatives and involve a representative sample of the population (Bryman, 2012). The Survey strategy is usually used to identify contributing factors among many data sets. It enables the collecting of massive amounts of data that will be utilised to address the study question. Case study strategy is concentrated on a single person or group of people. It may provide insight into the particulars of every example and demonstrate the significance of culture and context in the variations among examples (Silverman, 2000). Case study research involves analysing a single unit to determine its important characteristics and make generalisations (Bryman, 2012). Action research strategy is mostly employed to identify a potential answer to a problem. In fields like teaching or nursing, action research is frequent so that practitioners may evaluate how to better their professional approach and understanding among many others (Wiles, Crow and Pain, 2011). A research strategy called "grounded theory" focuses on developing theories that are "grounded" in evidence that has been methodically gathered and examined. Data collecting in grounded theory begins without the development of a prior theoretical framework. A theory is built using information gathered from several observations. An ethnographic research strategy uses observing and/or interacting with study participants in their natural surroundings. Administrative records and papers are used as the main source of data in archival research. This category includes using documentation as the primary source of information in systematic reviews and other forms of documentation.

5.6.1 Research Strategy Adopted

In this study, surveys in the form of questionnaires were adopted as research strategy. Surveys are commonly used to answer questions like who, what, where, how much, and how many. It permits data collection from a large sample that can then be analysed using descriptive or inferential statistics to clarify links between variables or to create models of the relationships. Organized interviews, structured observation, and questionnaires are all methods of gathering data. The survey adopted comprises a questionnaire on a Likert scale with a range of one to five connected together in such a way that their individual responses will cumulatively aid to arrive at a discovery. The use of questionnaire research signifies independent observation – implies the questionnaire is completed in the absence of the researcher, and since this research is out to establish the true (most) applicable driver to construction delay makes it positivist research. A positivist researcher is usually independent (of the subject) as an observer, reduces a phenomenon to simpler measurable drivers (causes of delay in construction projects to be deduced from several literature will be reduced using Likert scale.), explains the elements with regards to how they affect the phenomenon (cause and effect) and often uses large samples (Burrell and Morgan 2008; Easterby Smith et al. 1991). The questionnaire was divided into sections such that each section deals with a specific feature of event under investigation (delay risk drivers). The questionnaire was divided into five sections A to E, such that each section deals with a specific feature of event under investigation (delay risk drivers). Section A gathered the responder's information, section B asked for frequency of occurrence of thirty-four factors, section C enquired what percentage a responder would give to fifteen factors and section D enquired to what level of detail one factor had. All these made a total of fifty factors as drivers (features/ independent variables). Also, the responders were asked to rate how long the entire project delayed for in the final section E; this represents the target (dependent variable).

5.7 Research Choice

Research choice is a term used to describe the fourth layer of the research onion. This layer enables researchers to determine whether using either one approach or a combination of quantitative and qualitative research is appropriate. The research onion has three described choices, including the Mono, Mixed, and Multi method research choices, according to Saunders et al. (2019). When employing the Mono research choice, researchers are required to collect only one sort of information, which might be quantitative or qualitative as the two cannot be combined. To get a precise collection of data, one might combine quantitative and qualitative techniques in a study using the mixed method. According to (Flick, 2011), the mixed approach mixes methodologies to generate a single dataset, while the multi method is utilised when the study is separated into segments, with each creating a particular data set. Since both the mixed method and the multi-method include quantitative and qualitative technique in a study, they are comparable. Despite their similarities, they yet differ in certain ways. Multi-method does not integrate methodologies to establish a specific collection of data, whereas mixed method does.

5.7.1 Research Choice Adopted

This study adopts the Mono research choice since it uses quantitative method as a single method for gathering its quantitative data (numerical values) for prediction. Also as a positivist research (positivists predict) quantitative data is required (Fuller *et al.*, 2001). Secondly quantitative design helps to achieve the research objectives – use of statistical analysis methods to find correlations between construction delay and potential covariates. In quantitative research as this, the data is often presented in numerical form, and this information is examined using techniques for quantitative data analysis. The quantitative research design is first and foremost concerned with the act of "measuring variables or counting instances of a phenomena." The qualitative research approach, in addition, emphasises "the themes and patterns of meanings and experiences associated to the phenomenon." (Collis and Hussey, 2009).

5.8 Sampling Technique

The idea of sampling is to obtain high-quality data when access to the complete population is not possible (Saunders, Lewis and Thornhill, 2019). By directly monitoring a subset of the complete population, sampling enables the estimation of a population's characteristics. Researchers are more interested in what can be inferred about the overall population from the sample than they are in the sample itself. A sample survey needs to be well prepared and characterised. The obtained data won't help in achieving the survey's goals if the improper questions are asked. The statistics won't provide a good representative of the population if the questions are posed to the incorrect individuals thus can lead to skewed outcomes. As Neuman (2014) suggests, the sample techniques assist in gathering the necessary data in a manageable, affordable, and time-efficient manner. Probability sampling and non-probability sampling are the two categories into which sampling techniques are classified (Saunders, Lewis and Thornhill, 2019). Probability sampling is the process of selecting a sample from a population when the selection is based on chance, random selection, or the randomization principle (Bell, Bryman and Harley, 2018). Bell et al. (2018) claims that a research based on surveys addresses the idea of probability sampling. There is an equal opportunity to be chosen for the research sample, and it permits the statistical generalisations of a population that is necessary for the study (Saunders, Lewis and Thornhill, 2019). Conversely, the non-probability technique selects samples for a carefully designed direction. Comparing probability sampling to non-probability sampling entails greater complexity, longer processing times, and often higher costs. However, since the population's units are chosen at random and the selection probability for each unit can be determined, accurate estimates can be generated, and the population may be inferred statistically. While using a non-probability technique, the researcher is free to choose the study sample in accordance with their own assessment (Saunders, Lewis and Thornhill, 2019). For this study, probability sampling technique is adopted. The objective of selecting a probability sample design is to reduce the sampling error of the estimates for the most significant survey variables obtained from the construction project experts via online questionnaire while also decreasing the survey's time and expense. The features of the survey frame, for example, might influence that decision due to operational restrictions. More precisely, this study adopts the stratified probability sampling method over other methods such as simple random sampling, systematic sampling etc because foremostly in the case of stratified sampling, the population is split into homogenous, mutually exclusive groups known as strata, and independent samples are then chosen from each stratum. The second reason is that the sampling strategy can differ from one stratum to the next since any of the other sampling techniques previously stated can be utilised to sample within each stratum. One last justification is that stratified sampling guarantees an appropriate sample size for population subgroups of interest. A sample size is determined for each stratum when a population is stratified, which turns each stratum into an autonomous population. By using the standard formula, sample size can be systematically determined in a quantitative study. In which the determination of the sample size is based on the size of the intended population, the allowable margin of error, the degree of confidence, and the distribution of the responses. Determining an appropriate sample size is critical in survey-based research to ensure the validity and reliability of the findings. In this study, the sample size for the survey was determined using a standard formula, considering the total population, desired confidence level, and acceptable margin of error. The calculation of the sample size (n = 342) in this study follows the established principles of inferential statistics, particularly designed for survey-based research. The formula applied is:

$$\boldsymbol{n} = \frac{\boldsymbol{N} * \boldsymbol{Z}^{\mathbf{2}} * \mathbf{p} * (\mathbf{1} - \mathbf{p})}{\boldsymbol{E}^{\mathbf{2}} * (\boldsymbol{N} - \mathbf{1}) + \boldsymbol{Z}^{\mathbf{2}} * \boldsymbol{p} * (\mathbf{1} - \boldsymbol{p})} \dots \dots \dots \dots \dots \dots \dots \, equation\ 5.1$$

Where: n n denotes the sample size, N represents the population size, Z is the Z-score corresponding to the chosen confidence level, p signifies the estimated proportion of an attribute in the population, and E refers to the margin of error, also known as the confidence interval. The total population considered for this study encompasses professionals and organizations engaged in BIM-based construction projects estimated at N = 2060. Estimating the precise number involved was crucial to accurately determine the population size. A standard confidence level of 95% was adopted for this study, aligning with common practices in social sciences research. The Z-score associated with this confidence level is 1.96, ensuring that the sample accurately reflects the population within the bounds of the selected confidence interval. Given the absence of specific data on the proportion of the population familiar with or utilizing BIM, a conservative estimate of p = 0.5 p=0.5 was used. This choice is prudent as it maximizes the required sample size, thereby guaranteeing adequate representation across different segments of the population. The margin of error was established at ±5%, a commonly accepted threshold that strikes a balance between precision and practical constraints. This margin ensures the results are accurate while remaining within the feasible limits of survey distribution and data collection efforts. By incorporating these parameters, the sample size calculation ensures a robust and statistically significant sample, thereby supporting the reliability and validity of the research findings.

5.9 Unit of Analysis

The level at which data are utilised to represent a single data point in an analysis is known as the unit of analysis (Tainton B.E., 1990). It is the social unit about which data is gathered, theories are developed, and judgments are drawn (Yang and Miller, 2017). This is a complex task to identify since the data are frequently at various levels, the study design, and the analysis's underlying assumptions may call for distinct analysis units from the measurement units. The unit of analysis selected has an impact on the study design, the number of participants or classes required, and the level of confidence we may have in the findings and conclusions. The categories of unit of analysis listed by Bless et al. (2013) are as follows: Individuals: This refers to situations in which a group of people, such as young adults, Christians, or black Africans, are the subject of a research study and analysis. Each person is a unit, and this unit of analysis is the most common. Groups: This entails researching several groups and maybe contrasting the groups. In this instance, each group symbolises a unit rather than just its individual members. Organizations: One form of group that is frequently employed as an analytical unit in social science research is an organisation, with each organisation in the study serving as a unit. Some companies can be contrasted based on their earnings, percentage of workers from particular backgrounds, efficacy of policies, corporate social responsibility, etc. Social artefacts: These are "products of social beings" and can range from poetry and letters to cars and farming equipment. Such artefacts may yield insightful knowledge about the people and organisations that used or made them through a methodical investigation. Period of time: This includes examining how something has evolved throughout time. Despite the apparent simplicity of these explanations, it is easy to misunderstand what a study's real unit of analysis is. The main challenge in choosing and deciding on an acceptable unit of analysis is deciding what you want to be able to say about at the conclusion of the study (Grünbaum, 2007). With the aforementioned explanations, it was determined that the construction project itself would be this research's unit of analysis, falling within the unit of analysis's "social artefacts" category. This is so because the research's data collection and analysis focused on the construction project delay. Additionally, delay in construction projects were a dominant problem in the conclusions of this study. The unit of analysis in research may differ from the unit of observation, which is "the entity on which the initial measurements are performed," even if they are often the same (Tainton B.E., 1990). The initial measurements in this study, such as late payment by the client or ineffective project supervision, were established specifically to gauge the extent of construction project delay; hence, construction project itself was also the unit of observation.

5.10 Data Collection

The two basic kinds of data collection methods are primary data collection and secondary data collection. These data collection methods are mainly used by researchers for many research initiatives. The primary data collection method that are most frequently used include surveys, interviews, and observations. Since the primary data collection is done with a specific goal in mind, it may be far more accurate and relevant for the research topics that have been selected. However, the significance of collecting secondary data is still valid for research endeavours. Reassessing previous data collected for a different reason is the essence of secondary data collection. In order to conserve resources like time and money, researchers gather secondary data. Researchers have divided the secondary data into two categories: raw data and compiled data. According to this viewpoint, researchers obtain previously verified data and use it for the objectives of their projects. Howbeit, even while ethical issues can be included in the study article, the researchers who want to use the secondary data have no influence over how reliable the information has been acquired. For this study, the primary data collection method is adopted since the study set out a specific goal to survey experts in the construction industry on critical drivers of construction projects delay as a way of approaching its quantitative data collection. Furthermore, primary data collection method is adopted since this study’s research data were generated by the researcher through an expert survey in the form of a questionnaire (see Appendix A) created specifically for comprehending and resolving the research topic at hand. The questionnaire used in this study is methodically structured on a Likert scale into five sections, A through E, each designed to capture specific data points relevant to the research objectives. This approach ensures that the survey generates detailed and nuanced responses, facilitating a robust analysis of the factors affecting construction project delays and the adoption of AI technologies in BIM-based construction projects. Section A of the questionnaire focuses on the frequency with which various factors associated with construction delays occur. Respondents are asked to rate the frequency of specific events on a scale from 1 (Zero) to 5 (Very Frequently). This section directly supports Objective 1, which aims to identify and categorize the most common factors affecting construction project delays. By quantifying the frequency of occurrences such as equipment breakdowns, labour disputes, and natural disasters, the survey provides empirical data that help in ranking these factors based on their impact and prevalence in BIM-based projects. In Section B, respondents are asked to estimate the percentage impact of certain factors, such as the proportion of late payments by the owner or the percentage of project activities affected by ineffective government regulations. This section's design allows for a deeper exploration of the severity and scope of each delay factor, correlating with Objective 1's goal of establishing a comprehensive understanding of delay factors. The use of percentage-based questions helps quantify the extent to which these issues affect projects, providing a clearer picture of their relative significance. Section C investigates the detail level of the project's schedule, ranging from 'No Schedule' to 'Detailed and Frequently Updated Schedule.' This section addresses both Objectives 1 and 3, by highlighting the importance of detailed scheduling in mitigating delays and assessing how these schedules influence the effectiveness of AI predictive models. A detailed schedule can be a critical feature in developing these models, as it provides a structured dataset for analysis. Finally, section D asks respondents to indicate the percentage by which the project exceeded its initial schedule, providing a direct measure of project delays. This section directly feeds into Objective 1 by quantifying the delay's extent, allowing for a more precise analysis of the impact of various factors. It also sets the stage for Objective 3, where these data points become critical variables in developing and validating AI models. The research questions and strategies influence the data collection techniques. As a result, various study questions/objectives may necessitate different data collection techniques. For each of the study's objectives, Table 5.1 lists the methods for collecting and analysing data. The objectives and deliverables are included in Figure 5.2's illustration of the general methodological flow of the study.

Table 5.1: Methods for collecting and analysing data

Objective	Data collection method	Tool
Conduct a systematic review toward gathering the most common factors affecting construction project delays and use it to conduct survey of experts to establish the most applicable factors affecting construction project delays in BIM-based construction projects.	Systematic Review, Expert Survey	Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA)

Objective	Data collection method	Tool
Conduct a systematic review of AI in the construction industry and use it to establish the most appropriate AI technologies during construction	Systematic Review	Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA)
Utilize the applicable drivers in the first objective as independent features (variables) to develop hyperparameter optimised AI predictive model(s) established in the second objective.	Expert Survey	Reliability Analysis, Exploratory Data Analysis, Data cleaning, Outlier detection, feature selection, and engineering, hyperparameter tuning, AI technologies

As an approach to its data collection, this study used online survey software to implement the questions in the questionnaire to gather data from respondents. In line with government policies on data protection regulations enforced by the university of Hertfordshire for all research activities involving the use of human participation, this study has obtained ethics approval - BUS/PGR/UH/05077 for its data collection method from the Social Sciences, Arts and Humanities ECDA. Prior to distribution of the questionnaire, pilot testing was conducted over a period of one month by asking group of experts in construction via email to comment on the representativeness and suitability of the questions. This is to ensure thorough understanding of the questions to be shared amongst potential participants and to avoid errors when recording data, to assess questions validity and the likely reliability of data to be collected (Saunders et al, 2015: p.425). These experts are highly experienced construction professionals with over ten years' experience in construction industries, including contractors, quantity surveyor, architects, technical consultants, civil engineers, site engineer, procurement managers among many others. Feedback, suggestions, and recommendations from them were used to readdress the questionnaire before making available its final version to all participants online. All participants had four months period (from 28th of July 2021 to 1st of October 2021) to complete the questionnaire. These participants were highly experienced construction professionals with over five years experience in construction sector, including contractor, quantity surveyor, consultants, and architects amongst many others, completed the online questionnaire. Consequently, resulting in a total of 324 responses received from 430 online questionnaire distributed.

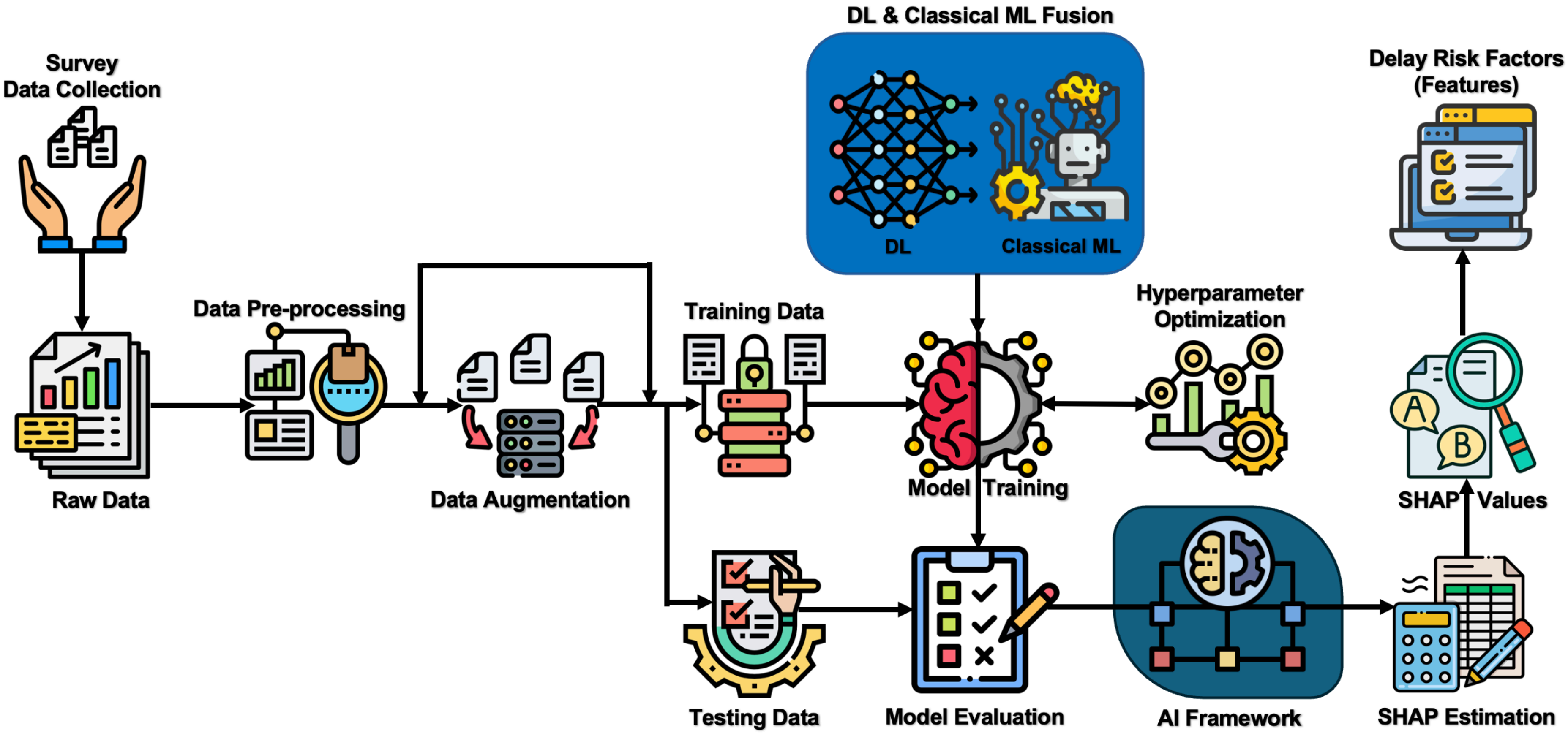

Figure 5.2: The general methodological flow of the study

5.11 Chapter Summary

This chapter talked about the research methodology for this study. First, it discussed the research philosophy. There are three parts to it - ontology (the study of reality), epistemology (the assumptions about knowledge), and axiology (the study of values and ethics). It then took a positivist, objectivist philosophical stance. Next, it covered the research paradigm. The main paradigms are positivism, realism, interpretivism, and pragmatism. The research adopted the positivism paradigm. It then discussed the research approach - whether to use deductive or inductive approaches. This study used a deductive approach. The research design refers to whether the study will be quantitative, qualitative, or mixed methods. This study adopted a quantitative research design. For the research strategy, this study used a survey method by administering an expert questionnaire. The research choice is whether to use a mono method (single data type), mixed methods, or multi-methods approach. This study used a mono-method quantitative approach. For sampling, it adopted a probability sampling technique, specifically stratified random sampling, to ensure representative samples. The unit of analysis explains what the study aims to analyse and draw conclusions about. For this study, the unit of analysis was construction project delays. For data collection and, the study used primary data collected through an expert survey questionnaire.

CHAPTER 6

6.0 DATA ANALYSIS[5]

6.1 Chapter Overview

By using clear visual aids like charts and tables together with their explanations, this chapter present the detailed data analysis of the research data collected through expert survey in the form of questionnaire. The data collected via the questionnaire (see chapter five) is analysed in this chapter to generate the quantitative variables (features) required for the development of artificial intelligence models for predicting potential delays in BIM-based construction projects, which will be developed in the following chapter. There are two major divisions in this chapter: exploratory data analysis detailing respondent profiles, and descriptive statistics.

6.2 Exploratory Data Analysis

One of the study's objectives, which was to provide answers to the research questions, was considered in the questionnaire design. The questionnaire was designed after much research and brainstorming. Meetings with representatives from the construction sector were held in order to determine the appropriate questions to ask and how best to deliver the answers. Additionally, special consideration was given to how well the questions would be understood by the responders. There were two main parts to the questionnaire. General background information about the respondents is provided in the first part, including their highest degree of education, position, and number of years of experience in the construction business and addressing the broad industry traits that respondents had in mind, such as the location and the type of construction project. A total of four hundred and thirty (430) online surveys were issued over a four-month period, and the researcher received three hundred and twenty-four (324) responses from them. This implies that the response rate was 75.35%. It was found while comparing their highest level of education that two hundred and five (205) (63.27%) of the respondents, who had the highest degree of education, held PhDs (see Figure 6.1).

---

[5] This chapter is primarily derived from the following journal articles:

Egwim, C.N., et al. (2021) 'Extraction of underlying factors causing construction projects delay in Nigeria', Journal of Engineering, Design and Technology, ahead-of-p(ahead-of-print). doi:10.1108/jedt-04-2021-0211.

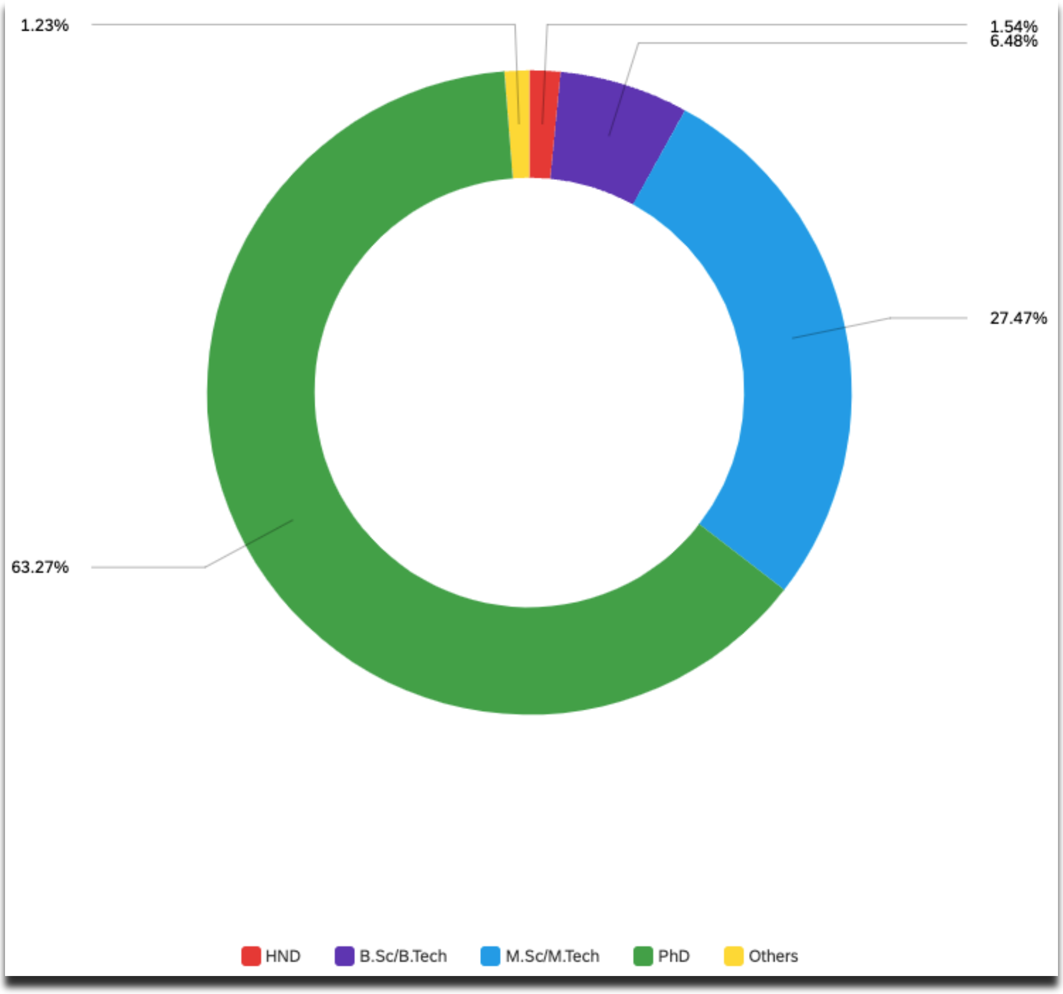

Figure 6.1: Highest level of education of respondents.

As shown in figure 6.2, 60% of the respondent year of experience was largely distributed between six and fifteen years – 33% fell between six and ten years and 27% between eleven and fifteen. This suggests that vast majority of the respondents are more aware of the working conditions on a construction site based on the construction project they had in mind when completing the survey and what their employers anticipated from them in terms of their job functions or performance.

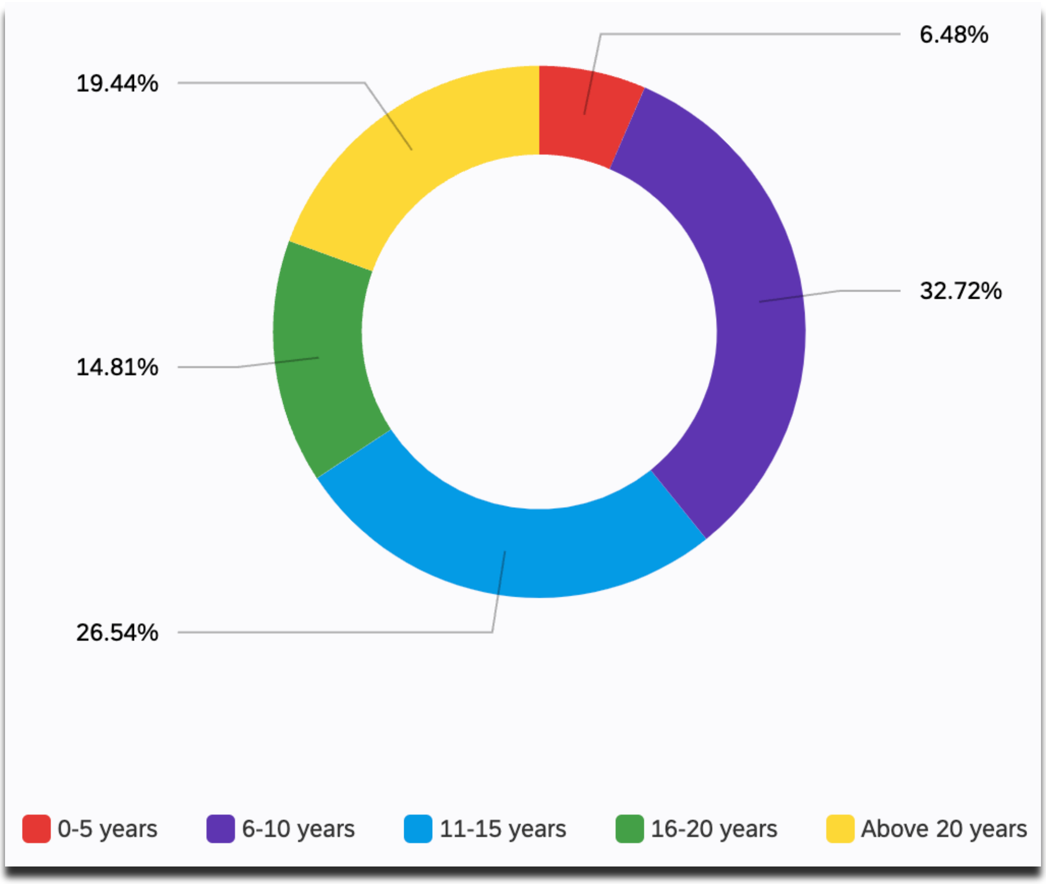

Figure 6.2: years of experience of the respondents.

For clarification, Figure 6.3 illustrated the specifics of the various construction project's professional cadres of respondents together with their percentages. More precisely, it shows that highly experienced construction professionals with over five years' experience in the construction sector including contractor (29.32%), quantity surveyor (27.78%), consultants (21.91%), and architects (17.59%) amongst many others completed the online survey. It is rather noteworthy that contractors make up a larger percentage of these experts that completed the survey as contractors are primarily responsible for carrying out and finishing major construction project tasks in accordance with the contract's terms, standards, specifications, timescale, and agreed-upon pricing(s).

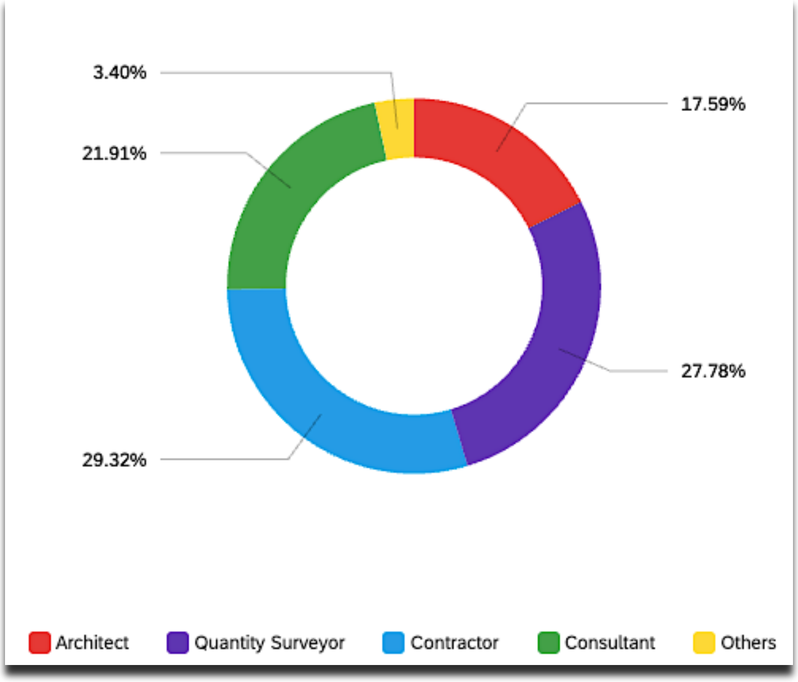

Figure 6.3: professional cadres of respondents

The majority of respondents represent stakeholders in the construction industry who have participated in project(s) with a budget of between one million and one hundred million pounds (see figure 6.4). This is very significant since it implies that the bulk of these construction projects are classified as medium or large projects, which is consistent with the European Commission's classification of project sizes (see chapter 1).

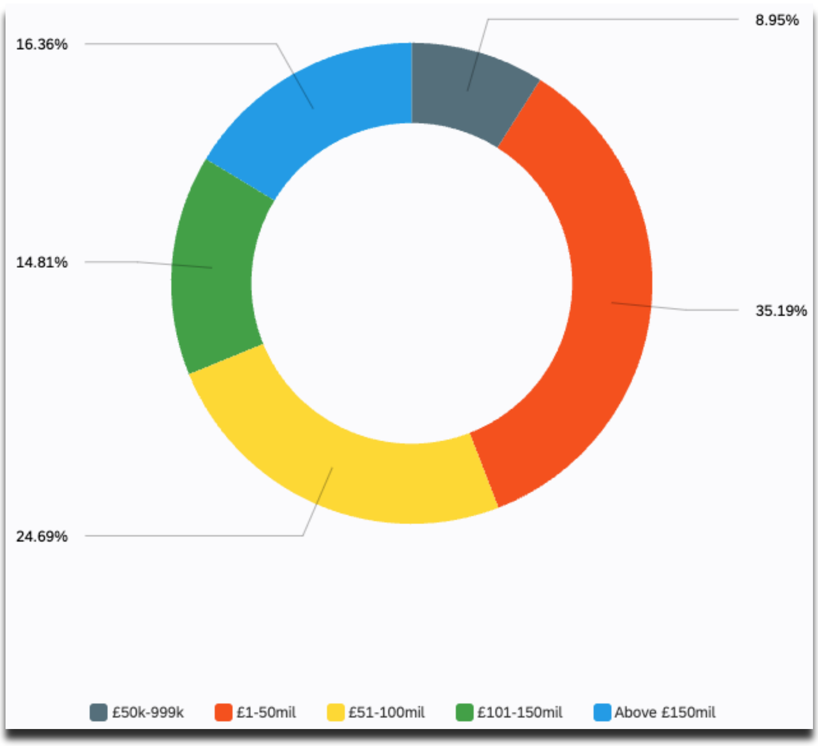

Figure 6.4: approximate construction project size.

Furthermore, an exploratory study of the survey results uncovered several types of construction projects that the respondents had previously worked on (see Figure 6.5). In more specific terms, it reveals that, out of the 324 responses, roughly 160 respondents (or 49.38%) have experience working on residential construction projects, followed by those who have experience working on infrastructure projects with 85 respondents (26.23%). This proves the viability of the proposed artificial intelligence model, which will be developed using this data as being capable of resolving delay issues in a range of construction projects. Additionally, it supports the conclusions of the first systematic review of this subject (see chapter 2), which identified infrastructure and residential projects as two of the three most crucial construction projects that encounter delays.

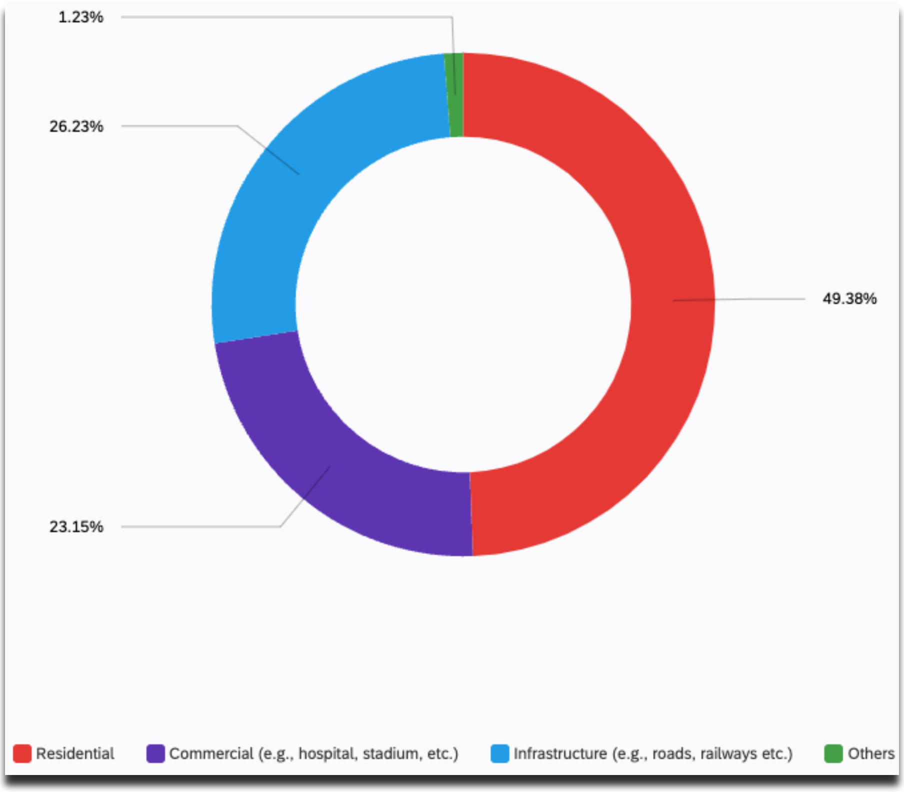

Figure 6.5: approximate construction project type

6.3 Descriptive Statistics

A comprehensive statistical analysis of the survey dataset, encompassing fifty-two factors as independent variable (F1 to F52) and project delay as dependent variable (F53), is undertaken (see Figure 6.6). This rigorous examination employs descriptive statistics (summary statistics) to provide an extensive understanding of the data. It includes **measures of central tendency, dispersion, and position**, offering a holistic view of the dataset's characteristics. Table 6.1 highlights the fifty-two applicable factors consolidated at the end of the survey exercise.

Table 6.1: Compilation of Independent Variables (Features) and the Dependent Variable (Target)

Factor ID	Factors
F1	Quality control issues
F2	Conflict between contractor and subcontractor
F3	Reworks due to error in construction
F4	Conflicts between consultant and contractor
F5	Inflation or sudden increase in commodities
F6	Labour dispute or strikes
F7	Unfavourable weather conditions
F8	Natural disasters like floods earthquakes etc.
F9	Contractors' financial difficulties
F10	Variation in structural design
F11	Late delivery of construction materials
F12	Change orders
F13	Discrepancies issues in contract documents
F14	Fluctuation in material prices
F15	Contractor cash flow issues
F16	Client cash flow issues
F17	Consultant cash flow issues
F18	Unforeseen site conditions
F19	Political Influence.
F20	Shortage of resources
F21	Space limitations at site for temporary equipment
F22	Space limitations at site for permanent equipment
F23	Changes in site topography after design
F24	Poor site investigation or management
F25	Shortage of materials
F26	Change of requirements by client
F27	Corruption issues

Factor ID	Factors
F28	Staff use of outdated construction methods
F29	*Staff use of new technologies*
F30	Site accident
F31	Change of specifications during construction
F32	Low productivity level of workers
F33	Poor decision making
F34	Change in economic conditions
F35	Equipment break
F36	Late payment by the owner
F37	Unskilled labourer
F38	Late delivery of ordered materials by suppliers
F39	Inaccurate budgeting
F40	Inaccurate resource planning
F41	Poor communication among stakeholders
F42	Ineffective project planning
F43	Poor contract management
F44	Ineffective government regulations
F45	Damaged storage materials
F46	*Nature of equipment(s) used*
F47	Ineffective project management
F48	Ineffective project supervision
F49	Pandemic
F50	Traffic restrictions
F51	Building permit
F52	*Schedule/program of work*
F53	***Project delay***

6.3.1 Measures of Central Tendency:

The measures of central tendency, namely mean, median, and mode, are fundamental to any statistical analysis. These measures provide a snapshot of the dataset by identifying a central value that best represents the data. For instance, the mean of unfavourable weather conditions (F7) as a delay driver is 2.69, suggesting that the average value of all observations by the respondents in the occurrence of unfavourable weather conditions during construction activities is approximately 2.69. This mean value on a scale of one to five suggests that the respondents' perceptions of unfavourable weather conditions during construction were neither extremely severe nor completely absent. This finding implies that the construction projects considered in the dataset were subject to varying degrees of unfavourable weather conditions. Additionally, it suggests that the respondents, who are individuals involved in the construction industry, reported encountering weather conditions that had some impact on the progress or execution of construction activities. The median, being the middle value when the data is arranged in ascending order, provides a robust measure of central location that is not skewed by outliers. From the survey data it is observed that the median of each independent variables also varies. The mode - the most frequently occurring value in a dataset, especially for categorical survey type of data, further enriches our understanding of the data by highlighting the most common observation. More so, the mode is not affected by extreme values or outliers in the dataset, making it a robust measure of central tendency. This is particularly beneficial when dealing with skewed distributions where the mean may be significantly influenced by extreme values, and the median may not adequately represent the most common outcome. The presence of multiple modes (bimodal or multimodal distributions) is eminent from the survey data (see Figure 6.6). For instance, independent variable F29 which represents adoption of new technologies among construction workers as delay risk driver share the mode value of 4.00. Similarly, independent variables F40, F42, and F43, which represents inaccurate resource planning, ineffective project planning, and poor contract manage as delay risk drivers respectively also shares the same mode value of 2.00 amongst many others. This suggests the presence of two or more distinct groups within the survey data, each centred around its own mode, indicating a concentration of respondents' ratings or perceptions in specific categories or levels of the independent variables. This can be indicative of underlying patterns or groupings that may not be immediately apparent from other measures of central tendency. More precisely, the adoption of new technologies, represented by F29, share the mode value of 4.00 suggests a notable group of construction workers who highly prioritize and extensively new technologies in their work. This finding implies the presence of a subgroup within the survey respondents who have embraced advanced technologies, such as BIM, and consider them integral to addressing delays and improving project outcomes. These construction workers likely have a shared perspective on the benefits and effectiveness of BIM and new technologies in mitigating delays and enhancing project efficiency. Similarly, the shared mode value of 2.00 for inaccurate resource planning (F40), ineffective project planning (F42), and poor contract management (F43) suggest a single factor underlying these responses.

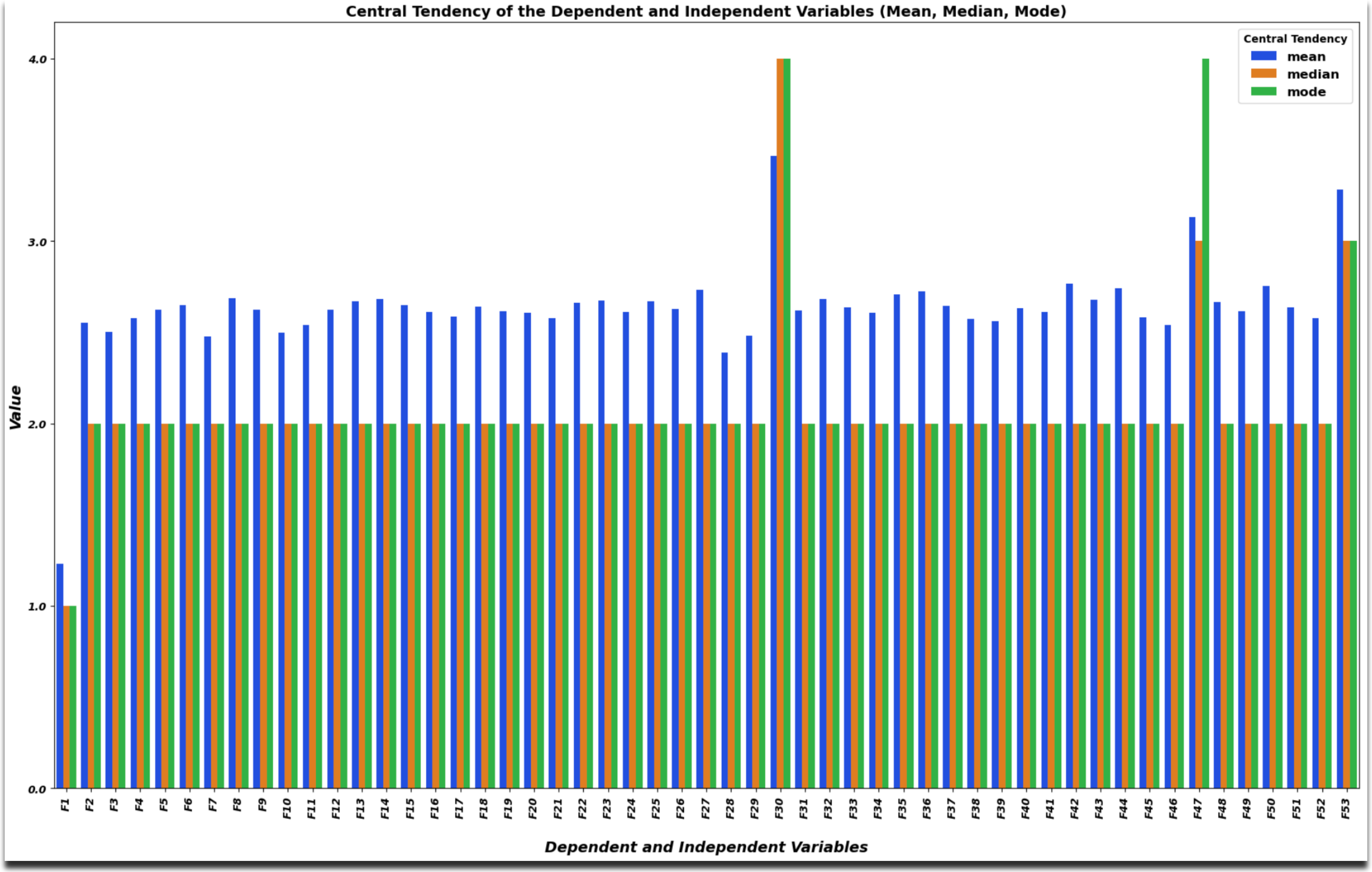

Figure 6.6: Central Tendency of the Dependent and Independent Variables (Mean, Median, Mode)

6.3.2 Measures of Dispersion:

The measures of dispersion, specifically the standard deviation, are equally critical in understanding the variability within the survey dataset because it quantifies the amount of variation or dispersion of a set of values. A low standard deviation indicates that the values tend to be close to the mean, while a high standard deviation suggests that the values are spread out over a wider range. This measure, therefore, provides a sense of the reliability of the mean, indicating whether the mean is representative of the data or if it is skewed by extreme values. Examining the survey data, the independent variable F41, which represents poor communication among construction stakeholders as a delay driver, exhibits a standard deviation value of 1.24 (see Figure 6.7) and a mean value of 2.76 (see Figure 6.6). This information has implications for the understanding of the data and the interpretation of the mean value. The values are spread out over a range, indicating that some respondents perceive poor communication as a significant delay driver, while others may not consider it as influential. The standard deviation of 1.1 was used as cut-off to visually distinguish between variables with relatively low and high variability. By using this threshold, variables that have higher dispersion in their values were highlighted, potentially indicating greater variability in responses. The higher standard deviation implies that there may be diverse perspectives or varying degrees of severity attributed to poor communication by the respondents. In relation to the mean value of 2.76, the standard deviation provides insights into the reliability of the mean as a representative measure of the data. With a relatively high standard deviation, there is an indication that the mean value may be influenced by extreme values or significant variations among the responses. It suggests that the mean value of 2.76 may not fully capture the range and diversity of opinions regarding the impact of poor communication on project delays. Hence, justifying the importance of considering the standard deviation alongside the mean to understand the distribution and variability of responses for a more comprehensive interpretation. Furthermore, the standard deviation helps identify potential outliers or extreme values that can significantly impact the mean. In the case of poor communication as a delay driver (F41), the standard deviation of 1.24 suggests the presence of substantial variation in perceptions, possibly influenced by unique or extreme viewpoints within the dataset. These outliers or extreme values, when present, can significantly influence the overall interpretation and decision-making processes related to addressing poor communication as a delay driver.

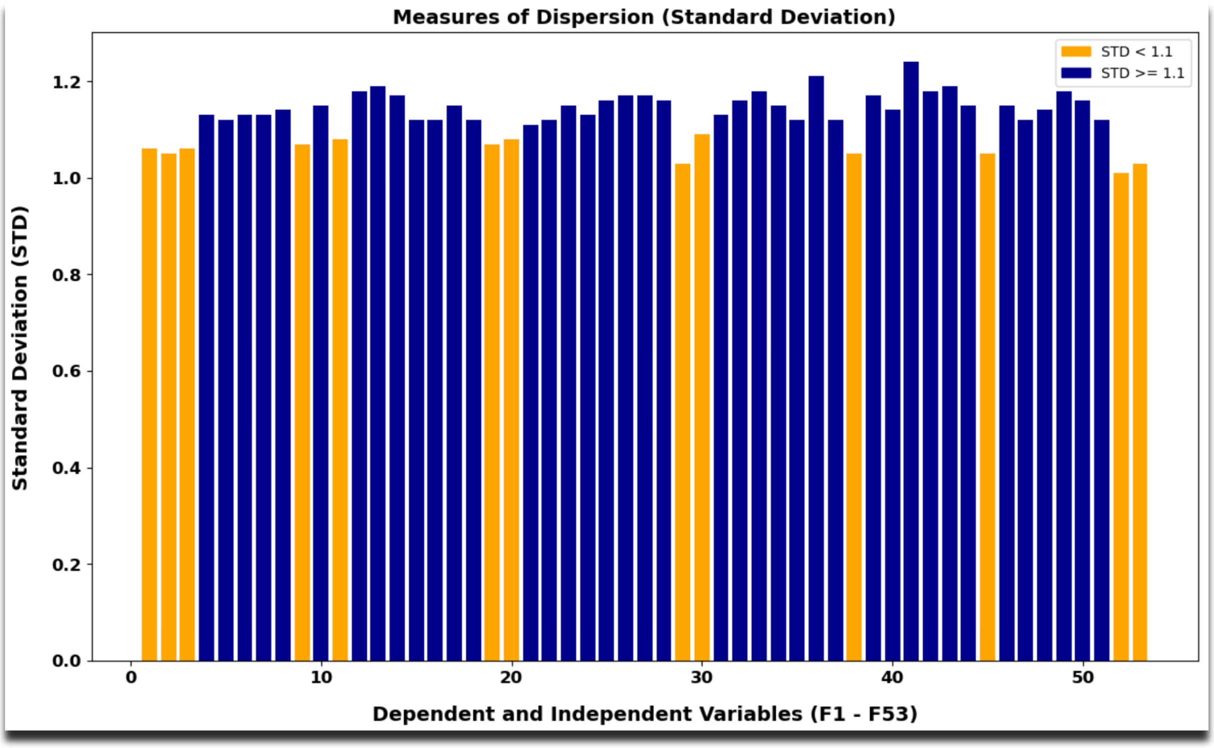

Figure 6.7: Measure of Dispersion of the Dependent and Independent Variables (Standard Deviation (STD))

6.3.3 Measures of Position:

The measures of position, including the minimum, 25th percentile, 50th percentile, 75th percentile, and maximum, provide further insights into the distribution of the data. These measures, also known as the five-number summary, provide a comprehensive overview of the data's spread. The minimum and maximum values define the range of the dataset, while the 25th, 50th, and 75th percentiles, also known as quartiles, divide the data into four equal parts. The 50th percentile is essentially the median which has been discussed in the previous subsection, providing a measure of central tendency that is not affected by outliers. The interquartile range, defined as the difference between the 75th and 25th percentiles, provides a measure of statistical dispersion that, like the median, is not influenced by outliers. Analysing the measures of position for these variables helps us grasp the distribution and relative severity of delay risk drivers in construction projects. The 25th percentile values provide insights into the lower quartile, indicating the level below which a significant portion of the projects fall. For instance, independent variable F8, which represents the occurrence of natural disasters, we observe that the 25th percentile value is 2.00 (see Figure 6.8). It is thus likely that less than half of the responses are more than 2.00 since the median of F18 is also 2.00. This information helps us understand the prevalence of natural disasters as a delay driver and provides a benchmark for comparison with other variables. On the other hand, the 75th percentile value represents the upper quartile or the value below which 75% of the projects experienced a frequency or severity equal to or lower than that value. As shown in Figure 6.8, the 75th percentile value of F8 (occurrence of natural disasters) being 3.00

indicates that 75% of the projects surveyed experienced a frequency of natural disasters equal to or lower than 3.00. This suggests that while the majority of the projects encountered a relatively low to moderate frequency of natural disasters, a significant portion still faced notable occurrences of these events, implying that natural disasters pose a substantial risk to construction projects, potentially causing delays and disruptions. In the same vein, as shown in Figure 6.8, the 75th percentile value of F34 (change in economic conditions) being 4.00 indicates that 75% of the projects experienced a change in economic conditions equal to or less severe than 4.00. This suggests that economic fluctuations have a considerable impact on construction projects, potentially leading to delays and uncertainties. Furthermore, comparing the 75th percentile values of F8 and F34 allows for a comparative analysis of the impact and severity of different delay risk drivers. In this case, the higher 75th percentile value of F34 (4.00) compared to F8 (3.00) suggests that economic conditions have a more significant influence on project delays than natural disasters. Additionally, comparing the 25th percentile value and the 75th percentile value of a particular delay driver, such as F8 (occurrence of natural disasters), can help identify potential patterns or trends in the impact of the delay driver(s) in construction projects.

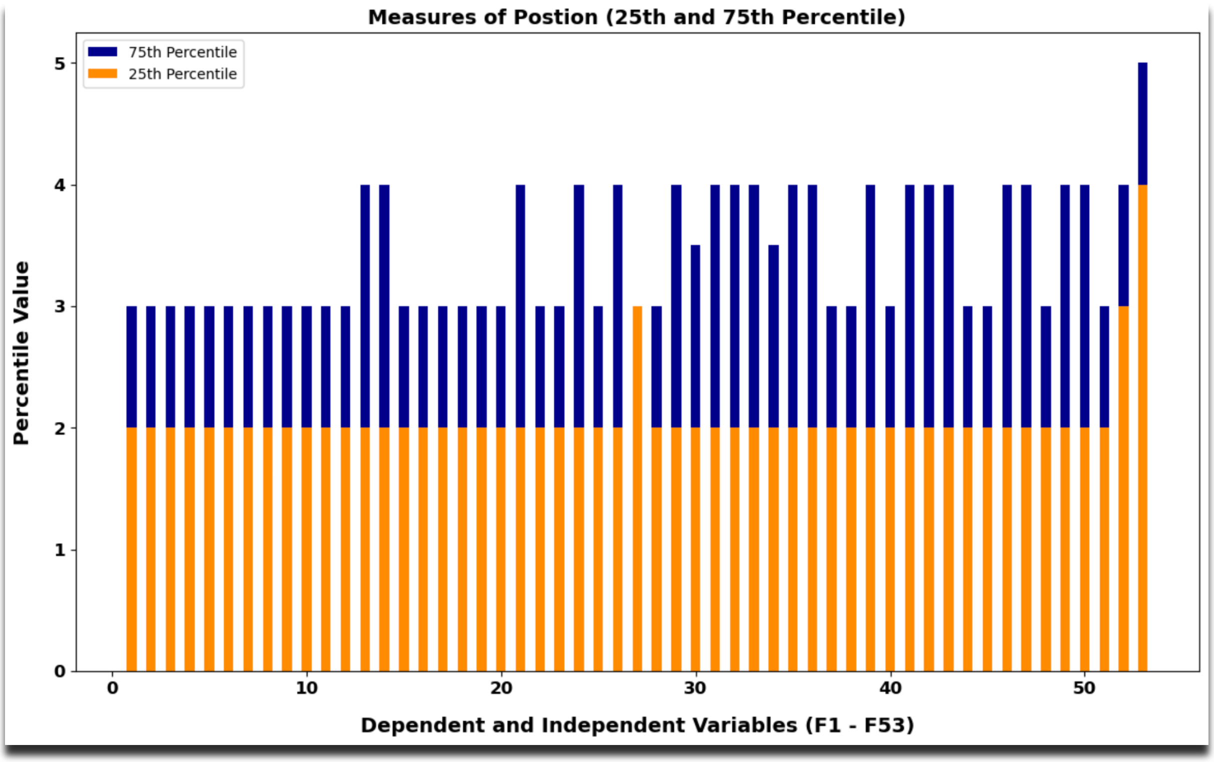

Figure 6.8: Measure of Position of the Dependent and Independent Variables (25th and 75th Percentiles)

6.4 Chapter Summary

This chapter presents a data analysis of the gathered research survey results with the principal objective of generating the requisite quantitative variables for the subsequent development of artificial intelligence models to predict potential delays in BIM-based construction projects. An exploratory study of the survey results reveals that the bulk of the erudite respondents, of whom approximately 63% held doctorates, had between 6 and 15 years of experience in the construction domain. This implies they likely possessed extensive discernment regarding the on-site working conditions and were cognizant of their employers' performance expectations based on the construction projects they had in mind when diligently completing the survey. Further analysis uncovers that contractor constituted the predominant share, approximately 29%, of the experts that completed the survey. This is particularly notable given that contractors bear principal responsibility for spearheading and concluding major construction project tasks in accordance with the contract's stipulations, benchmarks, specifications, timeframes, and agreed pricing(s). Additionally, an examination employing descriptive statistics, encompassing measures of central tendency, dispersion, and position, is undertaken to procure extensive comprehension of the dataset's characteristics. The measures of central tendency, namely the mean, median and mode provide a concise representation of the dataset by pinpointing a central value that best characterizes the data. For instance, the median facilitates a robust evaluation of the central tendency, being relatively unaffected by outliers and skewed data. Moreover, the presence of multiple modes in the dataset suggests two or more distinct respondent groups, each concentrated around a particular perspective regarding the impact of factors such as BIM usage. Furthermore, the measures of dispersion and position offer additional insights into the variability and spread of the data. The standard deviation quantifies the dispersion, indicating whether the mean reliably represents the data or is skewed by extreme values. The minimum, maximum, quartiles and percentiles outline the distribution range and spread, enabling comparative analysis of factor impacts across the projects.

CHAPTER 7

7.0 DEVELOPMENT OF AI PREDICTIVE MODELS FOR BIM-BASED AND NON-BIM-BASED CONSTRUCTION PROJECTS

7.1 Chapter Overview

This chapter explores into the intricate process by which the expert survey data, in conjunction with robust predictive algorithms, was harnessed to develop sophisticated predictive AI models aimed at addressing delays in both BIM-based and non-BIM-based construction projects. The chapter unfolds as a structured account of these endeavours, which commenced with a comprehensive data analysis. This process involved the judicious selection of relevant quantitative variables, as previously expounded upon in chapter six. This is achieved through processes such as data cleaning, the judicious selection of relevant features, and data engineering. Furthermore, the journey extends to the optimization of model parameters, fine-tuning them to enhance predictive accuracy. This iterative procedure aids in selecting the most suitable algorithmic frameworks, enriching our understanding of the complex dynamics within construction projects. This chapter plays a crucial role in transforming raw survey data into sophisticated AI models, which, in turn, enhance project efficiency by proactively addressing and mitigating delays.

7.2 Data Pre-processing

The process of data pre-processing assumes a paramount role as part of data preparation, encompassing various operations executed on raw data to render it amenable for subsequent data processing methodologies. It has long been underscored as an indispensable initial phase in the data mining continuum, poised not only for the training of artificial intelligence (AI) and machine learning models but also for the formulation of inferences derived from these models. This study, in its pursuit of comprehensive data pre-processing, explores into data profiling as a precursor, entailing a careful examination, evaluation, and review of data to extract statistical insights into its quality. The initial exploration of the expert survey data, unfolded through Exploratory Data Analysis (EDA). This analysis disclosed that the data manifests as a two-dimensional array, featuring 324 rows and 53 columns. The initial 52 columns (F1 – F52 factor IDs) represent features or independent variables, while the 53rd column (F53) serves as the target or dependent variable, as presented in Table 5.2 of section 5.

Descriptive statistics unveiled the discrete categorical nature of these columns, housing ordinal values ranging from one to five. Moreover, Figure 7.1 illustrates the distribution shape of the dataset, spotlighting that the median predominantly aligns with ordinal values two and three across most columns. The interquartile range, indicative of the concentration of central data, spans from approximately ordinal value two to three. Notably, the upper whisker extends predominantly to ordinal value four, encapsulating the majority of values, while the lower whisker reaches down to ordinal value one, signifying the presence of lower values that, however, do not qualify as outliers. Subsequently, the focus shifts to data cleaning, a process integral for rectifying quality issues within the dataset, such as addressing missing data, filling data gaps, and ensuring the overall adequacy of raw data for subsequent feature engineering endeavours. Notably, amidst the array of data imputation techniques available, including multiple imputation, interpolation, forward-fill, and backward-fill, this study adopts a judicious median imputation strategy for its ordinal data. The rationale behind this choice lies in the identification of missing values as completely at random during the survey, aligning with the suitability of median imputation for handling missing ordinal data in comparable studies (Graham, Olchowski, and Gilreath, 2007). Empirical evidence from studies, such as that by Enders and Peugh (2004), underscores the comparable performance of median imputation in terms of bias and standard errors, particularly when the extent of missing data remains low. Similarly, Cugnata and Salini (2017) affirm the supremacy of median imputation over alternative methods, citing its consistent outperformance in accuracy and reliability for handling missing values in ordinal data. The absence of **outliers** in the data distribution further attests to the robustness of the chosen imputation strategy, with all ordinal data values falling within the anticipated range (refer to figure 7.1). Post the exploratory phase, encompassing EDA, data profiling, and data cleaning, the analytical journey proceeds to a correlation analysis aimed at elucidating potential multicollinearity among predictors (features) in relation to the target variable. The correlation analysis relies on the computation of correlation coefficient values, which are subsequently visualized in Figure 7.2. This correlation matrix plot outlines the cross correlations between each feature (F1 – F52) and the target (F53). For instance, the positive correlation of 0.67 between F31 and F23, or the negative correlation of -0.27 between F29 and F28, signifies the interplay between different features. Notably, the criterion for identifying multicollinearity, an absolute correlation coefficient surpassing 0.7 among two or more predictors (Dormann et al., 2013), is rigorously adhered to in this analysis. The outcome is unequivocal as no evidence of multicollinearity exists among any pair of features or between the predictors (features) and the target variable, as discerned from the insights presented in Figure 7.2.

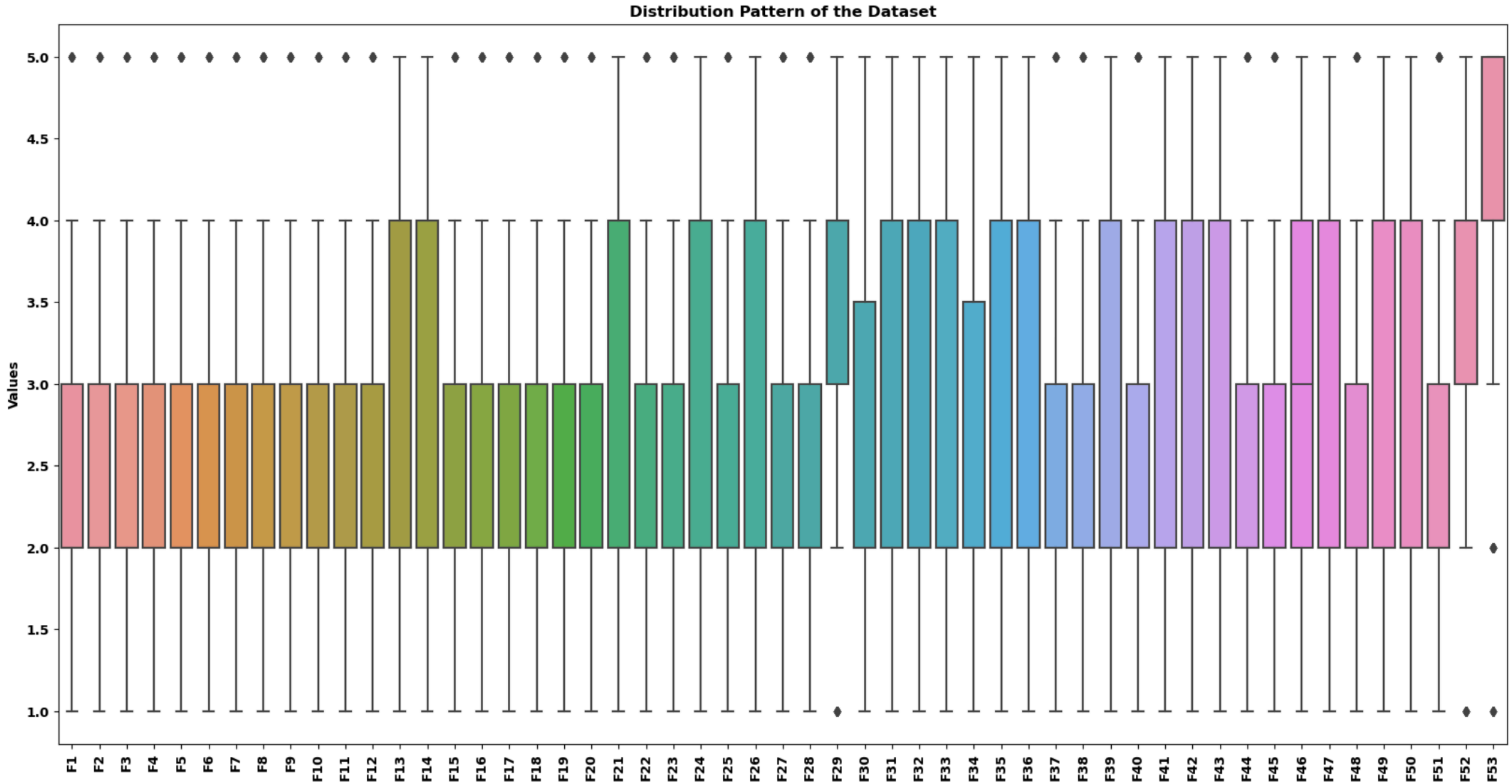

Figure 7.1: Distribution Pattern of the Dataset.

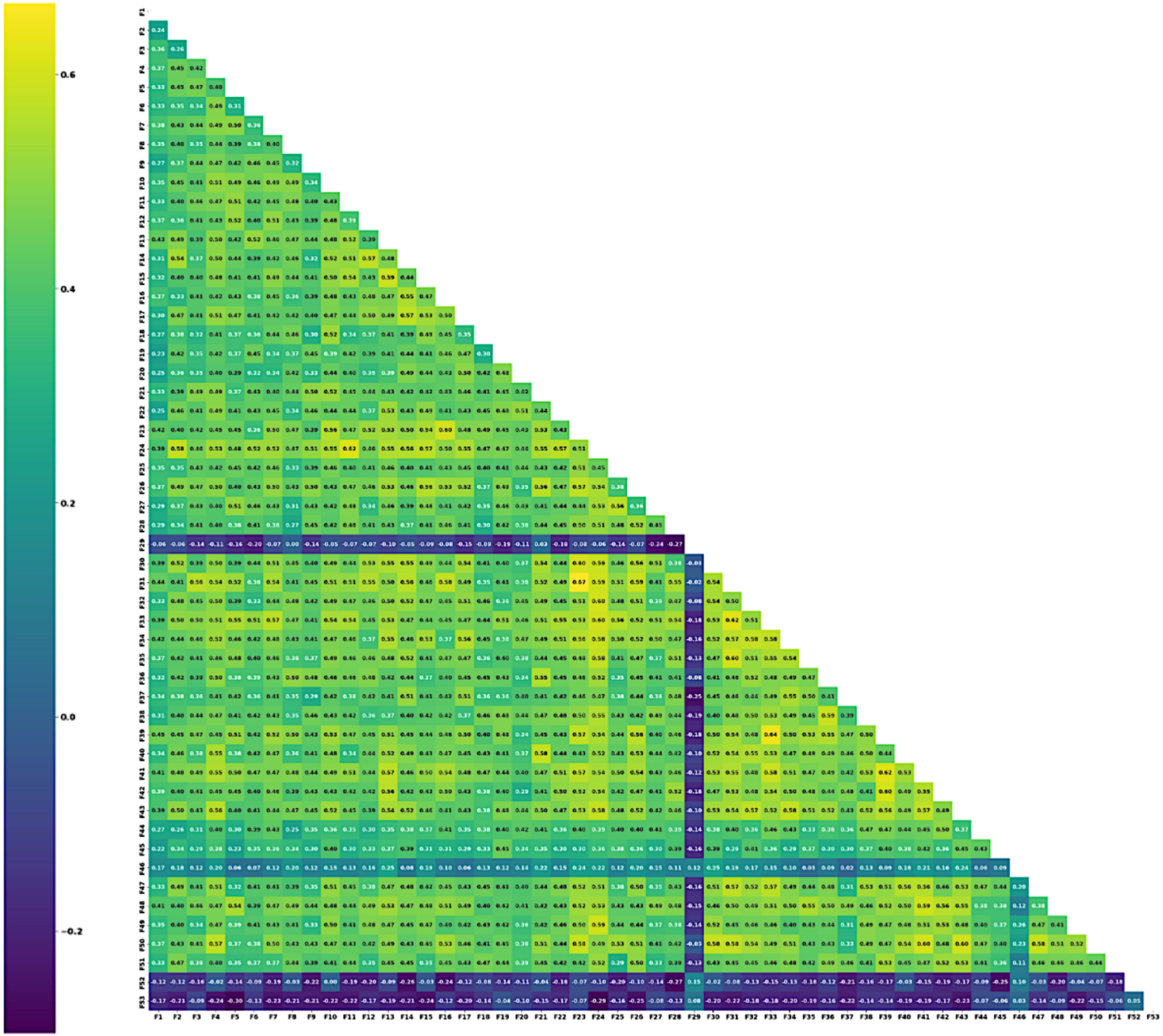

Figure 7.2: Visualization of the Correlation Matrix.

7.3 Feature engineering

Feature Engineering entails the systematic procedure of selecting, altering, or generating features (variables) derived from raw data with the aim of augmenting the efficacy of machine learning models. This process encompasses the conversion of the initial data into a refined representation, thereby enhancing the model's proficiency in discerning patterns, relationships, and underlying structures. Working with machine learning models necessitates feature engineering as a poor feature will directly affect your model, regardless of the data or architecture. Various tasks within the domain of feature engineering are extant, encompassing operations such as feature transformation, encoding, and scaling amongst many others. Furthermore, machine learning models often require numerical inputs, necessitating the conversion of categorical data into numeric representations. Encoding categorical data (such as the expert survey data for this study) serves the purpose of generating variables for model training and predictive features based on categories. Various encoding techniques exist, with this study opting for one-hot encoding for categorical data variables. One-hot encoding represents each category with a set of Boolean variables, indicating the presence or absence of a category for each observation. The chosen encoding technique retains all information of the categorical variable, does not assume category distribution, enhances data expressiveness, and allows for easy rescaling. Additionally, we more readily establish a probability for our values by utilising numeric numbers. In consequence, one-hot encoding is used for output values because it delivers more complex predictions than single labels. In this study, One-hot encoding (k-1 variant) was implemented as shown in Figure 7.3 on the target variable (F53). This procedure entailed encoding the data from column F53 into binary values: 0 (indicating no delay) for ordinal values less than or equal to 3, and 1 (indicating delay) for ordinal values of 4 and 5 (see Figure 7.3). This encoding facilitates the representation of the presence or absence of a specific category (delay) through the binary values of 0 or 1, respectively.

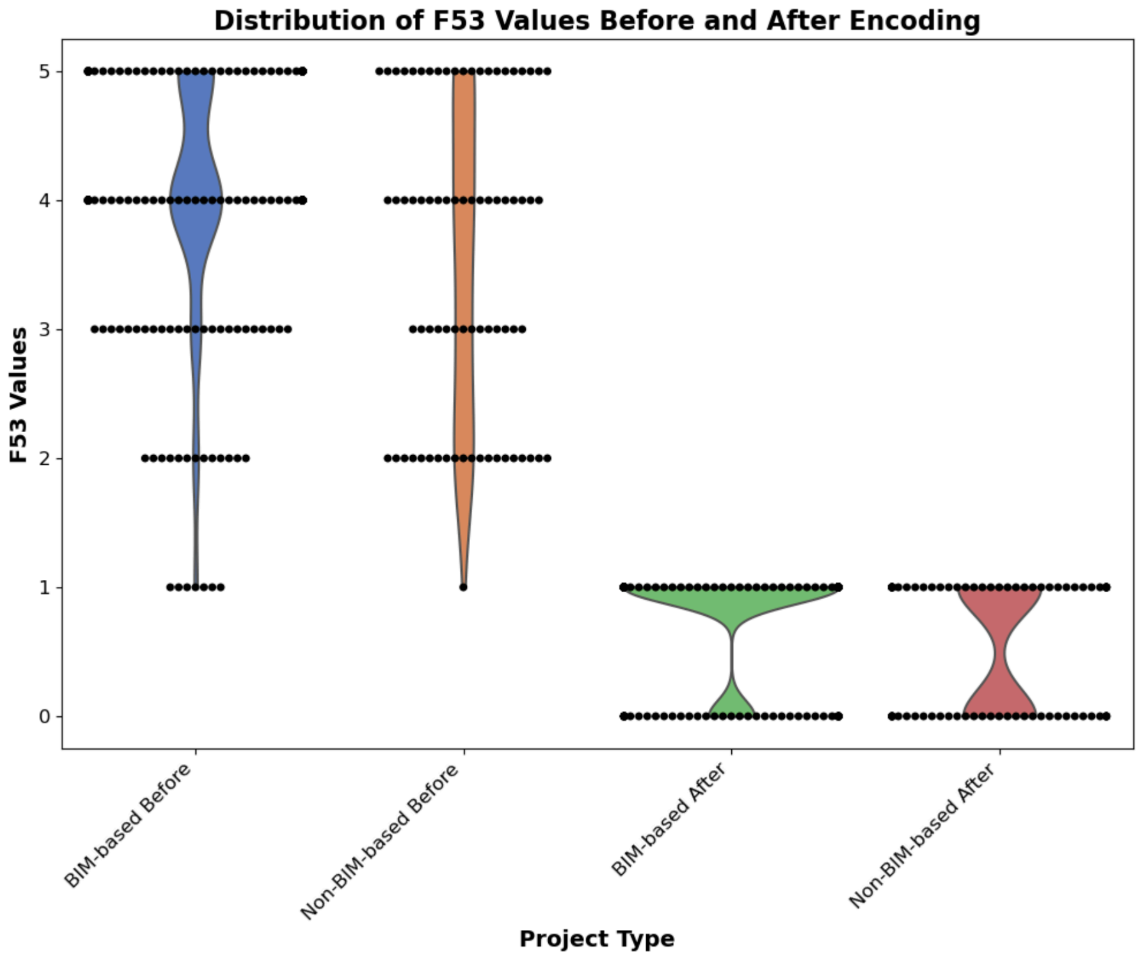

Figure 7.3: Visualizing One-Hot Encoding of the Target Variable

While one-hot encoding facilitates the bifurcation of the target variable into binary designations denoting the presence or absence of construction project delays, this encoding resulted in the introduction of class imbalance within the newly encoded target variable. Specifically, evaluation of the target variable distribution after one-hot encoding indicated that the 'delay' designation was substantially more represented within the dataset compared to the 'no delay' category. An analysis of the encoded column indicated the presence of 244 cases of project delays compared to only 79 instances of projects devoid of delays, highlighting a significant class imbalance in the ratio of 3:1 approximately (see Figure 7.3). The presence of skewed distributions with a predominant class poses an impediment during model development and evaluation, potentially encouraging overfitting towards the majority class. Class imbalance hampers model generalizability by diminishing performance on the minority class and associated edge cases. Furthermore, evaluation metrics conventionally employed under balanced class assumptions like classification accuracy become ineffective for imbalanced problems. Therefore, addressing this polarity in the bifurcated target variable distribution constitutes a vital step to facilitate an unbiased AI model, prevent majority class overfitting, and ensure generalization capability to unseen minority cases. Techniques commonly employed to address class skews include both data-level solutions of resampling the dataset and algorithmic methods of cost-sensitive learning. Oversampling the minority class through synthesis of new ‘no delay’ cases can ameliorate imbalance issues by balancing distributions. Specifically, the application of the Synthetic Minority Oversampling Technique (SMOTE) will generate synthetic samples of the minority category through interpolation along nearest neighbour instances. The key phases in SMOTE encompass the computational identification of k-nearest neighbours for each minority class sample based on feature space proximities, followed by stochastic linear interpolation to produce synthetic samples similar to but distinct from original neighbours. This strategic oversampling curtails overfitting by introducing acceptable data diversity, besides balancing distributions. Additionally, SMOTE will retain the original minority instances during this controlled oversampling process aimed at distribution parity between ‘delay’ and ‘no delay’ categories. Furthermore, cost-sensitive boosting or weighting techniques applicable during model training constitute complementary algorithmic approaches towards counteracting residual effects of imperfectly balanced classes.

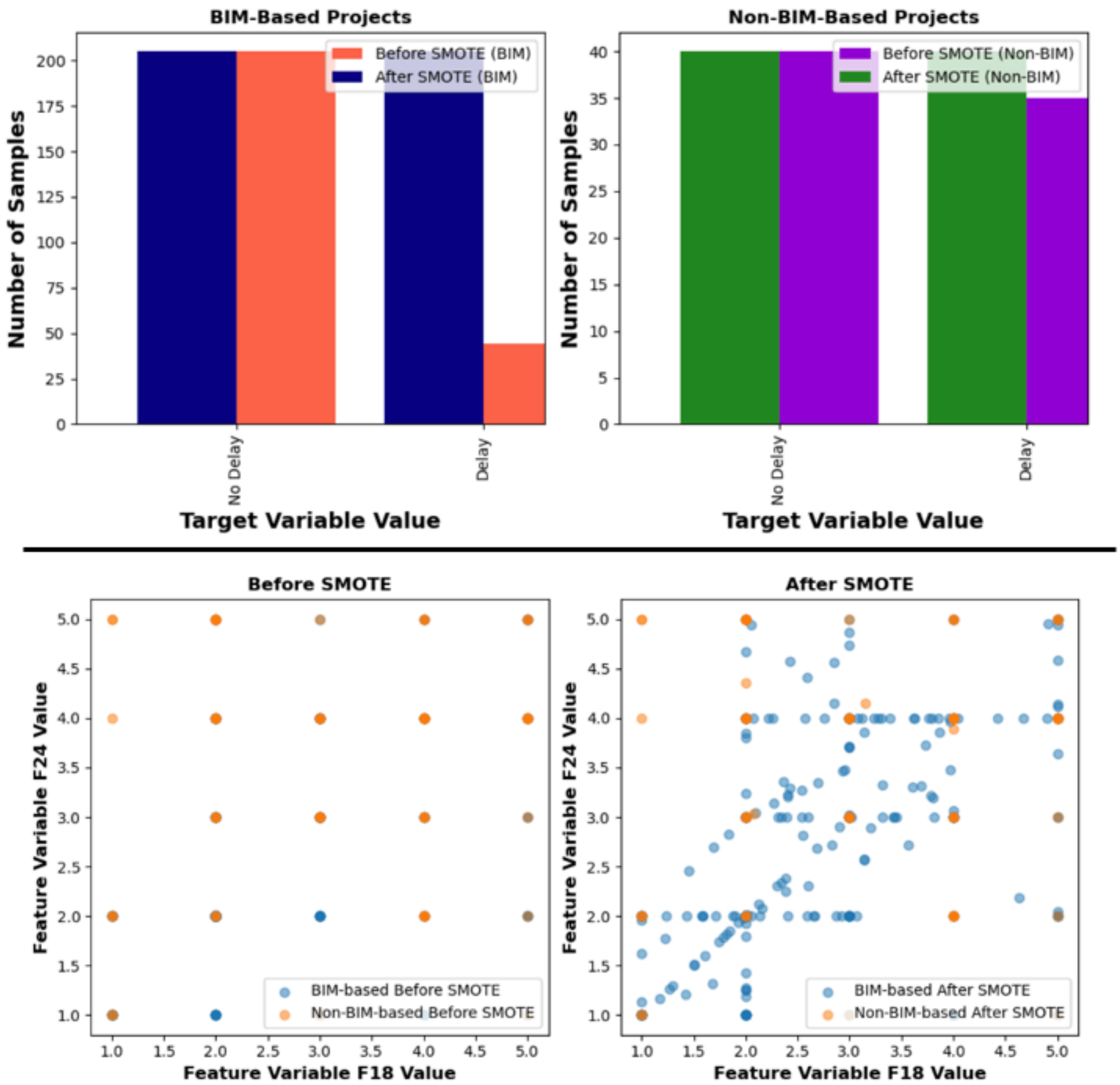

Figure 7.4: Feature and Target Distribution Before and After SMOTE Oversampling

In consideration of the threat imposed towards model efficacy and generalizability by skewed target variable distribution, this study implemented SMOTE based oversampling of the minority (no delay) class to achieve improved balance. Figure 7.4 illustrates the frequency distribution plots of the target variable both pre- and post-SMOTE implementation. Additionally, for the purpose of visually inspecting the synthetic sample data introduced to the feature variables, Figure 7.4 depict scatter plots comparing two specific features (F18 vs. F24) before and after SMOTE oversampling. This examination is of paramount importance for discerning the locations where synthetic samples are incorporated, thereby enhancing our understanding of how oversampling influences the feature space. The SMOTE process resulted in the generation of an additional 166 new samples, supplementing the initial set of 324 samples, thereby culminating in a comprehensive dataset comprising a total of 490 records. Machine learning models work better if the variables follow a normal distribution, often requiring specific transformations of features to achieve this normality (Raymaekers and Rousseeuw, 2021; Sun and Xia, 2024). However, primary datasets such as the data for this study commonly exhibit variables with skewed distributions. Consequently, enhancing the efficacy of machine learning models involves diversely transforming these variables and aligning their skewed distributions with a normal distribution. The procedures employed to standardize the range of values for independent variables are denoted as feature scaling. This is imperative due to the predisposition of variables with larger magnitude ranges to exert greater influence than those with smaller ranges. Additionally, expeditious convergence of gradient descent is facilitated when features possess comparable magnitudes. Therefore, maintaining uniformity in the scale or magnitude of all features is imperative to preclude disparate impacts, particularly when models are sensitive to magnitude. Various feature scaling techniques are available, encompassing mean normalization, min-max scaling, standardization, robust scaling, among others. This study adhered to the standardization feature scaling technique. This method was employed to fulfil the requirement by executing the subtraction of the mean from each observation of a feature and subsequently dividing the result by the standard deviation, as articulated in the equation and Figure 7.5 below:

$$\boldsymbol{X'} = \frac{\boldsymbol{X} - \bar{\mathbf{x}}}{\boldsymbol{\sigma}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots.\,\boldsymbol{equation\ 7.1}$$

Here, X' denotes the standardized value, X represents a specific feature observation, $\bar{\mathbf{x}}$ signifies the mean, and σ denotes the standard deviation. Consequently, the resultant dataset, post feature scaling, attains a variance of 1, centres its mean at 0, and exhibits a dynamic range between minimum and maximum values (see Figure 7.5). Subsequently, the dataset underwent partitioning, establishing a ratio of 60:40 for training (294 data points) and testing (196 data points). This rigorous approach ensures the model's proficiency in recognizing patterns and generalizing to unseen data instances, thereby contributing to the robustness of the analytical framework.

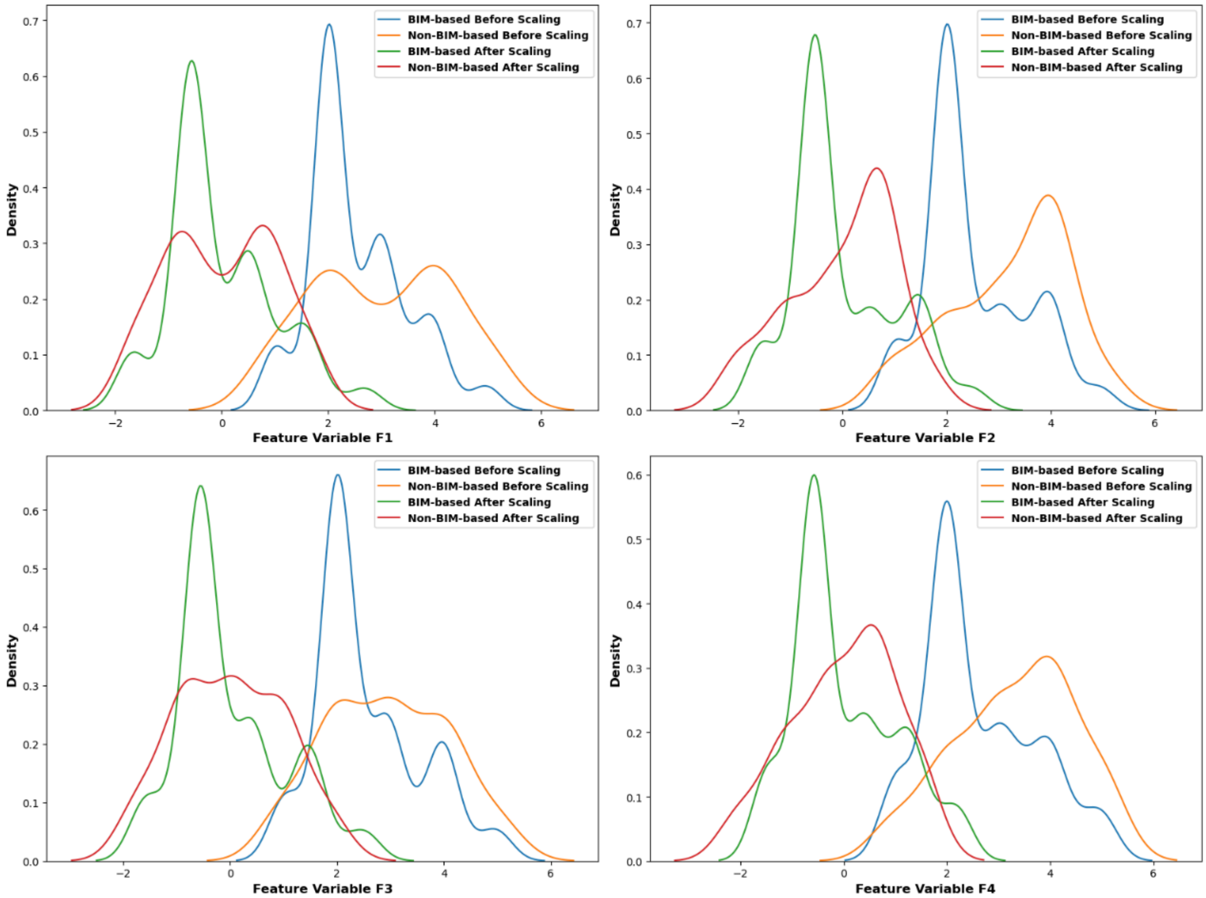

Figure 7.5: Feature Distribution Before and After Standardization Feature Scaling Technique

7.4 Feature selection

Feature selection constitutes an integral component within the machine learning pipeline by attenuating dimensionality and generalization error while enhancing model interpretability (Ling *et al.*, 2023). The identification and extraction of the most informative predictor variables from the raw dataset facilitates efficacious development of predictive models . As shown in Figure 7.6, two predominant paradigms of feature selection exist, supervised and unsupervised, differentiated by utilization of labelled output data. Supervised techniques harness the target variable to guide the selection process, evaluating predictor relevance founded on statistical relationships between features and responses. This enables the distillation of explanatory variables with maximal predictive capacity regarding the outcome of interest. Conversely, unsupervised methods disregard output data, instead detecting latent data patterns and intrinsic characteristics (Kong *et al.*, 2022). Thereby unsupervised feature selection explores feature relevance devoid of specified objectives in an undirected manner. Optimal feature selection for the present study necessitates deliberation of the survey dataset structure comprising the table of features F1 through F52. These features annotate potential indicators and drivers of delay. The target variable is manifested in the ordinal feature F53, quantifying construction delay on a Likert scale. Therefore, the predictive modelling objective entails prognosticating the construction delay response founded on the predictor features. This thus renders supervised feature selection techniques most suitable by leveraging correlations of features (F1 to F52) with the target (F53) to deduce explanatory variables relevant to delay prediction. Consequently, the study employs the supervised feature selection techniques encompass filter methods, wrapper methods, and embedded methods to evaluate and select features predictive of the target variable. The study thereby capitalizes on supervised feature selection to inform predictive model development through selection of delay-relevant predictors from the domain-specific expert survey data.

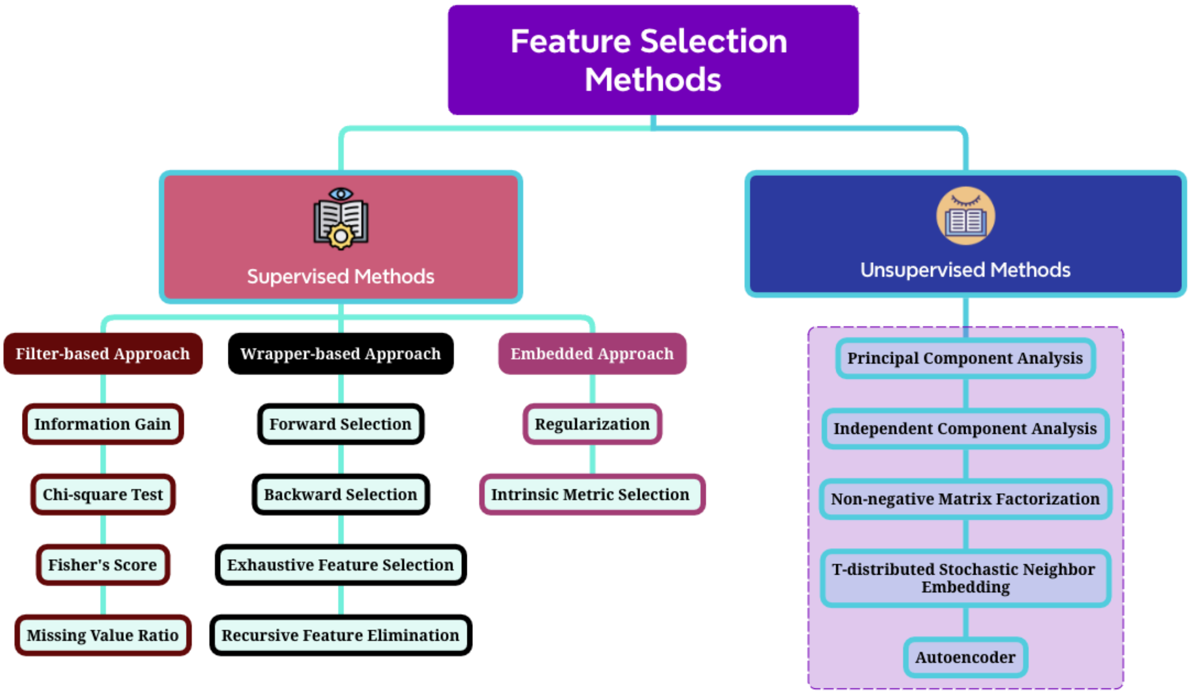

Figure 7.6: Feature Selection Methods

7.4.1 Filter-based Approach

Filter-based feature selection refers to techniques that apply a statistical measure to score and rank the relevance of features based on intrinsic data properties, independent of any machine learning model (Theng and Bhoyar, 2024a). The core premise relies on filtering features prior to model fitting guided solely by characteristics of the data itself through univariate relationships between the feature and target variable. These approaches quantify the descriptive utility of each feature in isolation and retain those demonstrating statistical significance for predicting the output. Common scoring criteria include distance, information, dependence, and consistency metrics that rank feature correlation, mutual information, or separability between classes (Gong *et al.*, 2024). Methods generally order features by the computed metric and eliminate those below a threshold. Many apply sequential searches greedily adding or removing based on ranking. As the methodology filters variables without reference to a model, advantages of pre-filter methods focus on their universality, efficiency, and reusability. Statistical evaluations scale well to ultra-high dimensions and run faster than wrappers or embedded routines. The capability to pre-process once and apply across diverse models makes this stage amenable to pipelining. However, filters assessing individual relevance overlook multivariate interactions and struggle with collinear groups of useful predictors. Next, we explore the predominant techniques and metrics leveraged in filter-based feature selection schemes. This study explores into the implementation of various statistical filter approaches, encompassing information gain, the chi-squared test, and Fisher's score. Notably, the missing value ratio, another filter method, is omitted from consideration due to prior handling of missing values during dataset pre-processing.

Information Gain

Information gain constitutes a prominent filter-based feature selection technique leveraging the principle of entropy from information theory to quantify the predictive signal within features (Zhou *et al.*, 2024). It operates by measuring the reduction in uncertainty about the target variable when knowing the feature versus when ignoring it. A higher decrease in output unpredictability in the presence of a given input signifies elevated information content encoding greater predictive utility. Mathematically, information gain computes the divergence between target variable's entropy without and with conditioning on that feature (Lim *et al.*, 2022). The term entropy quantifies unpredictability within a random variable, or the information required to describe it. For categorical outcomes, it tracks the homogeneity versus heterogeneity of class distribution. Higher entropy indicates more balanced class splits demanding more information to predict correctly. Figure 7.7 illustrates the outcome of the information gain implementation, revealing that features F5, F22, F35, F42, and F51 are among the top five features based on the feature importance score. Information gain, quantifying the reduction in output uncertainty achieved by knowledge of the predictor variable, highlights variables causing substantial decreases in target entropy as broadly relevant, irrespective of the model. Consequently, features useful for classification will reduce this target entropy thereby exhibiting higher information gains. The computation proceeds by first tabulating the unconditional entropy for just the output vector based on its class probability distribution. Thereafter, it incorporates conditioning on the candidate feature by averaging target entropy across the feature's values weighted by those probabilities. The final gain emerges from the deviation between the conditional and unconditional terms - larger drops in uncertainty signal variables more informative for the task. Beyond classification contexts, information gain generalizes for ranking usefulness of features with respect to continuous numerical regression problems too (Bolón-Canedo, Sánchez-Maroño and Alonso-Betanzos, 2015). Here, it measures the variance reduction for the target when incorporating knowledge of the predictor. Intuitively, features explaining more variation in the output demonstrate greater association and predictive merit. Calculation simply replaces entropy with variance while maintaining an identical relative conditioning framework. The innate sensitivity to both categorical and real-valued data coupled with low computational demands enable widespread adoption of information gain filters. Also, it avoids strict assumptions or distributional requirements while detecting monotonic and non-linear interactions.

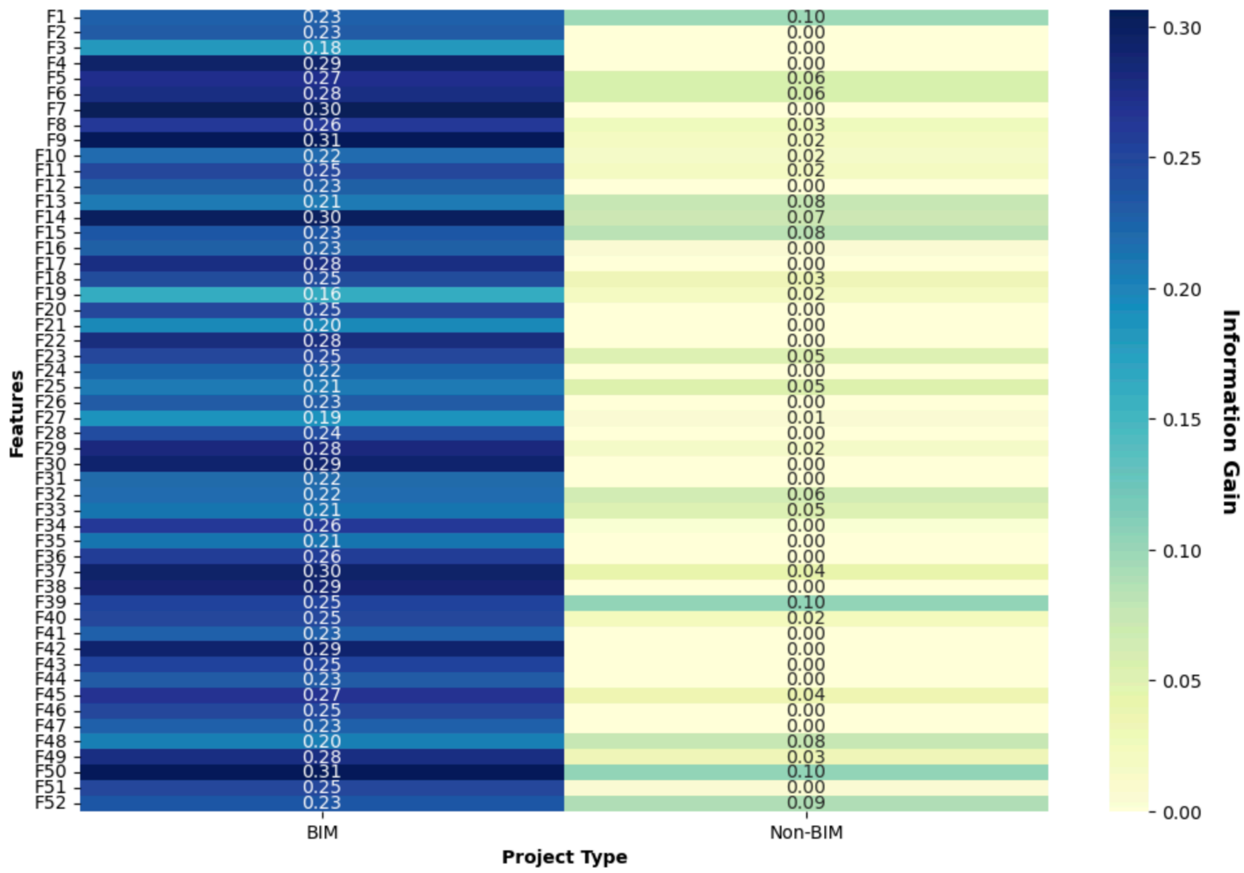

Figure 7.7: Information Gain Ranking of Features

Chi-Squared Test

The chi-squared test constitutes a statistical filter for evaluating association between categorical features and a target variable(Dey *et al.*, 2022). It tests divergence from the null hypothesis that a given categorical predictor and output are independent. High chi-squared values indicate dependence thus rejecting the independence assumption signalling useful classification features. Computationally, it contrasts observed frequencies of predictor-target combinations against expected counts if they were unrelated (De Caro *et al.*, 2023). Higher discrepancies reveal statistical dependency worthy of retention. Formally, it calculates normalized sum of squared differences between observed and expected frequencies across the contingency cells. Interpretation examines deviation between the resultant score distribution against a theoretical chi-squared distribution with degrees freedom equal to number of cells minus 1. Scores situated at distribution tails with low probability thresholds reject independence. Typical filtration retains features whose p-values from chi-squared test fall below 0.05 or 0.01 levels as thresholds for statistical significance of dependency. Figure 7.8 demonstrates its examination of the correlation between features and outcomes, resulting in features F50, F13, F2, F5, and F15 as the top features based on the feature score. Predictors with significant p-values indicate predictive associations with the response variable. As a univariate statistical filter, chi-squared analyser enjoys certain advantages aligned with filter methodology (Egwim, Alaka, Pan, *et al.*, 2023b). It avoids expensive model training and hyperparameter tuning. Robustness against collinearity and high dimensions enables excellent scalability. Also, the adjustment of the p-value cut-off provides smooth control over filtered subset size. Furthermore, subsequent multivariate validators on retained features improve optimality. However, limitations such as use of rigid thresholds and isolation from modelling context hinder selectivity relative to wrappers. Still, as the most common calculation for testing categorical association, chi-squared filter retains merit for rapidly pruning irrelevant high-cardinality spaces. Coupling with pre-binning methods for numerical features and ensemble aggregation to improve stability keeps it relevant amidst fierce filter competition. Integrative pipelines blending with other categories henceforth present a pragmatic path forward.

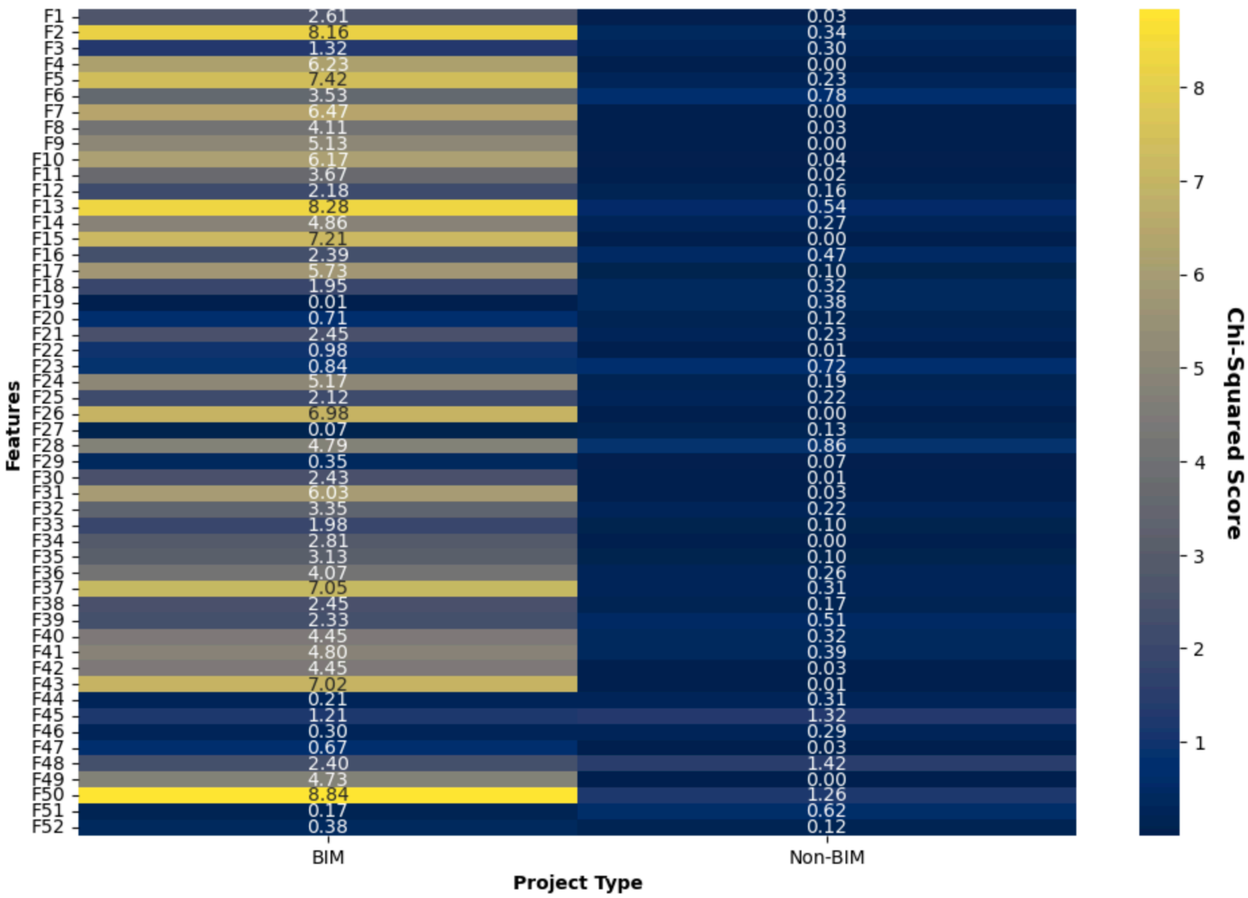

Figure 7.8: Chi-Squared Ranking of Features

Fisher's Score

Fisher's score operates as a supervised statistical metric for gauging relevance of features through the lens of class separability (Niwas *et al.*, 2015). Conceptually, it aims to quantify the extent of discrimination between subgroups partitioned by the target variable based on evaluation feature dispersion within those segmented strata. Mathematically, the formula factors class means showcasing separation and inverse in-class variances capturing compactness. Specifically, it evaluates the difference of means adjusted by a normalizing constant equalling arithmetic average of the variances of that feature restricted within samples from each target class. Intuitively, the numerator targeting mean divergence seeks higher values indicating greater subclass differentiation power for that predictor variable towards segregating the decision groups. Meanwhile, the denominator aims for lower intra-class variances implying tighter homogeneous clustering per group to aid partitioning. Division of the two terms thereby distils class distinction factor with balance. Fisher's score, computing a relevance metric based on the ratio of inter-class separation and intra-class cohesion, yields feature F2, F50, F5, F15, and F37 as the most important features in Figure 7.9. High scores, signifying discrimination between categories of the target variable along with homogeneity within groups, suggest informational relevance. Features manifesting elevated Fisher's scores demonstrate substantial shift or alteration between class-conditional densities aiding discrimination. Thus, filtration via Fisher's criterion retains variables exceeding a threshold that contribute most towards distributional divergence between target categories to assist classification. Selection relies on sorting scores across predictors and pruning away subsets associated with insignificant scores (Sun *et al.*, 2021). Although limited to labelled data as a supervised metric, advantages such as linear complexity, resilience to correlations and outlying values make Fisher's filter suitable for expediting high-dimensional searches. Limitations around univariate nature prone to missing multivariate interactions find mitigation in ensemble, bootstrap, and combined filtering remedies. Overall, as a pioneering discriminatory measure, Fisher's score persists relevant given its speed, scalability and sensitivity towards informing classifiers through computed estimates of class separability.

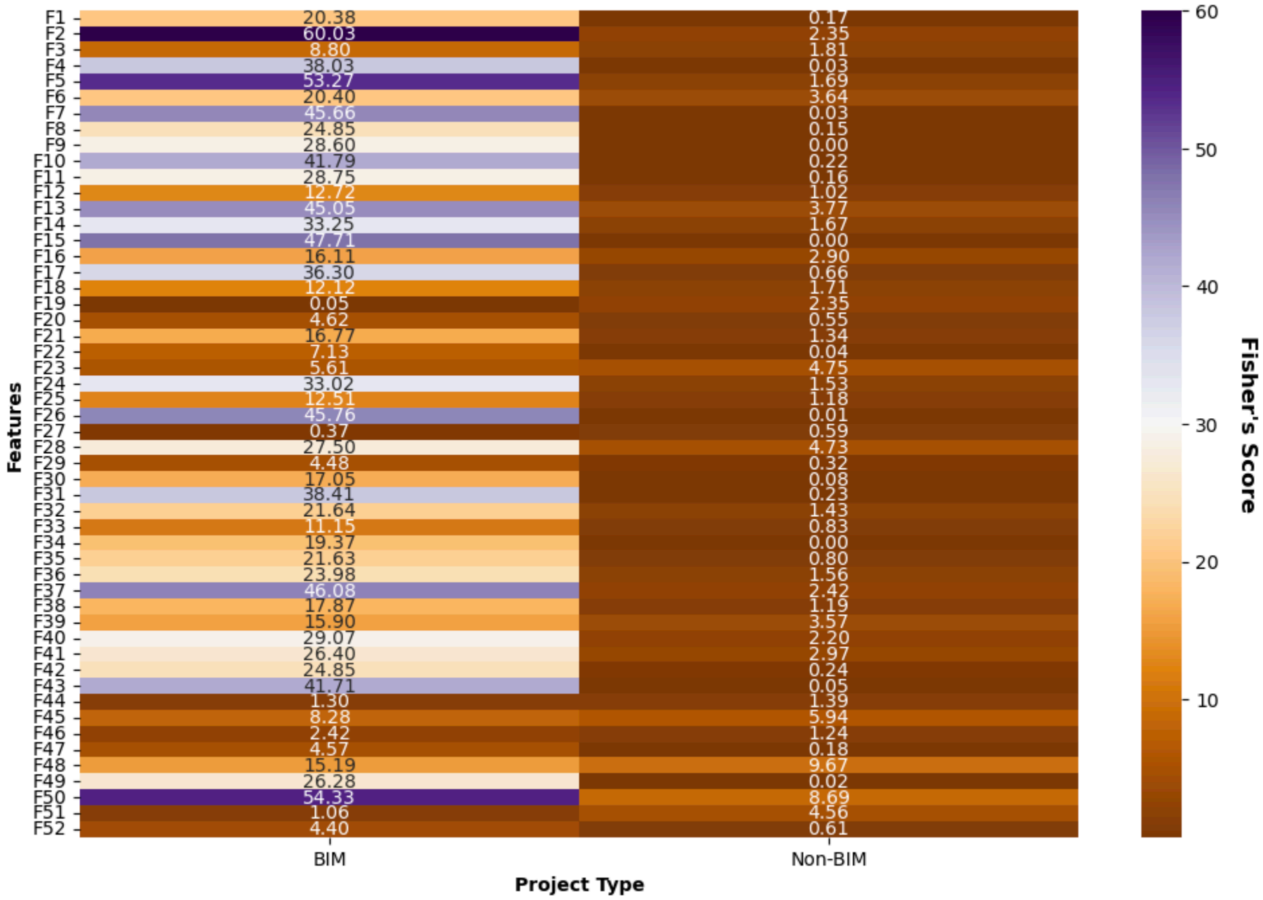

Figure 7.9: Fisher's Score Ranking of Features

7.4.2 Wrapper-based Approach

While these statistical filter approaches provide a rapid indication of feature relevance without relying on model induction or scoring on a specific validation dataset, they do not optimize relevance for the nuances of a particular model algorithm or peculiarities of dataset partitions. Instead, filtering techniques estimate generalized predictor utility, introducing the risk of retaining redundancies or excluding interacting variables relevant only for specific algorithms. As such, this study employs filtering as an initial screen to extract broadly useful features and eliminate consistently underperforming variables. Subsequent wrapper-based and embedded feature selection techniques refine relevance, tailored to the chosen model and intrinsic dataset structure (Maldonado, Riff and Neveu, 2022). This consolidation of filter, wrapper, and embedded approaches aims to achieve an optimal feature space, balancing computational efficiency, statistical power, generalized utility, and tailored model-specific selection. Wrapper methods constitute a systematic search approach that evaluates subsets of predictors by gauging their influence when presented to a particular machine learning model. Inherently, these methodologies intricately link the process of feature selection to the overall performance of the model, systematically exploring various combinations to derive the most optimal features tailored to the specific algorithm (Talukder *et al.*, 2024). The quintessence of wrapper methods lies in their ability to confer a personalized selection of features, aligning variables with the nuanced inductive biases of the model. Moreover, these techniques implement iterative approach, employing greedy sequential search methodologies such as forward selection, backward elimination, recursive feature elimination, and exhaustive feature selection. This iterative nature emphasizes a methodical exploration of the feature space, allowing the model to incrementally assess the relevance and impact of each variable. Such a deliberate process aims to unravel the intricacies and dependencies among predictors, ensuring a comprehensive understanding of their individual contributions to the model's predictive capacity. The essence of wrapper methods lies in their capacity to tailor the selection of features to the specific requirements and nuances of the machine learning model (Eskandari *et al.*, 2024). By integrating the assessment of subsets directly into the model's evaluation process, wrapper methods transcend mere feature filtering. Instead, they encapsulate a holistic approach that considers the interplay and synergies between variables, recognizing that their collective impact is greater than the sum of individual contributions. Furthermore, wrapper methods operate with a consideration of model performance, consistently seeking to enhance predictive accuracy and generalization capabilities(Sahebi *et al.*, 2020). The iterative nature of these methodologies facilitates an ongoing refinement process, where the model dynamically adapts its feature set based on the evolving understanding of predictor importance. This dynamic adaptability ensures that the selected features not only align with the intrinsic biases of the model but also evolve with the changing dynamics of the dataset. It is imperative to acknowledge that wrapper methods introduce an element of computational intensity due to their exhaustive exploration of feature subsets. However, this computational cost is justified by the nuanced and tailored selection process, which ultimately leads to improved model performance. The trade-off between computational resources and enhanced predictive accuracy underscores the strategic value of wrapper methods in the realm of feature selection. Moreover, wrapper methods offer a principled framework for navigating the vast landscape of potential feature combinations. By employing sequential search techniques, these methodologies systematically traverse the feature space, evaluating various configurations to pinpoint the most influential predictors. The figures 7.10, 7.11, 7.12, and 7.13 below uses radar charts to present complex multivariate survey data used in this study in a visual format. Radar charts are standard figures is appropriate and widely accepted in the discipline to compare multiple variables across different categories or groups and for summarizing and visualizing complex datasets in a clear and accessible format(Zhang *et al.*, 2015; Abeynayake *et al.*, 2023). These charts allow for an intuitive understanding of how different variables interact or contrast with each other, providing a comprehensive view that is easier to interpret. These radar charts illustrate the performance or characteristics of different features used in the study. For example, it compares the contributions of various delay factors in BIM and non-BIM construction projects / the effectiveness of different AI models in predicting project delays. By displaying each factor on a separate axis, radar charts visually convey the relative magnitude and differences across these dimensions, making it easier to identify patterns, strengths, and weaknesses (Chen *et al.*, 2022).

Forward Selection

Forward selection is a prominent technique in feature selection, distinguishes itself by its systematic approach to initiating the search process with an empty set, gradually incorporating features to optimize model performance (Saha, Patikar and Neogy, 2020). This method involves a careful inclusion of variables, prioritizing their individual contributions and aggregating them based on their significance in enhancing efficacy. An illustrative example of the strategic implementation of forward selection is depicted in Figure 7.10, where a line plot portrays the progressive augmentation of top features, resulting in a notable improvement in the model's accuracy to an impressive 79% and 77% for BIM-based and non-BIM-based construction projects respectively. At each iterative stage of forward selection, the model undergoes thorough scrutiny, evaluating the remaining variables to identify the most beneficial feature that maximizes the scoring metric. This incremental addition of features, though myopic, proves to be computationally efficient and tailored to the model's requirements, persisting until specific termination criteria, such as reaching a saturation point, are met. However, it is paramount to underscore the significance of selecting appropriate stopping criteria to forestall the pitfalls of overfitting and unnecessary computational burdens. Furthermore, the efficacy of forward selection hinges upon the discerning choice of stopping criteria, ensuring a delicate balance between model complexity and predictive performance (Puggini and McLoone, 2017). By adhering to stringent criteria, forward selection not only mitigates the risk of overfitting but also streamlines the computational overhead, thereby yielding a parsimonious yet robust feature subset conducive to superior model generalization.

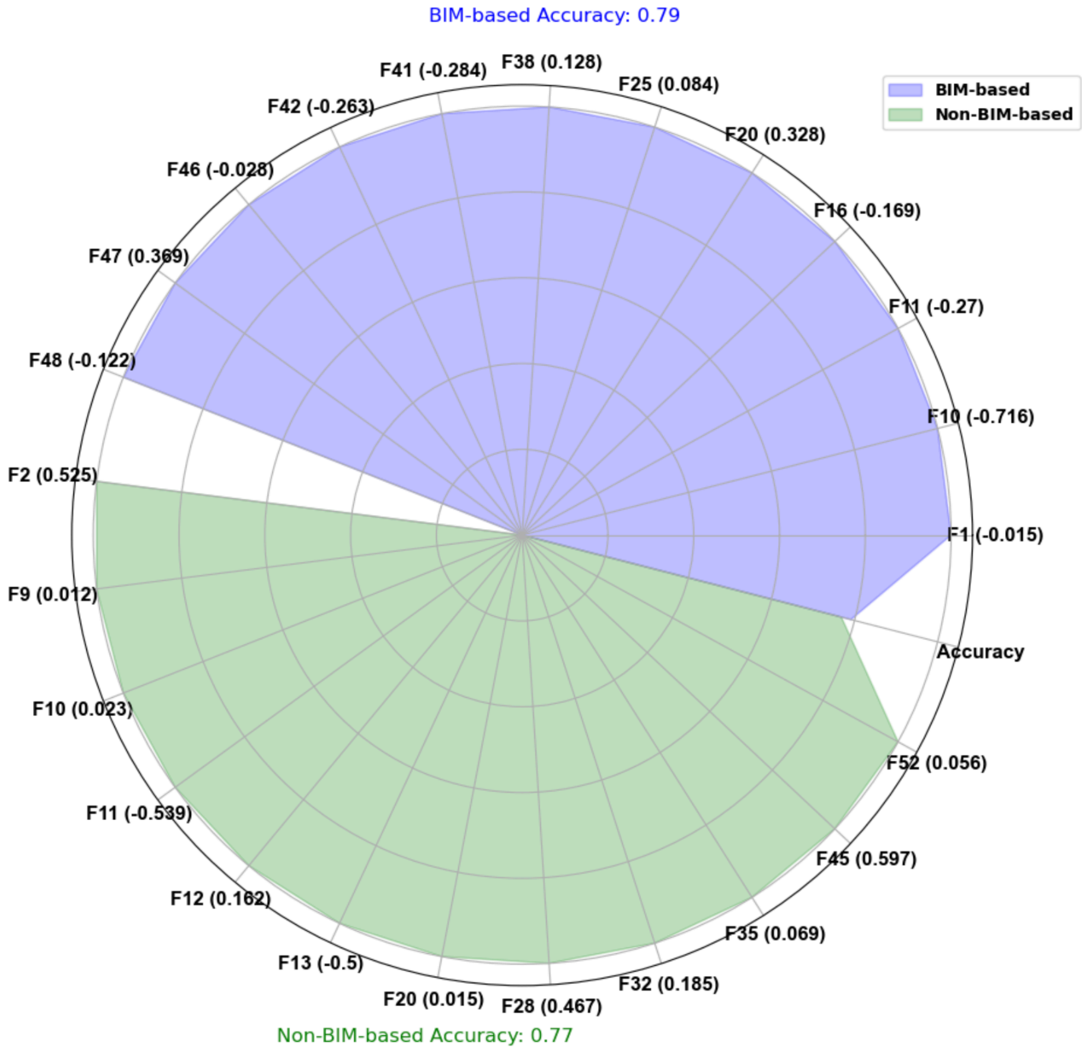

Figure 7.10: Forward Selection Ranking of Features

Backward Elimination

Backward elimination operates within the antithetical framework of wrapper search direction, commencing with the complete feature set and progressively discarding the least contributory features guided by their efficacy (Austin, 2008). Consequently, trivial variables are sequentially pruned, contingent upon the subsequent variance observed in performance metrics upon their exclusion from the feature set. As illustrated in Figure 7.11, the ramifications stemming from the execution of the backward selection methodology are striking, revealing a nuanced interplay between the selected features and the efficacy of a logistic model. The model's performance, set out through its accuracies across various feature subsets, is vividly portrayed on a unified plot, facilitating a comparative analysis between BIM-based and non-BIM-based projects. This iterative process intensifies computationally, particularly with larger initial feature sets. Furthermore, the elimination of feature clusters introduces instability, while allowing for the re-entry of previously removed variables can enhance the final outcomes. In essence, backward elimination serves to refine an approximate feature set obtained from filter methods through a process of tailored subtraction (Foithong, Pinngern and Attachoo, 2012).

However, the staggering dimensions inherent in certain datasets render the initialization with all features impractical, if not impossible. Thus, a judicious selection process becomes imperative, wherein the most pertinent features are identified and retained, while the less informative ones are systematically pruned. This thorough curation of features not only streamlines the computational complexity but also ensures that the final model is more robust and interpretable. Moreover, the iterative nature of backward elimination fosters a deeper understanding of the underlying relationships between features and the target variable, thereby facilitating insights into the dataset's structure and potential avenues for further refinement (Khodadadi *et al.*, 2023). Consequently, while challenging in its execution, backward elimination emerges as a potent technique for feature selection in high-dimensional datasets, offering a nuanced balance between computational efficiency and model performance optimization.

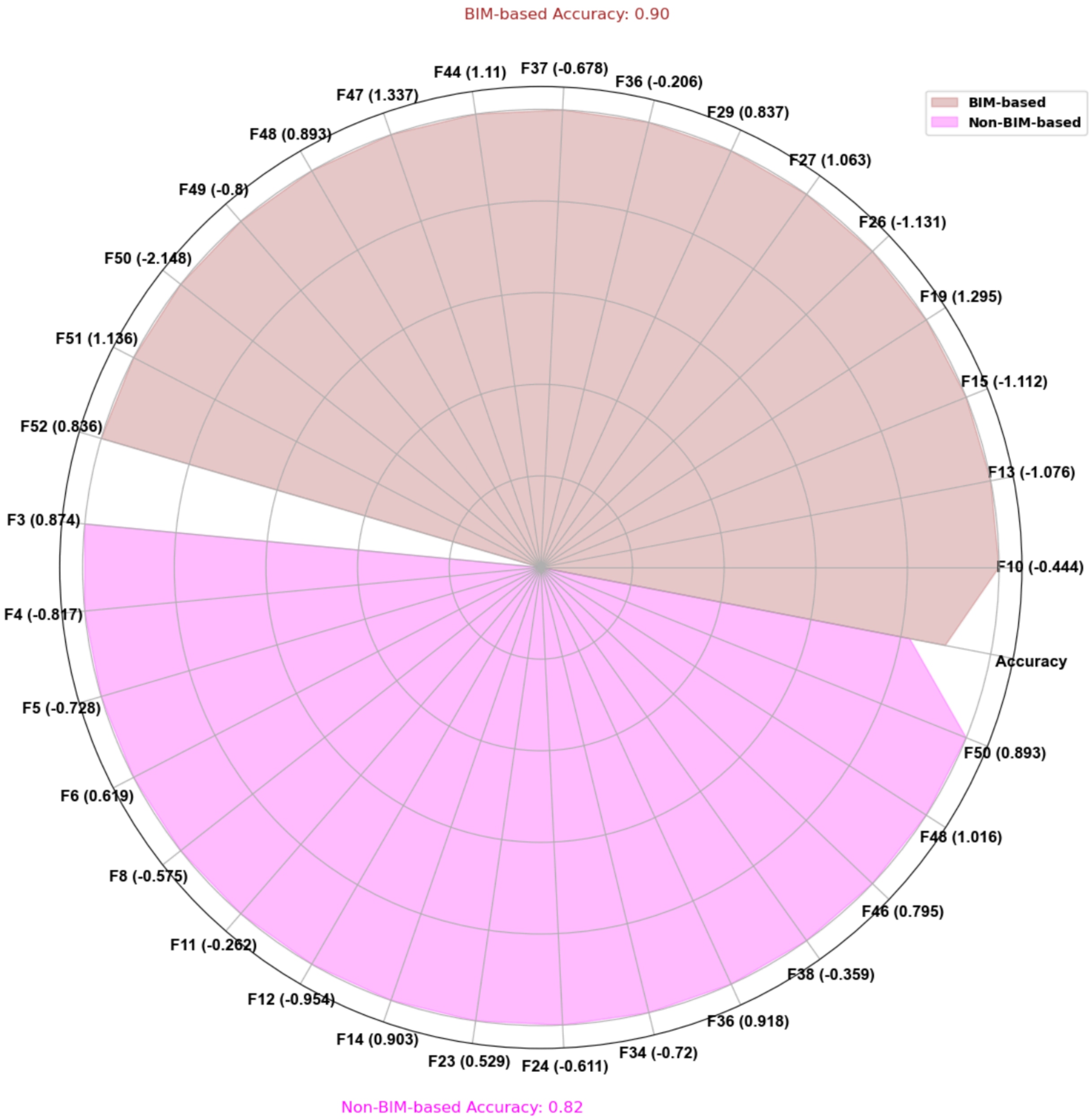

Figure 7.11: Backward Selection Ranking of Features

Exhaustive Feature Selection

Exhaustive feature selection is a wrapper-based technique that conducts an extensive search through all possible combinations of features in a dataset to determine the optimal subset for a given machine learning model (Uncu and Türkşen, 2007). This approach iteratively constructs models using different subsets of features, evaluates their performance, and compares them to select the top-performing subset. More specifically, exhaustive feature selection initializes by training separate models on each individual feature. It then progresses to evaluate dual combinations of features, followed by triple combinations, and so forth, until reaching models encompassing all features in the dataset. At each iteration, the technique records model accuracy metrics for the respective feature subsets. After completing the exhaustive traversal of the feature space, exhaustive selection chooses the subset yielding the best model performance based on the defined scoring metric. Therefore, an exhaustive search guarantees discovering the globally optimal feature subset for a particular model, unlike greedy iterative approaches such as forward selection or backward elimination. However, this comes at a computational cost exponential to the number of features. Accordingly, exhaustive selection remains feasible only for datasets with fewer than 30–40 features. For problems involving higher dimensionality, the exploding search space renders an exhaustive approach intractable. In such cases, an alternative suboptimal method may prove pragmatic (Mnich and Rudnicki, 2020). The most common scoring function for exhaustive feature selection is classification accuracy or AUC for supervised learning problems. However, other metrics like precision, sensitivity, specificity, or model interpretability may also constitute valid scoring schemes based on the project goals. For exhaustive selection, the choice of the underlying machine learning model is open (Deeba *et al.*, 2018). But tree-based ensemble methods like random forests and gradient boosting machines operate efficiently for this wrapping purpose. Their inherent multicollinearity handling ability and transparency of computed feature importance scores enable simplified exhaustive selection wrappers.

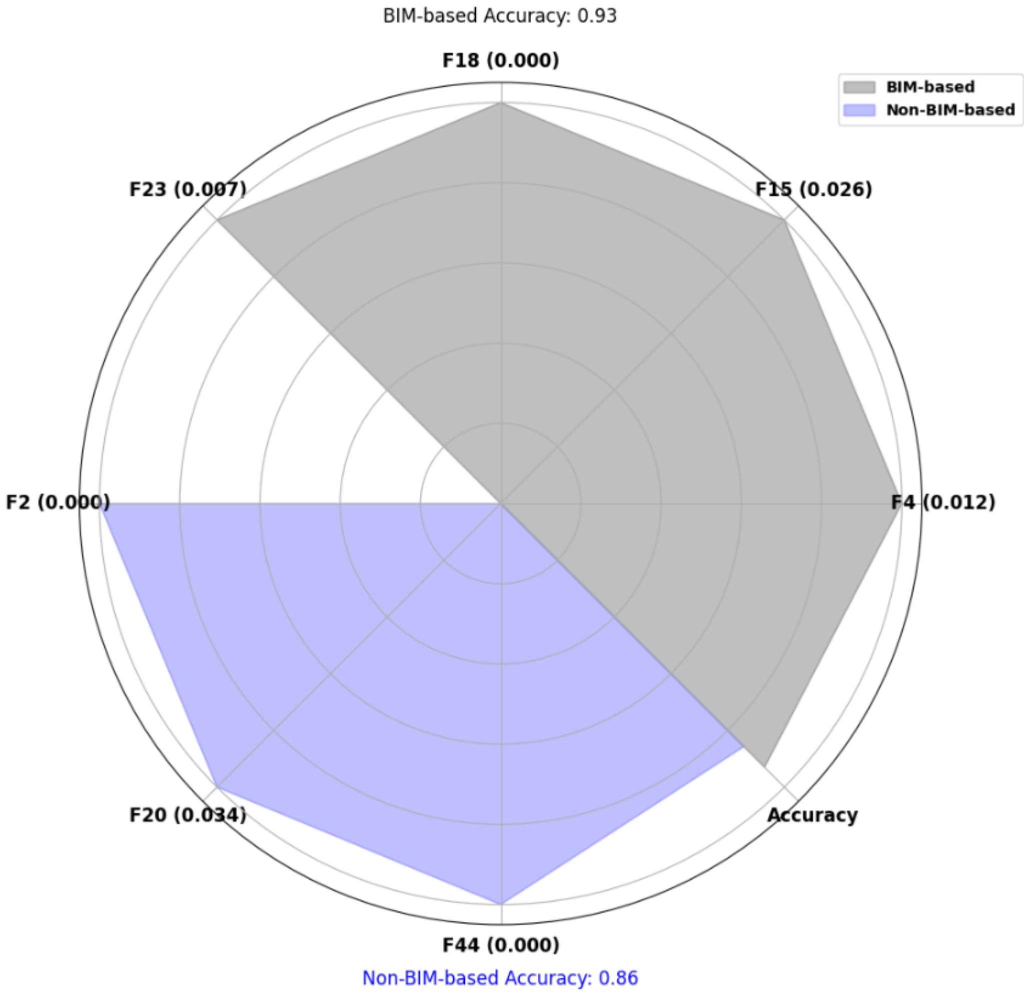

Figure 7.12: Exhaustive Feature Selection Ranking of Features

Recursive Feature Elimination

Recursive feature elimination (RFE) operates as a greedy, iterative wrapper method for feature selection in machine learning (Richhariya, Tanveer and Rashid, 2020). It repeatedly trains a model, ranks feature importance, and eliminates the least useful features until reaching a predefined number. The central notion of RFE relies on fitting models to differing subsets of features, assessing those contributing most to predicting the target, and progressively removing those having little predictive power (Liu and Wang, 2021). Initially, RFE trains the chosen model, ranks all features by importance, and removes those falling under a cut-off threshold. In subsequent iterations, RFE retrains the model on the remaining features, re-ranks their new importance, and again eliminates the least useful ones based on the updated ranking. Most implementations prune away 10-20% of the lowest ranking features per loop. This recursive procedure loops until the desired number of features persists or model metrics plateau. The final optimal subset contains the features still remaining after the successive pruning. Accordingly, unlike exhaustive search exploring all combinations, RFE only samples the space of possible feature subsets guided by model feedback. The incremental greediness provides computational efficiency but risks getting trapped in local optima. Still, for high-dimensional datasets, RFE stands as one of few pragmatically feasible selection routines. And ensemble approaches training multiple RFE iterations in parallel and aggregating their votes for feature rankings enhance robustness (Theerthagiri, 2022). The default feature importance ranker in RFE is the linear coefficient magnitude from logistic regression and SVM models or the Gini importance from tree classifiers. However, dependence on these metrics' biases RFE towards removing noisy but potentially useful features. Alternative scoring schemes assessing multivariate interactions, such as linear correlation, mutual information, or statistical tests can improve stability. Furthermore, combining RFE with complementary filter methods upfront reduces computational overhead. Filter pre-processing retains features meeting univariate statistical thresholds for the target before applying RFE model-based selection.

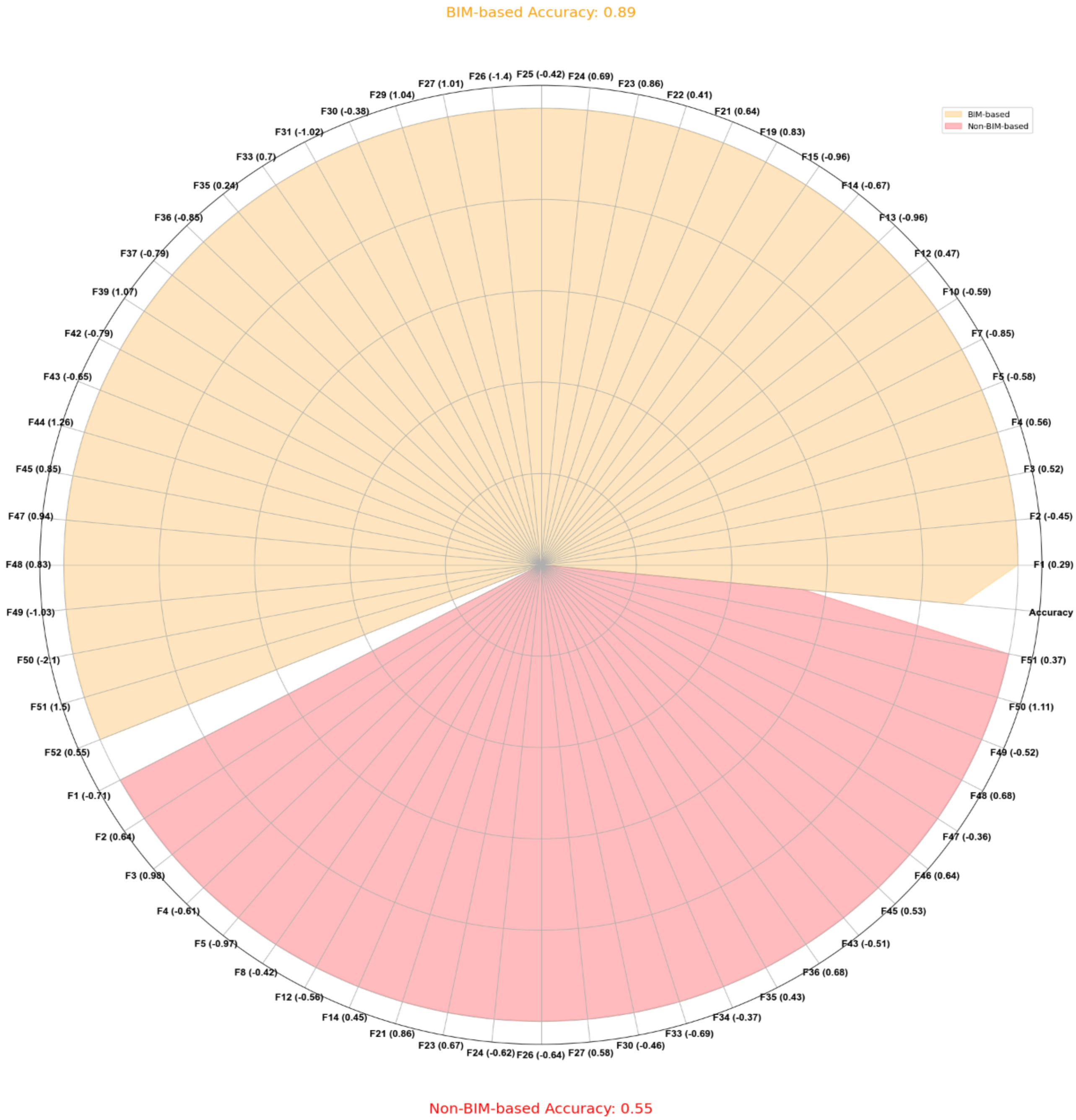

Figure 7.13: Recursive Feature Elimination Ranking of Features

Embedded Methods

Embedded feature selection refers to a set of techniques built into the model construction process that inherently perform feature selection concurrently while training (Liu, Zhou and Liu, 2019). These methods encode automatic feature selection as part of the objective function optimization. Unlike wrapper methods that employ an external model to test subsets of variables, embedded approaches directly encode feature selection into internal model optimization. Mathematically, the objective function consists of two opposing components - the loss function tracking model prediction error on training data, and a penalty term called regularization that encodes complexity control (Imani, Keyvanpour and Azmi, 2013). An overparameterized complex model with excess noisy variables tends to overfit exhibiting high variance. Two main subcategories exist - regularization-based embedding and intrinsic metric-based embedding. The regularization applies constraints to simplify the model preventing overfitting. It provides a knob for trading off between model fit and generalization. The most common form uses L1 norm regularization that penalizes sum of absolute magnitudes of model parameters. By imposing a constraint that limits sum of parameters, the solver has to push some parameters exactly to zero to satisfy the constraint (Theng and Bhoyar, 2024b). This applies aggressive shrinkage pruning away non-informative parameters. Features associated with zeroed out parameters automatically detach from model influence. Thereby, L1 regularization induces inherent feature selection by filtering unimportant variables whose coefficients become zero. Accordingly, by baking regularization directly into the loss, embedded methods accomplish joint optimization simultaneously fitting model and selecting features in one shot training paradigm. The level of sparsity or number of eliminated features depends on the regularization strength coefficient - higher penalty prunes more features. Popular L1 regularized algorithms containing inherent feature selection include lasso regression, elastic net, and sparse logistic models. The efficient solvers optimize the regularization path across full spectrum of penalty settings in mostly automatic black box fashion (Fan *et al.*, 2024). Beyond inducing sparsity in generalized linear models, intrinsic feature importance metrics native to certain model classes also facilitates embedded selection. Decision tree-based ensemble methods like random forest, gradient boosting and XGBoost efficiently calculate variable importance score of each feature. The importance signifies how much removing that feature degrades model accuracy thereby quantifying its predictive contribution. Following greedy backward elimination approach, features with importance below a cut-off threshold get iteratively pruned away until desired model sparsity reaches, or performance slows. Therefore, whether via regularization penalties or intrinsic importance computations, embedded methodologies fuse the processes of model optimization and feature selection instead of discrete wrapping stages. This helps limit overfitting and improves computational efficiency over wrappers. However, they are not guaranteed to achieve globally optimal subset and often get trapped in local solutions based on the embedded heuristics. Regularization approaches may also suffer from high collinearity dropping informative groups of correlated features. Recent innovations in this domain aim to overcome those limitations via global optimization or multivariate penalties.

Regularization-Based Embedding

Regularization-based approaches impose constraints on the objective function to shrink model coefficients towards zero (Armeshi, Sahebi and Aghababaei, 2024). This implicitly filters out features(see Figure 7.14) with little predictive contribution as their coefficients diminish to zero. The most universal variant, L1 regularization, penalizes the absolute sum of coefficients thereby forcing sparse solutions with many zeros. Consequently, features with zeroed coefficients automatically detach from the model. Though all coefficients shrink, only the least useful ones reduce fully to zero. Therefore, L1 regularization manifests as an embedded feature selection procedure. It removes redundant and noisy features without needing a separate selection step. The level of sparsity depends on the regularization strength - higher penalty coefficients induce greater sparsity and more feature elimination (Nokhwal and Kumar, 2023). Multiple optimized implementations of L1 regularized algorithms exist including lasso regression, elastic net, and sparse logistic regression. These models efficiently compute the entire regularization path of solutions for all penalty values. Such embedded regularized techniques garner wide adoption given their joint execution of feature selection and model fitting. They provide insightful, interpretable models that generalize well by removing non-informative variables. However, L1 methods exhibit limitations when features demonstrate high collinearity. Groups of correlated relevant variables often get entirely eliminated in favour of a single representative feature. Therein arises the need for intrinsic methods examining multivariate interactions (Khattab *et al.*, 2020).

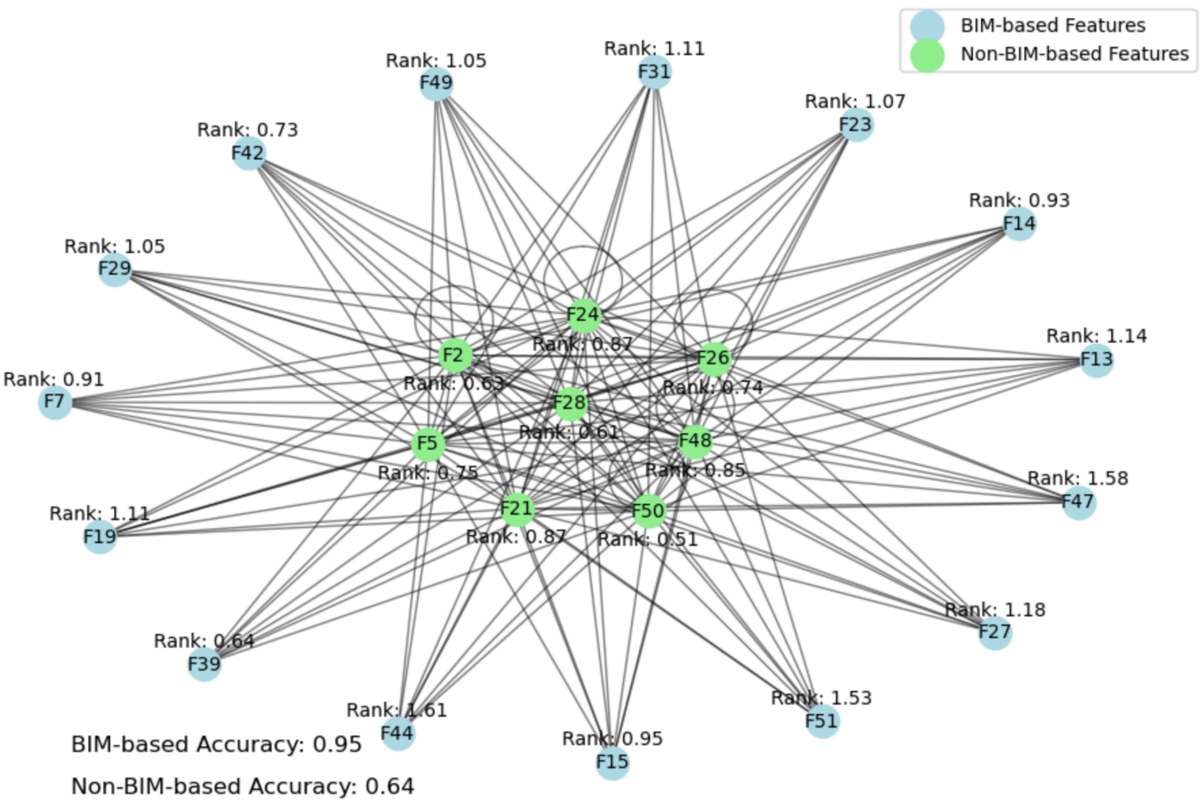

Figure 7.14: Regularization-Based Embedding Ranking of Selected Features

Intrinsic Metric-Based Embedding

This second variety of embedded selection stems from intrinsic variable importance metrics native to certain model classes (Liang *et al.*, 2019). While training complex non-linear models, intrinsic variable ranking procedures determine predictive utility of features based on how removing that feature impacts model performance. The feature importance extractor forms an inbuilt component of the model itself. For tree-based methods including random forest, gradient boosting machines, and XGBoost, the classifiers compute a variable importance score for each feature. Features producing nodes that reduce impurity most influence the predictions and thereby gain higher importance. Such tree intrinsic measures consider higher-order multivariate interactions overlooked by penalization. Embedded intrinsic selection follows a greedy backward elimination routine guided by variable importance (Wang, Wu and Kittler, 2021). First, the model trains on the full feature set, compiles importance, and removes all features (see Figure 7.15) below an importance threshold. Next, the pruned model refits and reassesses new importance on the remaining features to perform subsequent pruning iterations. The procedure repeats until few features stay or model metrics start degrading. Despite not guaranteeing global optima, intrinsic embedded techniques efficiently navigate complex feature spaces while handling nonlinear relationships and collinearity well. They deliver compact, performant models with capped compute times unlike wrappers. Their stability improves further with ensemble aggregation of multiple runs (Yuan and Yang, 2023). However, importance computations remain model-specific with questionable transferability across model families.

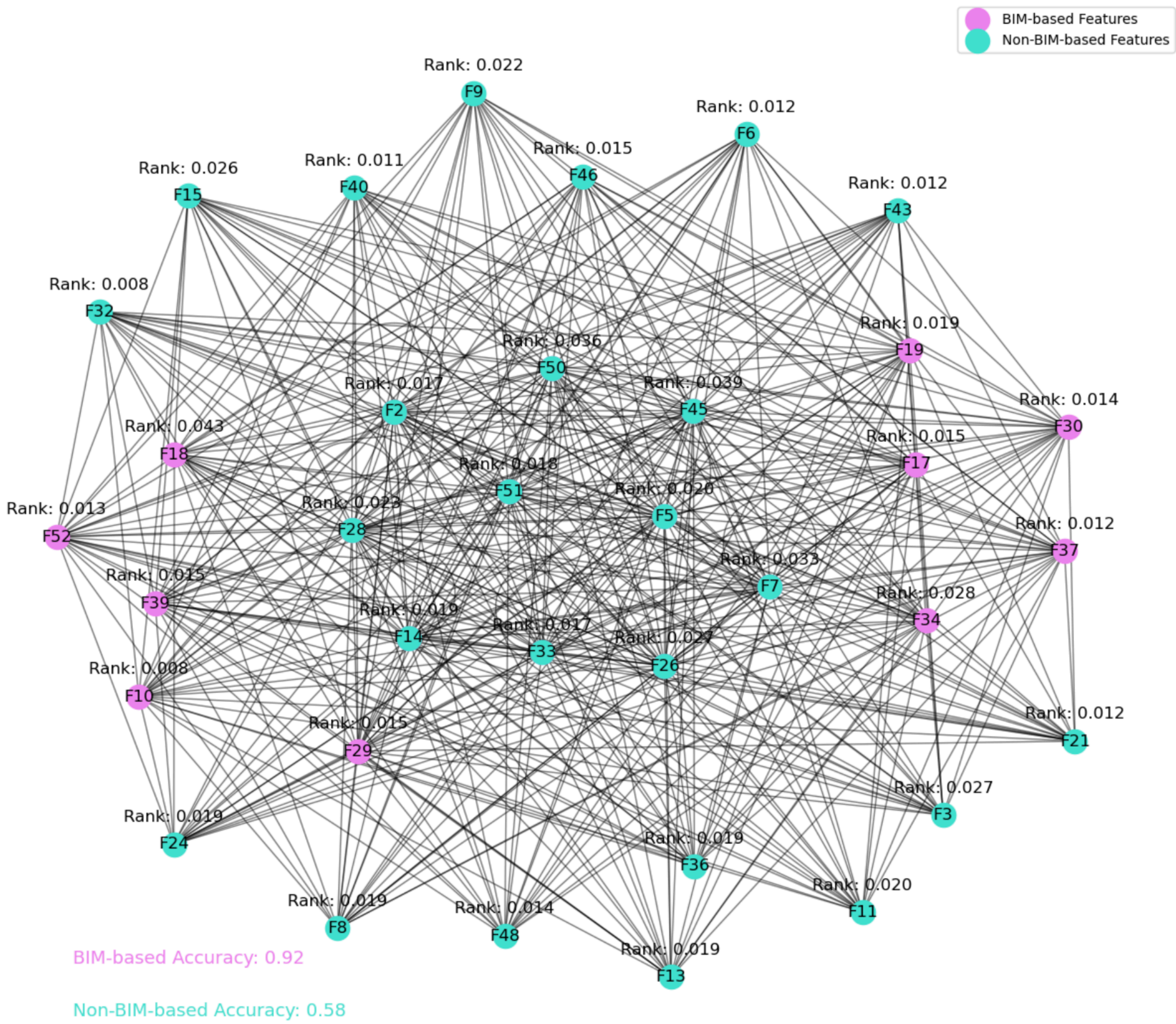

Figure 7.15: Intrinsic Metric-Based Embedding Ranking of Selected Features

7.4.3 Synthesizing the Feature Selection Techniques

Given the contentious nature of the discourse surrounding the optimal feature selection technique, this study embraces a comprehensive approach, employing two distinct ranking systems to evaluate features and ultimately converge on a final selection. It is noteworthy that Tables 7.3 and 7.4 present these ranking systems, while Table 7.5 amalgamates their outcomes, culminating in a refined feature selection for the development of artificial intelligence models, as showcased in Table 7.6. The percentage score ranking system underscores a thorough evaluation process, wherein each independent feature variable is assigned a quantitative score by various feature selection techniques. These scores, albeit diverse in scale across methods (as evidenced in Table 7.2), are harmonized into percentage values within each technique to ensure an equitable contribution. Consequently, the cumulative percentage scores afford a comprehensive basis for feature ranking (Table 7.3), transcending mere ordinal positioning to encapsulate nuanced differentials in feature importance, as elucidated by Lu et al. (2019) in their seminal work on feature selection methodologies. Conversely, the summed ranking system espouses a pragmatic approach, wherein feature variables are individually ranked within each selection method, and subsequently, these rankings are aggregated across all techniques. This summation yields a consolidated ranking, wherein lower values denote a higher precedence of feature variables (see Table 7.4). Notably, this system accords parity to all selection techniques, mitigating potential biases and affording a holistic appraisal of feature significance, in alignment with the principles espoused by Jia et al. (2020) in their influential research on feature ranking strategies. The culmination of these disparate systems coalesces in the final ranking system, wherein the average value of the percentage score ranking and the summed ranking informs a comprehensive re-ranking of feature variables (see Table 7.5). This amalgamation, epitomizing a synthesis of diverse evaluation paradigms, begets a refined hierarchy of feature importance, facilitating the judicious selection of the top twenty feature variables for model development. Notably, this culmination underscores the study's commitment to methodological rigor and empirical robustness, reflective of contemporary best practices in feature selection methodologies, as elucidated by Zhao et al. (2021) in their authoritative treatise on the subject. The column denoted as 'BIM Status' served as a differentiating criterion, facilitating the delineation of ranking scores for the two distinct project typologies that constituted the focus of this empirical investigation. To elucidate further, a BIM Status value of 1 was assigned to represent the respective ranking scores attributed to all feature variables pertaining to projects executed through the implementation of BIM methodologies. Conversely, a BIM Status value of 0 was employed to denote the corresponding ranking scores ascribed to all feature variables associated with projects undertaken through traditional, non-BIM-based approaches. This binary classification system enabled a comprehensive comparative analysis of the ranking profiles exhibited by the feature variables across the two project delivery paradigms under scrutiny. The culmination of this analysis yielded an empirically validated framework elucidating the critical determinants of project delays, a pivotal prerequisite for the development of predictive models aimed at quantifying delay risks across both BIM-based and traditional non-BIM construction projects, as depicted in Figure 7.16. This rigorously substantiated framework serves to corroborate the initial conceptual framework formulated in Section 2, thereby fortifying the theoretical foundations underpinning this research initiative. Additionally, through this analytical process, the salient factors contributing to project delays have been systematically identified and integrated into a coherent validated schema. This validated framework constitutes an indispensable foundation for the formulation of predictive models capable of anticipating and evaluating the potential risks of schedule delays, irrespective of whether the construction project adheres to cutting-edge BIM methodologies or more conventional non-BIM approaches.

Table 7.2: Allocation of Raw Scores to Features Across Selection Techniques

Feature	BIM Status	Information Gain	Chi-square Test	Fisher's Score	Forward Selection	Backward Selection	Exhaustive Feature Selection	Recursive Feature Selection	Regularization	Intrinsic Metric Selection
F1	1	0.23	2.61	20.38	-0.015	0	0	0.29	0	0
	0	0.1	0.03	0.17	0	0	0	-0.71	0	0
F2	1	0.23	8.16	60.03	0	0	0	-0.45	0	0
	0	0	0.34	2.35	0.525	0	0	0.64	0.63	0.017
F3	1	0.18	1.32	8.80	0	0	0	0.52	0	0
	0	0	0.30	1.81	0	0.874	0	0.98	0	0.027
F4	1	0.29	6.23	38.03	0	0	0.012	0.56	0	0
	0	0	0	0.03	0	-0.817	0	-0.61	0	0
F5	1	0.27	7.42	53.27	0	0	0	-0.58	0	0
	0	0.06	0.23	1.69	0	-0.728	0	-0.97	0.75	0.020
F6	1	0.28	3.53	20.40	0	0	0	0	0	0
	0	0.06	0.78	3.64	0	0.619	0	0	0	0.012
F7	1	0.30	6.47	45.66	0	0	0	-0.85	0.91	0
	0	0	0	0.03	0	0	0	0	0	0.033
F8	1	0.26	4.11	24.85	0	0	0	0	0	0

Feature	BIM Status	Information Gain	Chi-square Test	Fisher's Score	Forward Selection	Backward Selection	Exhaustive Feature Selection	Recursive Feature Selection	Regularization	Intrinsic Metric Selection
	0	0.03	0.03	0.15	0	-0.575	0	-0.42	0	0.019
F9	1	0.31	5.13	28.75	0	0	0	0	0	0.008
	0	0.02	0	0	0.012	0	0	0	0	0.022
F10	1	0.22	6.17	41.79	-0.716	-0.444	0	-0.59	0	0
	0	0.02	0.04	0.22	0.023	0	0	0	0	0
F11	1	0.25	3.67	28.75	-0.270	0	0	0	0	0
	0	0.02	0.02	0.16	-0.539	-0.262	0	0	0	0.020
F12	1	0.23	2.18	12.72	0	0	0	0.47	0	0
	0	0	0.16	1.02	0.162	-0.954	0	-.056	0	0
F13	1	0.21	8.28	45.05	0	-1.076	0	-0.96	1.14	0
	0	0.08	0.54	3.77	-0.5	0	0		0	0.019
F14	1	0.30	4.86	33.25	0	0	0	-0.67	0.93	0
	0	0.07	0.27	1.67	0	0.903	0	0.45	0	0.019
F15	1	0.23	7.21	47.71	0	-1.112	0.026	-0.96	0.95	0
	0	0.08	0	0	-0.169	0	0	0	0	0.026
F16	1	0.23	2.39	16.11	0	0	0	0	0	0
	0	0	0.47	2.90	0	0	0	0	0	0

Feature	BIM Status	Information Gain	Chi-square Test	Fisher's Score	Forward Selection	Backward Selection	Exhaustive Feature Selection	Recursive Feature Selection	Regularization	Intrinsic Metric Selection
F17	1	0.28	5.73	36.30	0	0	0	0	0	0.015
	0	0	0.10	0.66	0	0	0	0	0	0
F18	1	0.25	1.95	12.12	0	0	0	0	0	0.043
	0	0.03	0.32	1.71	0	0	0	0	0	0
F19	1	0.16	0.01	0.05	0	1.295	0	0.83	1.11	0.019
	0	0.02	0.38	2.35	0	0	0	0	0	0
F20	1	0.25	0.71	4.62	0.328	0	0.034	0	0	0
	0	0	0.12	0.55	0.015	0	0	0	0	0
F21	1	0.20	2.45	16.77	0	0	0	0.64	0	0
	0	0	0.23	1.34	0	0	0	0.86	0.87	0.012
F22	1	0.28	0.98	7.13	0	0	0	0.41	0	0
	0	0	0.01	0.04	0	0	0	0	0	0
F23	1	0.25	0.84	5.61	0	0	0.007	0.86	1.07	0
	0	0.05	0.72	4.75	0	0.529	0	0.67	0	0
F24	1	0.22	5.17	33.02	0	0	0	0.69	0	0
	0	0	0.19	1.53	0	-0.611	0	-0.62	0.87	0.019
F25	1	0.21	2.12	12.51	0.084	0	0	-0.42	0	0

Feature	BIM Status	Information Gain	Chi-square Test	Fisher’s Score	Forward Selection	Backward Selection	Exhaustive Feature Selection	Recursive Feature Selection	Regularization	Intrinsic Metric Selection
	0	0.05	0.22	1.18	0	0	0	0	0	0
F26	1	0.23	6.98	45.76	0	-1.131	0	-1.4	0	0
	0	0	0	0.01	0	0	0	-0.64	0.74	0.027
F27	1	0.19	0.07	0.37	0	0	0	1.01	1.18	0
	0	0.01	0.13	0.59	0	0	0	0.58	0	0
F28	1	0.24	4.79	27.50	0	0	0	0	0	0
	0	0	0.86	4.73	0.467	0	0	0	0	0.023
F29	*1*	*0.28*	*0.35*	*4.48*	*0*	*0.837*	*0*	*1.04*	*1.05*	*0.015*
	0	*0.02*	*0.07*	*0.32*	*0*	*0*	*0*	*0*	*0*	*0*
F30	1	0.29	2.43	17.05	0	0	0	-0.38	0	0.014
	0	0	0.01	0.08	0	0	0	-0.46	0	0
F31	1	0.22	6.03	38.41	0	0	0	-1.02	1.11	0
	0	0	0.03	0.23	0	0	0	0	0	0
F32	1	0.22	3.35	21.64	0	0	0	0	0	0
	0	0.06	0.22	1.43	0.185	0	0	0	0	0.008
F33	1	0.21	1.98	11.15	0	0	0	0.7	0	0
	0	0.05	0.10	0.83	0	0	0	-0.69	0	0.027

Feature	BIM Status	Information Gain	Chi-square Test	Fisher's Score	Forward Selection	Backward Selection	Exhaustive Feature Selection	Recursive Feature Selection	Regularization	Intrinsic Metric Selection
F34	1	0.26	2.81	19.37	0	0	0	0	0	0.028
	0	0	0	0	0	-0.72	0	-0.37	0	0
F35	1	0.21	3.13	21.63	0	0	0	0.24	0	0
	0	0	0.10	0.80	0.069	0	0	0.43	0	0
F36	1	0.26	4.07	23.98	0	-0.206	0	-0.85	0	0
	0	0	0.26	1.56	0	0.918	0	0.68	0	
F37	1	0.30	7.05	46.08	0	-0.678	0	-0.79	0	0.012
	0	0.04	0.31	2.42	0	0	0	0	0	0
F38	1	0.29	2.45	17.87	0.128	0	0	0	0	0
	0	0	0.17	1.19	0	-0.359	0	0	0	
F39	1	0.25	2.33	15.90	0	0	0	1.07	0.64	0.015
	0	0.10	0.51	3.57	0	0	0	0	0	0
F40	1	0.25	4.45	29.07	0	0	0	0	0	0
	0	0.02	0.32	2.20	0	0	0	0	0	0.011
F41	1	0.23	4.80	26.40	-0.284	0	0	0	0	0
	0	0	0.39	2.97	0	0	0	0	0	0
F42	1	0.29	4.45	24.85	-0.263	0	0	-0.79	0.73	0

Feature	BIM Status	Information Gain	Chi-square Test	Fisher's Score	Forward Selection	Backward Selection	Exhaustive Feature Selection	Recursive Feature Selection	Regularization	Intrinsic Metric Selection
	0	0	0.03	0.24	0	0	0	0	0	0
F43	1	0.25	7.02	41.71	0	0	0	-0.65	0	0
	0	0	0.01	0.05	0		0	-0.51	0	0
F44	1	0.23	0.21	1.30	0	1.11	0	1.26	1.61	0
	0	0	0.31	1.39	0.597	0	0	0	0	0.012
F45	1	0.27	1.21	8.28	0	0	0	0.85	0	0
	0	0.04	1.32	5.94	0	0	0	0.53	0	0.039
F46	*1*	*0.25*	*0.30*	*2.42*	*-0.028*	*0*	*0*	*0*	*0*	*0*
	0	*0*	*0.29*	*1.24*	*0*	*0.795*	*0*	*0.64*	*0*	*0.015*
F47	1	0.23	0.67	4.57	0.369	1.337	0	0.94	1.58	0
	0	0	0.03	0.18			0	-0.36	0	0
F48	1	0.20	2.40	15.19	-0.122	0.893	0	0.83	0	0
	0	0.08	1.42	9.67	0	1.016	0	0.68	0.85	0.014
F49	1	0.28	4.73	26.28	0	-0.8	0	-1.03	1.05	0
	0	0.03	0	0.02	0	0	0	-0.52	0	0
F50	1	0.31	8.84	54.33	0	-2.148	0	-2.1	0	0
	0	0.10	1.26	8.69	0	0.893	0	1.11	0.51	0.036

Feature	BIM Status	Information Gain	Chi-square Test	Fisher's Score	Forward Selection	Backward Selection	Exhaustive Feature Selection	Recursive Feature Selection	Regularization	Intrinsic Metric Selection
F51	1	0.25	0.17	1.06	0	1.136	0	1.5	1.53	0
	0	0	0.62	4.56	0	0	0	0.37	0	0
F52	*1*	*0.23*	*0.38*	*4.40*	*0*	*0.836*	*0.001*	*0.55*	*0*	*0.013*
	0	*0.09*	*0.12*	*0.61*	*0.056*	*0*	*0*	*0*	*0*	*0.018*
Total	*1/0*	*12.84/1.33*								

Table 7.3: Ranking Determined by the Percentage Score of Features within Each Selection Method

Feature	BIM Status	Information Gain	Chi-square Test	Fisher's Score	Forward Selection	Backward Selection	Exhaustive Selection	Recursive Selection	Regularization	Intrinsic Metric Selection	Total Percentage	Ranking
F1	1	1.791	1.395	1.692	1.901	0	0	37.662	0	0	44.441	19th
	0	7.519	0.229	0.190	0	0	0	-60.169	0	0	-52.231	46th
F2	1	1.791	4.360	4.983	0	0	0	-58.441	0	0	-47.307	39th
	0	0	2.591	2.633	58.140	0	0	54.237	12.069	3.424	133.094	4th
F3	1	1.402	0.705	0.730	0	0	0	67.532	0	0	70.369	16th
	0	0	2.287	2.028	0	57.4622	0	83.051	0	5.455	150.283	3rd
F4	1	2.259	3.329	3.157	0	0	0.150	72.727	0	0	81.622	15th
	0	0	0	0.034	0	-53.715	0	-51.695	0	0	-105.376	52nd
F5	1	2.103	3.965	4.422	0	0	0	-75.325	0	0	-64.835	41st
	0	4.511	1.753	1.894	0	-47.863	0	-82.203	14.368	4.040	-103.5	51st
F6	1	2.181	1.881	1.693	0	0	0	0	0	0	5.755	32nd
	0	4.511	5.945	4.079	0	0.407	0	0	0	2.424	17.366	19th
F7	1	2.336	3.457	3.790	0	0	0	-110.390	5.485	0	-95.322	52nd
	0	0	0	0.034	0	0	0	0	0	6.667	6.701	28th
F8	1	2.025	2.196	2.063	0	0	0	0	0	0	6.284	30th
	0	2.256	0.229	0.168	0	-37.804	0	-35.593	0	3.838	-66.906	48th

Feature	BIM Status	Information Gain	Chi-square Test	Fisher's Score	Forward Selection	Backward Selection	Exhaustive Selection	Recursive Selection	Regularization	Intrinsic Metric Selection	Total Percentage	Ranking
F9	1	2.414	2.741	2.386	0	0	0	0	0	4.396	11.937	27th
	0	1.504	0	0	1.329	0	0	0	0	4.444	7.277	24th
F10	1	1.714	3.297	3.469	90.748	294.040	0	-76.623	0	0	316.645	7th
	0	1.504	0.305	0.247	2.547	0	0	0	0	0	4.603	31st
F11	1	1.947	1.961	2.386	34.221	0	0	0	0	0	40.515	21st
	0	1.504	0.152	0.178	-59.690	-9.928	0	0	0	4.040	-63.744	47th
F12	1	1.791	1.165	1.056	0	0	0	61.039	0	0	65.051	17th
	0	0	1.220	1.143	17.940	-0.954	0	-47.458	0	0	-28.109	40th
F13	1	1.636	4.424	3.739	0	712.583	0	-124.675	6.872	0	604.579	3rd
	0	6.015	4.116	4.225	-55.371	0	0	0	0	3.838	-37.177	43rd
F14	1	2.336	2.597	2.760	0	0	0	-87.013	5.606	0	-73.714	42nd
	0	5.263	2.058	1.871	0	59.369	0	38.136	0	3.838	110.535	7th
F15	1	1.791	3.853	3.960	0	736.424	0.325	-124.675	5.726	0	627.404	2nd
	0	6.015	0	0	-18.715	0	0	0	0	5.253	-7.447	38th
F16	1	1.791	1.277	1.337	0	0	0	0	0	0	4.405	34th
	0	0	3.582	3.259	0	0	0	0	0	0	6.841	26th
F17	1	2.181	3.062	3.013	0	0	0	0	0	8.242	16.498	26th

Feature	BIM Status	Information Gain	Chi-square Test	Fisher's Score	Forward Selection	Backward Selection	Exhaustive Selection	Recursive Selection	Regularization	Intrinsic Metric Selection	Total Percentage	Ranking
	0	0	0.762	0.740	0	0	0	0	0	0	1.502	34th
F18	1	1.950	1.042	1.006	0	0	0	0	0	23.626	27.624	24th
	0	2.256	2.439	1.916	0	0	0	0	0	0	6.611	29th
F19	1	1.246	0.005	0.056	0	-857.616	0	107.792	6.691	10.440	-731.386	50th
	0	1.504	2.896	2.633	0	0	0	0	0	0	7.033	25th
F20	1	1.947	0.379	0.383	-41.572	0	0.425	0	0	0	-38.438	38th
	0	0	0.915	0.616	1.661	0	0	0	0	0	3.192	32nd
F21	1	1.558	1.309	1.637	0	0	0	83.117	0	0	87.621	14th
	0	0	1.753	1.502	0	0	0	72.881	16.667	2.424	95.227	9th
F22	1	2.181	0.524	0.696	0	0	0	53.247	0	0	56.648	18th
	0	0	0.076	0.045	0	0	0	0	0	0	0.121	37th
F23	1	1.947	0.449	0.547	0	0	0.088	111.688	6.450	0	121.169	10th
	0	3.759	5.488	5.323	0	34.780	0	56.780	0	0	106.13	8th
F24	1	1.713	2.763	3.222	0	0	0	89.610	0	0	97.308	12th
	0	0	1.448	1.714	0	-40.171	0	-52.542	16.667	3.838	-69.046	49th
F25	1	1.636	1.133	1.221	-10.646	0	0	-54.545	0	0	-61.201	40th
	0	3.759	1.677	1.322	0	0	0	0	0	0	6.758	27th

Feature	BIM Status	Information Gain	Chi-square Test	Fisher's Score	Forward Selection	Backward Selection	Exhaustive Selection	Recursive Selection	Regularization	Intrinsic Metric Selection	Total Percentage	Ranking
F26	1	1.791	3.730	4.466	0	749.007	0	-181.818	0	0	577.176	4th
	0	0	0	0.011	0	0	0	-54.237	14.176	5.455	-34.595	42nd
F27	1	1.480	0.037	0.036	0	0	0	131.169	7.113	0	139.835	9th
	0	0.752	0.991	0.661	0	0	0	49.153	0	0	51.557	13th
F28	1	1.869	2.560	2.684	0	0	0	0	0	0	7.113	29th
	0	0	6.555	5.300	51.717	0	0	0	0	4.646	68.218	12th
F29	*1*	*2.181*	*0.187*	*0.437*	*0*	*-553.305*	*0*	*135.065*	*6.329*	*8.242*	*-400.864*	*45th*
	0	*1.504*	*0.534*	*0.359*	*0*	*0*	*0*	*0*	*0*	*0*	*2.397*	*33rd*
F30	1	2.259	1.298	1.664	0	0	0	-49.351	0	7.692	-36.438	37th
	0	0	0.080	0.090	0	0	0	-38.983	0	0	-38.813	44th
F31	1	1.714	3.222	3.748	0	0	0	-132.468	6.691	0	-117.093	44th
	0	0	0.229	0.258	0	0	0	0	0	0	0.487	36th
F32	1	1.714	1.790	2.112	0	0	0	0	0	0	5.616	33rd
	0	4.511	1.677	1.602	20.487	0	0	0	0	1.616	29.893	17th
F33	1	1.636	1.058	1.088	0	0	0	90.909	0	0	94.691	13th
	0	3.760	0.762	0.930	0	0	0	-0.585	0	5.455	10.322	21st
F34	1	2.025	1.502	1.890	0	0	0	0	0	15.385	20.802	25th

Feature	BIM Status	Information Gain	Chi-square Test	Fisher's Score	Forward Selection	Backward Selection	Exhaustive Selection	Recursive Selection	Regularization	Intrinsic Metric Selection	Total Percentage	Ranking
	0	0	0	0	0	-47.337	0	-31.356	0	0	-78.693	50th
F35	1	1.636	1.673	2.222	0	0	0	31.169	0	0	36.700	22nd
	0	0	0.762	0.896	7.641	0	0	36.441	0	0	45.740	14th
F36	1	2.025	2.175	2.340	0	136.424	0	-110.390	0	0	32.574	23rd
	0	0	1.982	1.748	0	60.355	0	57.627	0	0	121.712	5th
F37	1	2.336	3.767	4.497	0	449.006	0	-102.597	0	6.593	363.602	6th
	0	3.008	2.363	2.712	0	0	0	0	0	0	8.083	23rd
F38	1	2.259	1.310	1.744	-16.223	0	0	0	0	0	-10.91	35th
	0	0	1.296	1.333	0	-23.603	0	0	0	0	-20.974	39th
F39	1	1.947	1.245	1.552	0	0	0	138.961	3.858	8.242	155.805	8th
	0	7.519	3.887	4.000	0	0	0	0	0	0	15.406	20th
F40	1	1.947	2.378	2.837	0	0	0	0	0	0	7.162	28th
	0	1.504	2.439	2.465	0	0	0	0	0	2.222	8.63	22nd
F41	1	1.791	2.565	2.576	35.995	0	0	0	0	0	42.927	20th
	0	0	2.973	3.328	0	0	0	0	0	0	6.301	30th
F42	1	2.259	2.378	2.425	33.333	0	0	-66.949	4.400	0	-22.154	36th
	0	0	0.229	0.269	0	0	0	0	0	0	0.498	35th

Feature	BIM Status	Information Gain	Chi-square Test	Fisher's Score	Forward Selection	Backward Selection	Exhaustive Selection	Recursive Selection	Regularization	Intrinsic Metric Selection	Total Percentage	Ranking
F43	1	1.947	3.751	4.070	0	0	0	-84.416	0	0	-74.648	43rd
	0	0	0.076	0.056	0	0	0	43.220	0	0	43.352	15th
F44	1	1.791	0.112	0.127	0	-735.099	0	163.636	9.705	0	-559.728	49th
	0	0	2.363	1.558	66.113	0	0	0	0	2.424	72.458	10th
F45	1	2.103	0.647	0.808	0	0	0	110.390	0	0	113.948	11th
	0	3.008	9.984	6.656	0	0	0	44.915	0	7.879	72.442	11th
F46	*1*	*1.947*	*0.160*	*0.236*	*3.549*	*0*	*0*	*0*	*0*	*0*	*5.892*	*31st*
	0	*0*	*2.210*	*1.390*	*0*	*52.268*	*0*	*54.237*	*0*	*3.030*	*113.135*	*6th*
F47	1	1.791	0.358	0.446	-46.768	-885.430	0	122.078	9.524	0	-798.001	51st
	0	0	0.229	0.202	0	0	0	-30.508	0	0	-30.077	41st
F48	1	1.558	1.282	1.482	15.463	-591.391	0	107.792	0	0	-463.814	46th
	0	6.015	10.823	10.836	0	66.798	0	57.627	16.284	2.828	171.211	2nd
F49	1	2.181	2.528	2.565	0	529.801	0	-133.766	6.329	0	409.638	5th
	0	2.256	0	0.022	0	0	0	-44.068	0	0	-41.79	45th
F50	1	2.414	4.724	5.302	0	1422.517	0	-272.727	0	0	1162.23	1St
	0	7.519	9.604	9.738	0	58.711	0	94.068	9.770	7.272	196.682	1st
F51	1	1.947	0.090	0.103	0	-752.318	0	194.805	9.222	0	-546.151	48th

Feature	BIM Status	Information Gain	Chi-square Test	Fisher's Score	Forward Selection	Backward Selection	Exhaustive Selection	Recursive Selection	Regularization	Intrinsic Metric Selection	Total Percentage	Ranking
	0	0	4.726	5.110	0	0	0	31.356	0	0	41.192	16th
F52	*1*	1.791	*0.203*	*0.429*	*0*	*-553.642*	*0.013*	*71.429*	*0*	*7.143*	*-472.634*	*47th*
	0	*6.767*	*0.915*	*0.684*	*6.202*	*0*	*0*	*0*	*0*	*3.636*	*18.204*	*18th*

Table 7.4: Ranking Determined by the Summed Ranking of Features within Each Selection Method

Feature	BIM Status	Information Gain	Chi-square Test	Fisher's Score	Forward Selection	Backward Selection	Exhaustive Selection	Recursive Selection	Regularization	Intrinsic Metric Selection	Sum of all positions	Summed Ranking
F1	1	9	28	28	5	17	2	18	14	9	130	20th
	0	1	30	40	13	17	4	23	8	17	153	32nd
F2	1	9	3	1	13	17	2	22	14	9	90	2nd
	0	11	13	13	2	17	4	6	5	11	82	7th
F3	1	13	38	40	13	17	2	15	14	9	161	37th
	0	11	16	16	13	5	4	2	8	4	79	5th
F4	1	3	10	12	13	17	2	13	14	9	93	4th
	0	11	34	46	13	15	4	19	8	17	167	42nd
F5	1	5	4	3	13	17	2	23	14	9	90	2nd
	0	5	20	18	13	14	4	24	3	8	109	16th
F6	1	4	24	27	13	17	2	35	14	9	145	30th
	0	5	5	8	13	7	4	25	8	14	89	9th
F7	1	2	9	7	13	17	2	28	11	9	98	6th
	0	11	34	46	13	17	4	25	8	3	161	36th
F8	1	6	21	22	13	17	2	35	14	9	139	26th
	0	8	30	42	13	11	4	21	8	9	146	28th

Feature	BIM Status	Information Gain	Chi-square Test	Fisher's Score	Forward Selection	Backward Selection	Exhaustive Selection	Recursive Selection	Regularization	Intrinsic Metric Selection	Sum of all positions	Summed Ranking
F9	1	1	15	18	13	17	2	35	14	8	123	16th
	0	9	34	50	10	17	4	25	8	7	164	39th
F10	1	10	11	9	12	10	2	24	14	9	101	7th
	0	9	39	38	8	17	4	25	8	17	165	40th
F11	1	7	23	17	10	17	2	35	14	9	134	22nd
	0	9	31	41	12	9	4	25	8	8	147	29th
F12	1	9	34	36	13	17	2	16	14	9	150	33rd
	0	11	24	28	5	16	4	18	8	17	131	25th
F13	1	11	2	8	13	13	2	30	5	9	93	4th
	0	3	8	7	11	17	4	25	8	9	92	10th
F14	1	2	16	14	13	17	2	26	10	9	109	10th
	0	4	18	19	13	3	4	9	8	9	87	8th
F15	1	9	5	4	13	14	1	30	9	9	94	5th
	0	3	34	50	8	17	4	25	8	5	154	33rd
F16	1	9	32	33	13	17	2	35	14	9	164	38th
	0	11	10	11	13	17	4	25	8	17	116	18th
F17	1	4	13	13	13	17	2	35	14	4	115	12th

Feature	BIM Status	Information Gain	Chi-square Test	Fisher's Score	Forward Selection	Backward Selection	Exhaustive Selection	Recursive Selection	Regularization	Intrinsic Metric Selection	Sum of all positions	Summed Ranking
	0	11	27	31	13	17	4	25	8	17	153	32nd
F18	1	7	37	38	13	17	2	35	14	1	164	38th
	0	8	14	17	13	17	4	25	8	17	123	23rd
F19	1	14	50	52	13	2	2	9	6	3	151	34th
	0	9	12	14	13	17	4	25	8	17	119	20th
F20	1	7	42	44	2	17	2	35	14	9	172	39th
	0	11	26	34	9	17	1	25	8	17	148	30th
F21	1	12	29	32	13	17	2	12	14	9	140	27th
	0	11	20	24	13	17	4	3	1	14	107	15th
F22	1	4	40	42	13	17	2	17	14	9	158	36th
	0	11	33	45	13	17	4	25	8	17	173	43rd
F23	1	7	41	43	13	17	3	7	7	9	147	31st
	0	6	6	4	13	8	4	5	8	17	71	4th
F24	1	10	14	15	13	17	2	11	14	9	105	9th
	0	11	22	21	13	12	4	20	1	9	113	17th
F25	1	11	35	37	4	17	2	14	14	9	143	29th
	0	6	21	27	13	17	4	25	8	17	138	28th

Feature	BIM Status	Information Gain	Chi-square Test	Fisher's Score	Forward Selection	Backward Selection	Exhaustive Selection	Recursive Selection	Regularization	Intrinsic Metric Selection	Sum of all positions	Summed Ranking
F26	1	9	8	6	13	15	2	33	14	9	109	10th
	0	11	34	49	13	17	4	21	4	4	157	35th
F27	1	13	49	51	13	5	2	5	4	9	151	34th
	0	10	25	33	13	17	4	7	8	17	134	26th
F28	1	8	18	19	13	17	2	35	14	9	135	23rd
	0	11	4	5	3	17	4	25	6	6	81	6th
F29	*1*	*4*	*45*	*46*	*13*	*7*	*2*	*4*	*8*	*4*	*133*	*21st*
	0	*9*	*28*	*35*	*13*	*17*	*4*	*25*	*8*	*17*	*156*	*34th*
F30	1	3	30	31	13	17	2	20	14	5	135	23rd
	0	11	33	43	13	17	4	15	8	17	161	36th
F31	1	10	12	11	13	17	2	31	6	9	111	11th
	0	11	30	37	13	17	4	25	8	17	162	37th
F32	1	10	25	25	13	17	2	35	14	9	150	33rd
	0	5	21	22	4	17	4	25	8	16	122	22nd
F33	1	11	36	39	13	17	2	10	14	9	151	34th
	0	6	27	29	13	17	4	22	8	4	130	24th
F34	1	6	27	29	13	17	2	35	14	2	145	30th

Feature	BIM Status	Information Gain	Chi-square Test	Fisher's Score	Forward Selection	Backward Selection	Exhaustive Selection	Recursive Selection	Regularization	Intrinsic Metric Selection	Sum of all positions	Summed Ranking
	0	11	34	50	13	13	4	13	8	17	163	38th
F35	1	11	26	26	13	17	2	19	14	9	137	25th
	0	11	27	30	6	17	4	10	8	17	130	24th
F36	1	6	22	24	13	9	2	29	14	9	128	19th
	0	11	19	20	13	2	4	4	8	17	98	12th
F37	1	2	6	5	13	11	2	27	14	7	87	1st
	0	7	15	12	13	17	4	25	8	17	118	19th
F38	1	3	29	30	3	17	2	35	14	9	142	28th
	0	11	23	26	13	10	4	25	8	17	137	27th
F39	1	7	33	34	13	17	2	3	13	4	126	18th
	0	1	9	9	13	17	4	25	8	17	103	14th
F40	1	7	20	16	13	17	2	35	14	9	133	21st
	0	9	14	15	13	17	4	25	8	15	120	21st
F41	1	9	17	20	11	17	2	35	14	9	134	22nd
	0	11	11	10	13	17	4	25	8	17	116	18th
F42	1	3	20	23	9	17	2	27	12	9	122	15th
	0	11	30	36	13	17	4	25	8	17	161	36th

Feature	BIM Status	Information Gain	Chi-square Test	Fisher's Score	Forward Selection	Backward Selection	Exhaustive Selection	Recursive Selection	Regularization	Intrinsic Metric Selection	Sum of all positions	Summed Ranking
F43	1	7	7	10	13	17	2	25	14	9	104	8th
	0	11	33	44	13	17	4	16	8	17	163	38th
F44	1	9	47	49	13	4	2	2	1	9	136	24th
	0	11	15	23	1	17	4	25	8	14	118	19th
F45	1	5	39	41	13	17	2	8	14	9	148	32nd
	0	7	2	3	13	17	4	8	8	1	63	3rd
F46	*1*	*7*	*46*	*48*	*6*	*17*	*2*	*35*	*14*	*9*	*184*	*40th*
	0	*11*	*17*	*25*	*13*	*6*	*4*	*6*	*8*	*12*	*102*	*13th*
F47	1	9	43	45	1	1	2	6	2	9	118	13th
	0	11	30	39	13	17	4	12	8	17	151	31st
F48	1	12	31	35	7	6	2	9	14	9	125	17th
	0	3	1	1	13	1	4	4	2	13	42	2nd
F49	1	4	19	21	13	12	2	32	8	9	120	14th
	0	8	34	48	13	17	4	17	8	17	166	41st
F50	1	1	1	2	13	16	2	34	14	9	92	3rd
	0	1	3	2	13	4	4	1	7	2	37	1st
F51	1	7	48	50	13	3	2	1	3	9	136	24th

Feature	BIM Status	Information Gain	Chi-square Test	Fisher's Score	Forward Selection	Backward Selection	Exhaustive Selection	Recursive Selection	Regularization	Intrinsic Metric Selection	Sum of all positions	Summed Ranking
	0	11	7	6	13	17	4	11	8	17	94	11th
F52	*1*	*9*	*44*	*47*	*13*	*8*	*2*	*14*	*14*	*6*	*157*	*35th*
	0	*2*	*26*	*32*	*7*	*17*	*4*	*25*	*8*	*10*	*131*	*25th*

Table 7.5: Ranking Determined by the Mean of the Percentage and Summed Ranking System

Feature	BIM Status	Percentage Score Ranking	Summed Ranking	Mean of the Percentage and Summed Ranking	Ranking Derived from the Mean Values
F1	1	19	20	19.5	8th
	0	46	32	39	37th
F2	1	39	2	20.5	9th
	0	4	7	5.5	4th
F3	1	16	37	26.5	17th
	0	3	5	4	3rd
F4	1	15	4	9.5	4th
	0	52	42	47	41st
F5	1	41	2	21.5	11th
	0	51	16	33.5	31st
F6	1	32	30	31	23rd
	0	19	9	14	13th
F7	1	52	6	29	21st
	0	28	36	32	28th
F8	1	30	26	28	20th
	0	48	28	38	35th
F9	1	27	16	21.5	11th
	0	24	39	31.5	27th
F10	1	7	7	7	3rd
	0	31	40	35.5	32nd
F11	1	21	22	21.5	11th
	0	47	29	38	35th
F12	1	17	33	25	14th
	0	40	25	32.5	29th
F13	1	3	4	3.5	2nd
	0	43	10	26.5	24th

Feature	BIM Status	Percentage Score Ranking	Summed Ranking	Mean of the Percentage and Summed Ranking	Ranking Derived from the Mean Values
F14	1	42	10	26	16th
	0	7	8	7.5	7th
F15	1	2	5	3.5	2nd
	0	38	33	35.5	32nd
F16	1	34	38	36	29th
	0	26	18	22	20th
F17	1	26	12	19	7th
	0	34	32	33	30th
F18	1	24	38	31	23rd
	0	29	23	26	23rd
F19	1	50	34	42	33rd
	0	25	20	22.5	21st
F20	1	38	39	38.5	31st
	0	32	30	31	26th
F21	1	14	27	20.5	9th
	0	9	15	12	11th
F22	1	18	36	27	18th
	0	37	43	40	38th
F23	1	10	31	20.5	9th
	0	8	4	6	5th
F24	1	12	9	10.5	5th
	0	49	17	33	30th
F25	1	40	29	34.5	27th
	0	27	28	27.5	25th
F26	1	4	10	7	3rd
	0	42	35	38.5	36th
F27	1	9	34	21.5	11th

Feature	BIM Status	Percentage Score Ranking	Summed Ranking	Mean of the Percentage and Summed Ranking	Ranking Derived from the Mean Values
	0	13	26	19.5	17th
F28	1	29	23	26	15th
	0	12	6	9	9th
F29	*1*	*45*	*21*	33	*26th*
	0	*33*	*34*	33.5	*31st*
F30	1	37	23	30	22nd
	0	44	36	40	38th
F31	1	44	11	27.5	19th
	0	36	37	36.5	34th
F32	1	33	33	33	26th
	0	17	22	19.5	17th
F33	1	13	34	23.5	12th
	0	21	24	22.5	21st
F34	1	25	30	27.5	19th
	0	50	38	44	40th
F35	1	22	25	23.5	12th
	0	14	24	19	16th
F36	1	23	19	21	10th
	0	5	12	8.5	8th
F37	1	6	1	3.5	2nd
	0	23	19	21	18th
F38	1	35	28	31.5	24th
	0	39	27	33	30th
F39	1	8	18	13	6th
	0	20	14	17	15th
F40	1	28	21	24.5	13th
	0	22	21	21.5	19th

Feature	BIM Status	Percentage Score Ranking	Summed Ranking	Mean of the Percentage and Summed Ranking	Ranking Derived from the Mean Values
F41	1	20	22	21	10th
	0	30	18	24	22nd
F42	1	36	15	25.5	15th
	0	35	36	35.5	32nd
F43	1	43	8	25.5	15th
	0	15	38	26.5	24th
F44	1	49	24	36.5	30th
	0	10	19	14.5	14th
F45	1	11	32	21.5	11th
	0	11	3	7	6th
F46	*1*	*31*	*40*	35.5	*28th*
	0	*6*	*13*	9.5	*10th*
F47	1	51	13	32	25th
	0	41	31	36	33rd
F48	1	46	17	31.5	24th
	0	2	2	2	2nd
F49	1	5	14	9.5	4th
	0	45	41	43	39th
F50	1	1	3	2	1st
	0	1	1	1	1st
F51	1	48	24	36	29th
	0	16	11	13.5	12th
F52	*1*	*47*	*35*	41	*32nd*
	0	*18*	*25*	21.5	*19th*

Table 7.6: The Top Twenty Ranked BIM and Non-BIM Features Selected for AI Model Development

Feature Category	Serial Number	Feature Name	Feature ID
BIM-Based Features (BIM Status = 1)	1	Traffic restrictions	F50
	2	Unskilled labourer	**F37**
	3	Variation in structural design	F10
	4	Conflicts between consultant and contractor	F4
	5	Poor site investigation or management	F24
	6	Inaccurate budgeting	F39
	7	Consultant cash flow issues	F17
	8	Quality control issues	F1
	9	Conflict between contractor and subcontractor	F2
	10	Late payment by the owner	**F36**
	11	Corruption issues	F27
	12	Poor decision making	F33
	13	Inaccurate resource planning	**F40**
	14	Change orders	F12
	15	Staff use of outdated construction methods	F28
	16	Fluctuation in material prices	F14
	17	Reworks due to error in construction	**F3**
	18	Space limitations at site for permanent equipment	**F22**
	19	Change of specifications during construction	F31
	20	Natural disasters like floods earthquakes etc.	F8
Non- BIM-Based Features(BIM Status = 0)	1	Traffic restrictions	F50
	2	Ineffective project supervision	F48
	3	Reworks due to error in construction	**F3**

Feature Category	Serial Number	Feature Name	Feature ID
	4	Conflict between contractor and subcontractor	F2
	5	Changes in site topography after design	F23
	6	Damaged storage materials	F45
	7	Fluctuation in material prices	F14
	8	Late payment by the owner	F36
	9	Staff use of outdated construction methods	F28
	10	*Nature of equipment(s) used*	F46
	11	Space limitations at site for temporary equipment	F21
	12	Building permit	F51
	13	Labour dispute or strikes	F6
	14	Ineffective government regulations	F44
	15	Inaccurate budgeting	F39
	16	Equipment break	F35
	17	Corruption issues	F27
	18	Unskilled labourer	**F37**
	19	Inaccurate resource planning	**F40**
	20	Client cash flow issues	F16

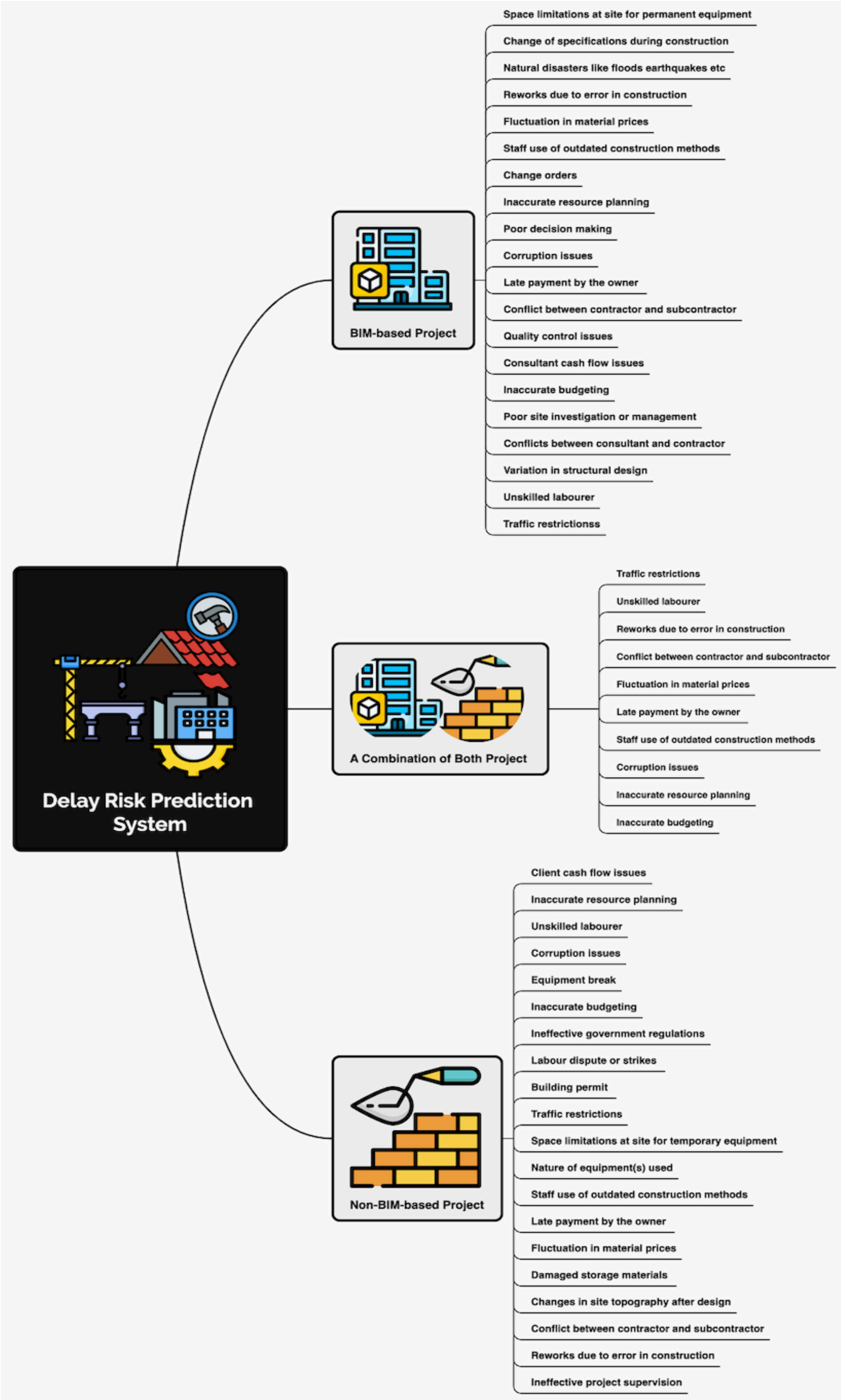

Figure 7.16: Validated framework of delay factors essential for developing predictive models assessing delay risks in both BIM and Non-BIM construction projects.

7.5 The AI technologies/ Algorithms Employed for Developing the Predictive Models.

As established from the findings of the second systematic review conducted in section 3 of this study, the two most powerful Artificial Intelligence (AI) technologies/algorithms widely employed to tackle issues in the construction industry are both supervised and deep learning. Consequently, this study employs these paradigms for developing predictive models. More precisely, therefore, the following subtypes of supervised and deep learning algorithms will be employed They include, **Decision Tree, Logistic Regression, K-Nearest Neighbour, Support Vector Machine, Ensemble Method** (including Random Forest , Gradient Boosting Machine, Adaptive Boosting, Naïve Bayes, Extreme Gradient Boosting, Extra Trees, and Light Gradient Boosting Machine), and **Artificial Neural Network** (including Multi-Layer Perceptron (MLP), Radial Basis Function Network (RBFN), and Fully Connected Neural Network (FCNN)) (Egwim *et al.*, 2024).

I. **Decision Tree (DT):** DT is a versatile supervised learning algorithm that can perform both classification and regression tasks(Egwim and Alaka, 2021). It partitions the feature space into disjoint regions and assigns a class label or predicts a continuous value for each region. The algorithm recursively splits the data based on the features that best separate the classes or minimize the impurity in each partition. This process continues until a stopping criterion is met, such as reaching a maximum depth or minimum number of samples in a leaf node. The decision tree can be represented as a tree structure, where each internal node represents a decision based on a feature, and each leaf node represents the predicted class or value.

II. **Logistic Regression (LR):** LR is a popular supervised learning algorithm used for binary classification tasks. Despite its name, it is a linear model that predicts the probability of an instance belonging to a particular class (Charizanos, Demirhan and İçen, 2024).The logistic function, also known as the sigmoid function, is used to map the output of the linear model to a probability between 0 and 1. Mathematically, the logistic regression model can be expressed as:

$$P(y = 1|x; w) = \frac{1}{1+ e^{-w^T x}} \text{ --------------- Equation 7.2}$$

where:
$P(y = 1|x; w)$ is the probability of class 1 given input vector x,
w is the weight vector, and
e is the base of the natural logarithm.
LR optimizes the parameters (weights) using techniques like gradient descent to minimize the logistic loss function.

III. **K-Nearest Neighbour (KNN):** KNN is a simple, yet effective supervised learning algorithm used for classification and regression tasks. It makes predictions based on the majority class or the average value of the k-nearest data points in the feature space(Yuk Carrie Lin, 2024). Given a new instance, KNN calculates the distance to all training instances and selects the k closest ones. For classification, the class label is determined by majority voting among the k-nearest neighbours, while for regression, the predicted value is the average of the target values of the k-nearest neighbours. KNN's performance heavily relies on the choice of distance metric and the value of k.

IV. **Support Vector Machine (SVM):** SVM is a powerful supervised learning algorithm used for classification and regression tasks (Roy and Chakraborty, 2023). It works by finding the hyperplane that best separates the data points into different classes. The hyperplane is chosen in such a way that it maximizes the margin between the classes, thereby enhancing the generalization ability of the model. Mathematically, the objective of SVM is to find the optimal hyperplane that separates the data points into two classes. This hyperplane can be represented as:

$$w^T x + b \ = 0 ------------------ \text{Equation 7.3}$$

where:
w is the weight vector,
x is the input vector, and
b is the bias term.
The distance between the hyperplane and the closest data points from each class is known as the margin. SVM aims to maximize this margin while minimizing the classification error.

V. **Ensemble Methods (EMs):** EMs combine multiple base models to improve predictive performance (Egwim *et al.*, 2022b). By aggregating the predictions of diverse models, ensemble methods can reduce overfitting and increase robustness. This study examines seven distinct EMs, identified as leading techniques for solving construction-related problems, as presented in Table [reference] in Section 3. They include the following:

- **Random Forest:** A collection of decision trees trained on random subsets of the data and features, with the final prediction being the mode of the individual tree predictions.
- **Gradient Boosting Machine:** Builds a series of decision trees sequentially, with each tree correcting the errors of the previous ones, leading to a highly accurate model.
- **Adaptive Boosting (AdaBoost):** Focuses on misclassified instances by assigning higher weights to them, allowing subsequent models to pay more attention to these instances.

- **Naïve Bayes:** A probabilistic classifier based on Bayes' theorem with strong independence assumptions between features.
- **Extreme Gradient Boosting (XGBoost):** An advanced implementation of gradient boosting with regularization and parallel processing capabilities.
- **Extra Trees:** Similar to Random Forest but with random splits at each node, leading to a higher degree of randomness.
- **Light Gradient Boosting Machine (LightGBM):** Another gradient boosting framework designed for large-scale datasets with faster training and lower memory usage.

VI. **Artificial Neural Network (ANN):** ANN is a computational model inspired by the biological neural networks in the human brain (Egwim, Alaka, Toriola-Coker, Balogun and Sunmola, 2021b). It consists of interconnected nodes, called neurons, organized in layers: an input layer, one or more hidden layers, and an output layer. ANNs are capable of learning complex patterns and relationships from data through a process called backpropagation, where the errors in predictions are propagated backward through the network to adjust the weights and biases. Within the broader scope of ANN techniques, which have been recognized as leading methodologies for addressing construction-related challenges (see section 3), this study will incorporate the following specialized neural network architectures for further investigation and analysis:

- **Multi-Layer Perceptron (MLP):** A feedforward neural network with one or more hidden layers between the input and output layers. Each neuron in the network is connected to every neuron in the adjacent layers, allowing for nonlinear transformations of the input data. The output of an MLP can be calculated as follows:

$$y_j = f\left(\sum_{i=1}^{n} w_{ij} x_i + b_j\right) ------------ \text{Equation 7.4}$$

where:
y_j is the output of neuron j in the output layer,
$f()$ is the activation function,
w_{ij} is the weight of the connection between neuron i in the previous layer and neuron j in the current layer,
x_i is the input to neuron i,
b_j is the bias of neuron j, and
n is the number of neurons in the previous layer.

The architecture of an MLP typically consists of an input layer, one or more hidden layers, and an output layer. Each layer is composed of multiple neurons, and connections between neurons are represented by weighted edges.

- **Radial Basis Function Network (RBFN):** A type of neural network that uses radial basis functions as activation functions in the hidden layer (Dai, Wu and Zhang, 2024). The output of an RBFN can be calculated as follows:

$$y(x) = \sum_{i=1}^{N} w_i \varnothing(||x - c_i||) ------------ \text{Equation 7.5}$$

where:
$y(x)$ is the output of the RBFN for input x,
N is the number of neurons in the hidden layer,
w_i are the weights associated with each neuron,
c_i are the centres of the radial basis functions,
$||x - c_i||$ represents the distance between input x and centre c_i , and
$\varnothing()$ is the radial basis function.

The architecture of an RBFN typically consists of an input layer, a hidden layer with radial basis functions as activation functions, and an output layer. The centres of the radial basis functions are often determined through clustering algorithms such as k-means.

- **Fully Connected Neural Network (FCNN):** Also known as a dense neural network, it is characterized by every neuron in one layer being connected to every neuron in the subsequent layer (Bertschinger *et al.*, 2023; Sun *et al.*, 2024). FCNNs are capable of learning complex nonlinear mappings but may suffer from overfitting on large datasets. Mathematically the output of an FCNN is expressed as follows:

$$y = f(W^T x + b\,) ----------------- \text{Equation 7.6}$$

where:
y is the output vector,
W is the weight matrix,
x is the input vector,
b is the bias vector, and
$f()$is the activation function.

The architecture of an FCNN consists of an input layer, one or more hidden layers, and an output layer. Each layer is fully connected to the next layer, forming a dense network of connections.

7.6 Performance Evaluation Metrics

The AI technologies/ algorithms employed exhibit a spectrum of transparency, each offering distinct levels of insight. While certain approaches yield tangible models—albeit not always readily interpretable—others operate without producing a discernible model. Consequently, within this subsection, this study explores into the intricacies underpinning the resultant AI models tailored for both BIM and Non-BIM-based construction projects. Furthermore, it details the specifics pertaining to these developed models, shedding light on their inner workings and implications via performance evaluation metrics below. These metrics provide insights into different aspects of the model's behaviour and effectiveness in classifying instances correctly. Also, these metrics provide critical insights into which features most significantly impact delay predictions. For instance, if certain features consistently result in high precision but low recall, this may indicate a need for additional features that capture missed delay cases. Conversely, features leading to high recall, but low precision might indicate a need for more granular or specific features to reduce false positives.

I. **Accuracy:** Accuracy measures the proportion of correctly classified instances out of the total number of instances in the dataset (Padilla, Netto and Da Silva, 2020). It provides a general assessment of the model's overall performance. While accuracy provides an overall snapshot of the model's performance, it can sometimes be misleading, especially in cases of class imbalance where one class significantly outnumbers the other. In this study, focusing on BIM and Non-BIM-based construction projects, accuracy alone may not fully capture the model's effectiveness. This is because accuracy can be inflated by the predominant class's correct predictions. Therefore, accuracy should be considered alongside other metrics that provide a more nuanced understanding of the model's performance across all classes.

$$Accuracy\ = \frac{TN +\ TP}{TP +\ TN + FP + FN}) - - - - - - - - \text{Equation 7.7}$$

where:

TP is the number of true positives (correctly predicted positive instances),

TN is the number of true negatives (correctly predicted negative instances),

FP is the number of false positives (incorrectly predicted positive instances), and

FN is the number of false negatives (incorrectly predicted negative instances).

II. **Precision:** Precision measures the proportion of true positive predictions among all positive predictions made by the model (Padilla *et al.*, 2021). It indicates the model's ability to avoid false positive predictions. Precision is particularly relevant in scenarios where false positives have significant consequences. For instance, in predicting delays in construction projects, a false positive (predicting a delay where there isn't one) could lead to unnecessary reallocation of resources or adjustments in project planning. High precision is thus crucial in ensuring that when a delay is predicted, it is indeed likely to occur, thereby optimizing decision-making processes and resource allocation in project management.

$$Precision = \frac{TP}{TP + FP} \text{ -------- Equation 7.8}$$

III. Recall (Sensitivity): Recall measures the proportion of true positive predictions among all actual positive instances in the dataset (Tharwat, 2018). It indicates the model's ability to capture positive instances. Recall is crucial in contexts where capturing all relevant instances is essential. In this study, high recall is vital to ensure that most, if not all, potential delays are identified. This metric is particularly important for risk management in construction projects, where failing to identify a delay (false negative) could result in unanticipated costs and disruptions. Thus, a model with high recall is effective in minimizing the risk of missing critical delay indicators.

$$Recall = \frac{TP}{TP + FN} \text{ -------- Equation 7.9}$$

IV. F1 Score: The F1 score is the harmonic mean of precision and recall. It provides a balanced measure that considers both false positives and false negatives (Miao and Zhu, 2022). The F1 score balances precision and recall, providing a single metric that considers both the false positives and false negatives. This balance is crucial in construction project management, where both over-predicting and under-predicting delays can have significant ramifications. The F1 score helps in understanding the trade-offs between precision and recall, ensuring that the model's predictions are both accurate and comprehensive.

$$F_1\, score = \frac{2}{\frac{1}{Precision} + \frac{1}{Recall}} = 2\, x \frac{Precision\, x\, Recall}{Precision + Recall} \text{ --- Equation 7.10}$$

V. Specificity: Specificity measures the proportion of true negative predictions among all actual negative instances in the dataset (Tharwat, 2018). It indicates the model's ability to correctly identify negative instances. Specificity, or the true negative rate, is particularly important in scenarios where it is crucial to correctly identify non-delayed projects. High specificity ensures that projects not experiencing delays are not incorrectly flagged, which is important for maintaining trust and efficiency in project management. In the context of this study, specificity helps prevent the misallocation of resources that might occur if non-delayed projects are incorrectly identified as delayed.

$$Specificity = \frac{TN}{TN + FP} \text{ ------- Equation 7.11}$$

VI. Receiver Operating Characteristic Area Under the Curve (ROC-AUC): ROC- AUC measures the area under the ROC curve, which plots the true positive rate (sensitivity) against the false positive rate (1 - specificity) at various threshold settings (Miao and Zhu, 2022). It provides a measure of the model's ability to discriminate between positive and negative instances across different threshold values. This is particularly useful in construction delay prediction, where the decision threshold can vary based on the project's criticality and the acceptable risk levels. A high ROC-AUC value indicates that the model is effective at distinguishing between delayed and non-delayed projects across various decision thresholds, providing flexibility in its application.

VII. **Log Loss (Cross-Entropy Loss):** Log loss measures the performance of a classification model based on the predicted probability of the true class (Mei *et al.*, 2024; Mushava and Murray, 2024). It is commonly used as the loss function during training and can be used for evaluation, with lower values indicating better performance. Log loss is essential for assessing the confidence level of predictions regarding project delays. Lower log loss values indicate that the model assigns higher probabilities to the correct class, thus providing more reliable predictions. This metric is critical when the model's output is used for making probabilistic decisions in resource allocation and risk management.

VIII. **Confusion Matrix:** A confusion matrix is a tabular representation (see Table 7.7) of the model's predictions compared to the actual class labels (Egwim, Alaka, Toriola-Coker, Balogun and Sunmola, 2021b). It provides detailed information about true positive, true negative, false positive, and false negative predictions, allowing for a deeper understanding of the model's performance. The confusion matrix offers a detailed breakdown of the model's performance, showcasing the distribution of true positives, true negatives, false positives, and false negatives. This detailed view is crucial for understanding specific areas where the model may be overestimating or underestimating delays. By analyzing the confusion matrix, stakeholders can identify specific patterns or biases in the model's predictions, allowing for targeted improvements in model training and feature selection.

Table 7.7 Confusion Matrix

		Prediction	
		Negative (delay < threshold limit) = 0	Positive (delay > threshold limit) = 1
Actual	Negative (delay < threshold limit) = 0	*True Negative*	*False Positive*
	Positive (delay > threshold limit) = 1	*False Negative*	*True Positive*

7.7 AI Model Development

In this study, the respective AI models were developed by importing a running instance of Jupiter Notebook using Scikit-learn - an integral Python programming language module with a broad spectrum of state-of-the-art algorithms for supervised and unsupervised medium-scale problems (Pedregosa *et al.*, 2011). Furthermore, the experimentation was performed on an Apple MacBook Pro macOS Monterey version 12.4 with an Apple M1 chip, 16 gigabyte random access memory, and 8 cores hardware. The training dataset (70% of the total dataset) was used to fit the fourteen different algorithms listed in the previous subsection, while their hyperparameters were optimized during successive runs to further improve the performance for making predictions on the unseen test dataset (30% of the total dataset). The model development process began with data pre-processing, which involved handling missing values, encoding categorical variables, and scaling numerical features as detailed in previous sections. The model development process involved creating the separate models for BIM and non-BIM-based construction projects. The dataset was split accordingly, and the pre-processing, training, and evaluation steps were performed independently for each project type. The pre-processed data was then split into training and testing sets, with the training set used for model fitting and the testing set reserved for evaluating the models' performance on unseen data. Each of the fourteen algorithms (four classical algorithms, seven ensemble methods, and three artificial neural networks) was implemented using the respective libraries and functions provided by Scikit-learn. Hyperparameter tuning was performed using techniques such as grid search and randomized search cross-validation to optimize the models' performance. For instance, in the case of Decision Trees, parameters like the maximum depth of the tree, the minimum number of samples required to split a node, and the criteria for splitting nodes (e.g., Gini impurity or entropy) were tuned. Similarly, for Logistic Regression, the regularization strength and the solver algorithm were optimized. The number of neighbours in the K-Nearest Neighbour algorithm and the kernel function in Support Vector Machines were also fine-tuned. Ensemble Methods required careful selection and tuning of the base learners and their respective hyperparameters. For example, in Random Forest, the number of trees, the maximum depth of each tree, and the number of features to consider at each split were optimized. In Gradient Boosting techniques like XGBoost and LightGBM, parameters like the learning rate, the maximum depth of each tree, and the regularization strengths were tuned. Artificial Neural Networks, being more complex models, required extensive hyperparameter optimization. The number of hidden layers, the number of neurons in each layer, the activation functions, the optimization algorithm, the learning rate, and the regularization techniques were all carefully tuned to prevent overfitting and improve generalization performance. After training and tuning the models for both BIM and non-BIM-based construction projects, their performances were evaluated on the testing dataset using various evaluation metrics, including Accuracy, Confusion Matrix, Precision, Recall, F1-Score, and the ROC-AUC. These metrics provided insights into different aspects of the models' behaviour and effectiveness in correctly classifying instances. The results of these evaluation metrics were visualized using appropriate plots and figures (Figures 7.17 to 7.24) to facilitate a comprehensive analysis and comparison of the models' performances. The visualization techniques employed included charts for ROC-AUC for assessing the trade-off between true positive and false positive rates. Finally, Table 7.8 presents a comprehensive comparison of the fourteen AI models developed. These models, categorized as supervised learning models, will be thoroughly discussed in the subsequent chapter. As described earlier in subsection 7.4, the computation of overall accuracy, specificity, sensitivity, precision, and recall is based on the analysis of the confusion matrix associated with each model.

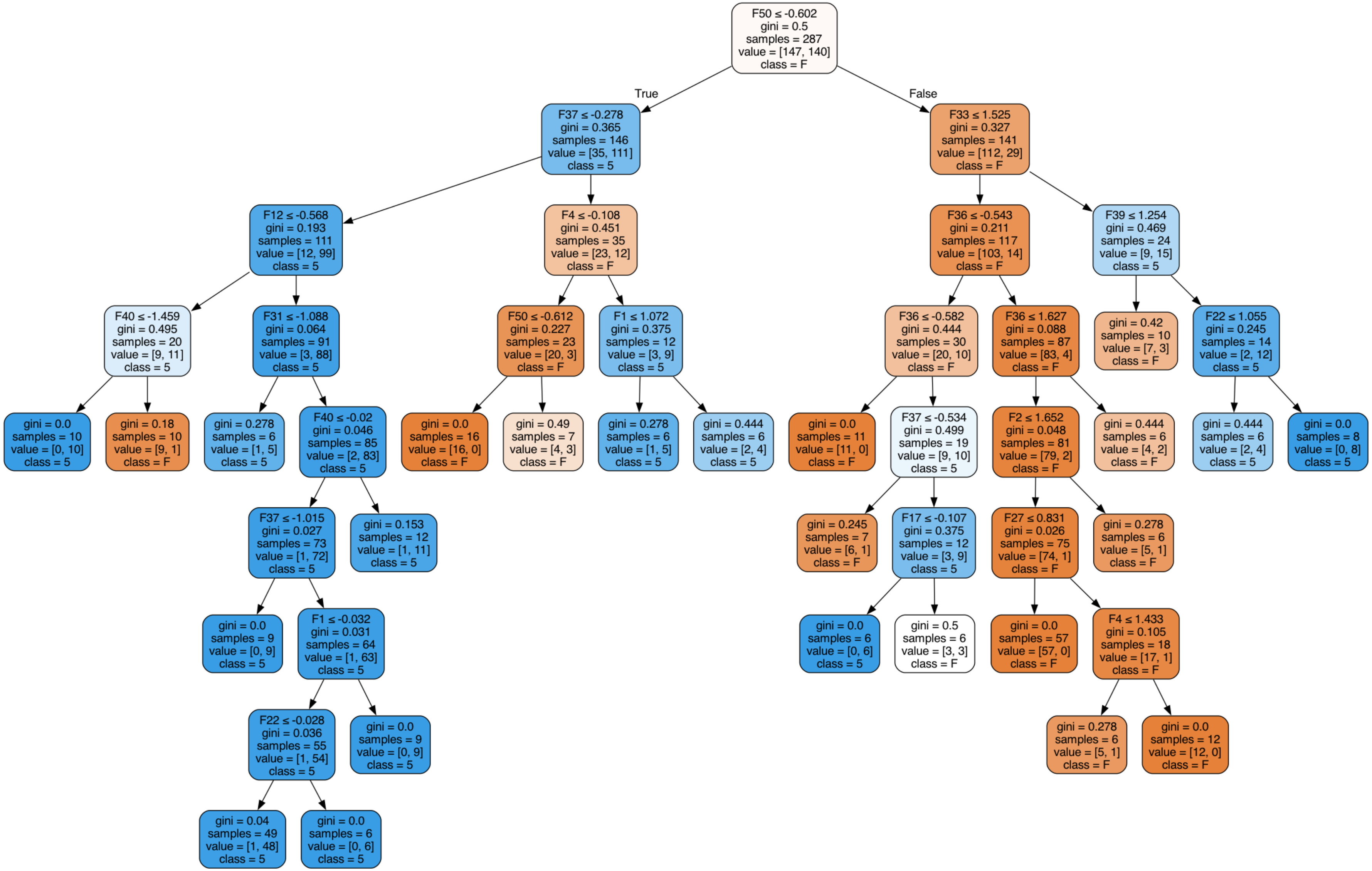

Figure 7.17: Decision making process of the decision tree model for BIM-based project.

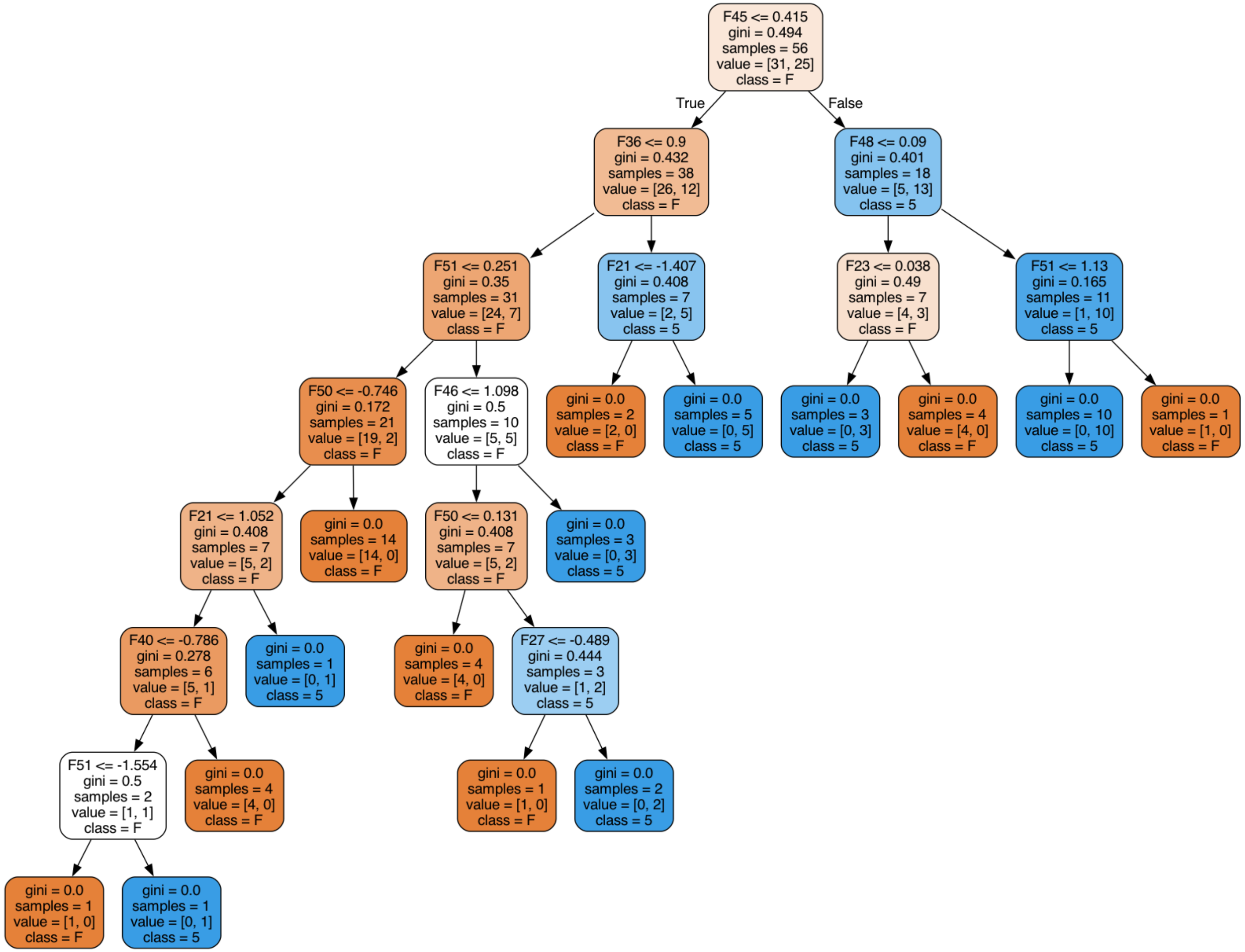

Figure 7.18: Decision making process of the decision tree model for non-BIM-based project.

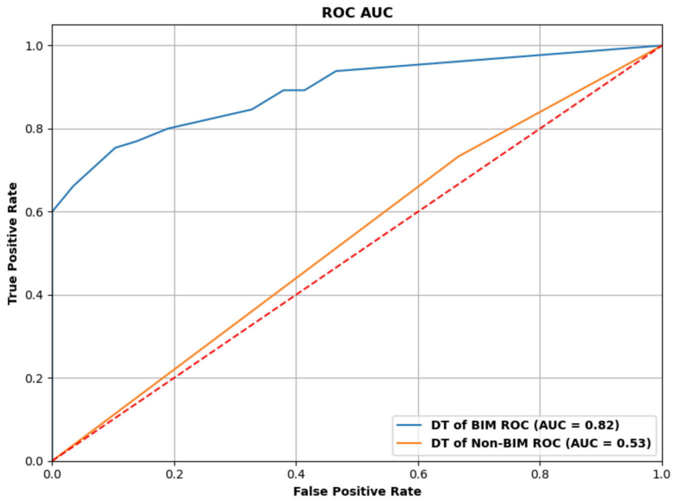

Figure 7.19: DT ROC AUC Plot

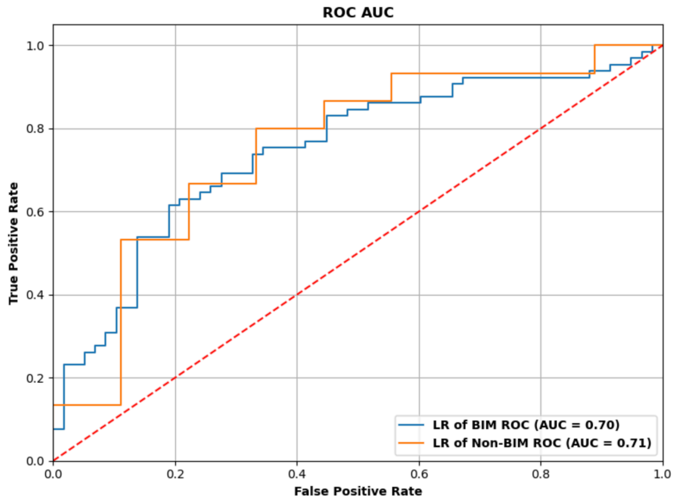

Figure 7.20: LR ROC AUC Plot

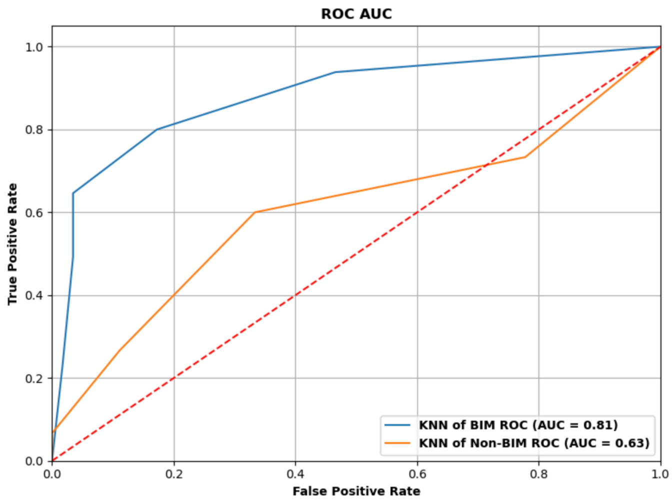

Figure 7.21: KNN ROC AUC Plot

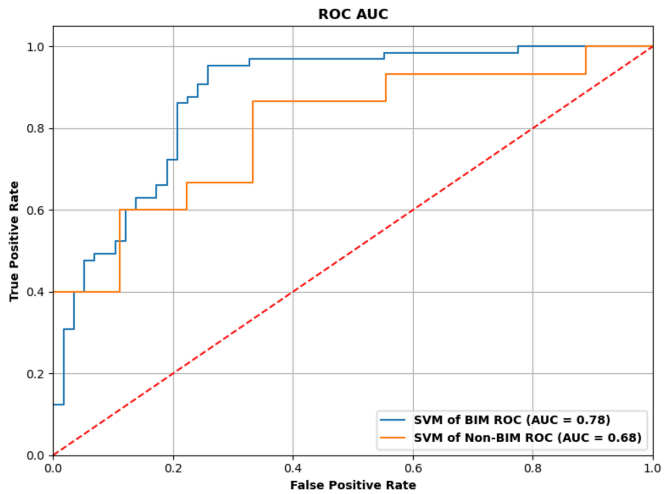

Figure 7.22: SVM ROC AUC Plot

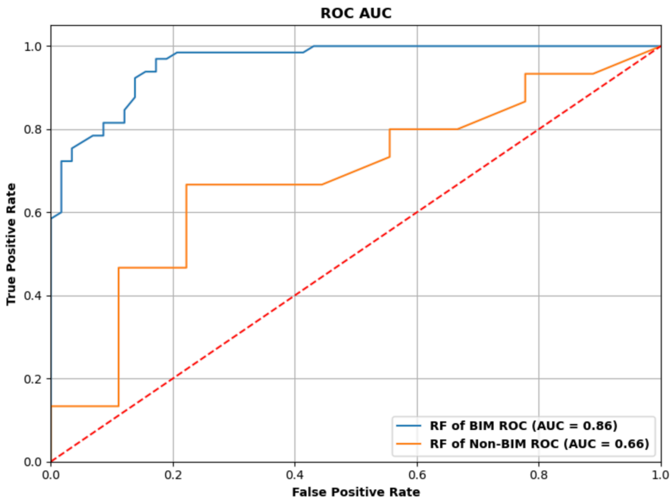

Figure 7.23: RF ROC AUC Plot

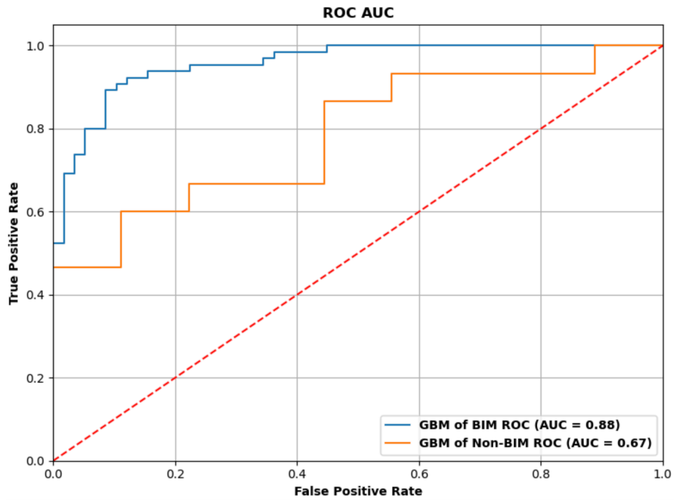

Figure 7.24: GBM ROC AUC Plot

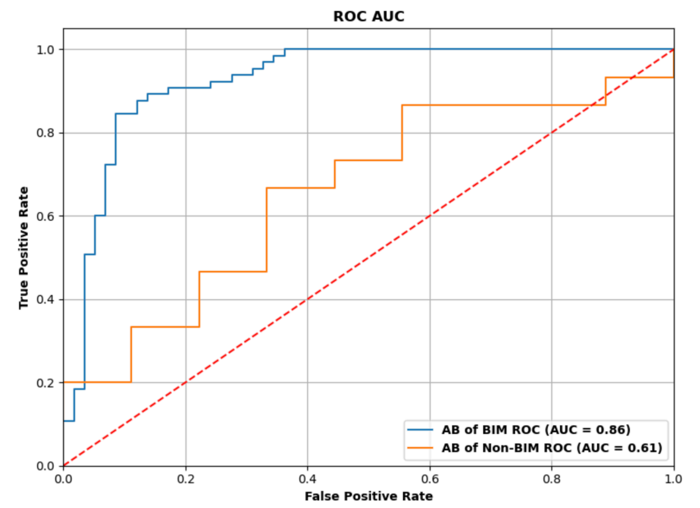

Figure 7.25: AB ROC AUC Plot

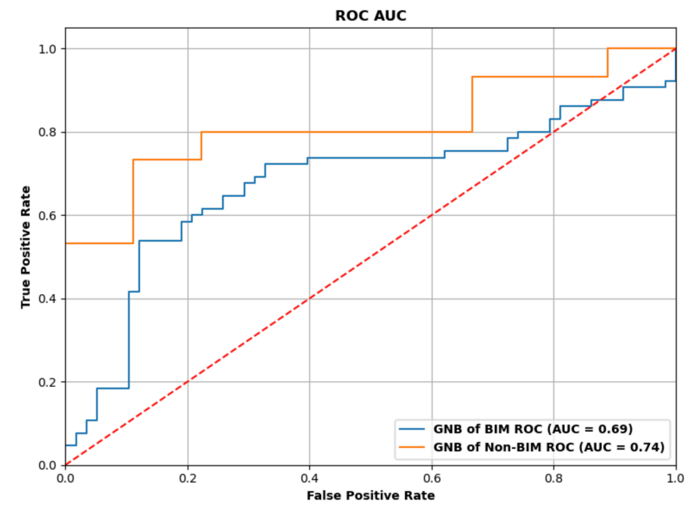

Figure 7.26: GNB ROC AUC Plot

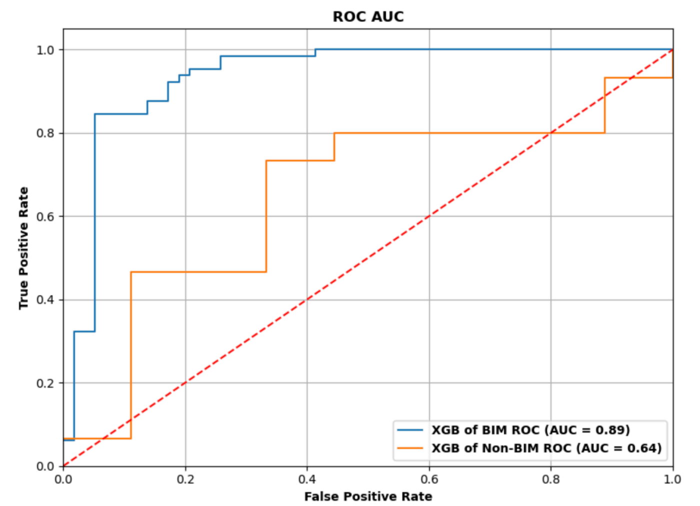

Figure 7.27: XGB ROC AUC Plot

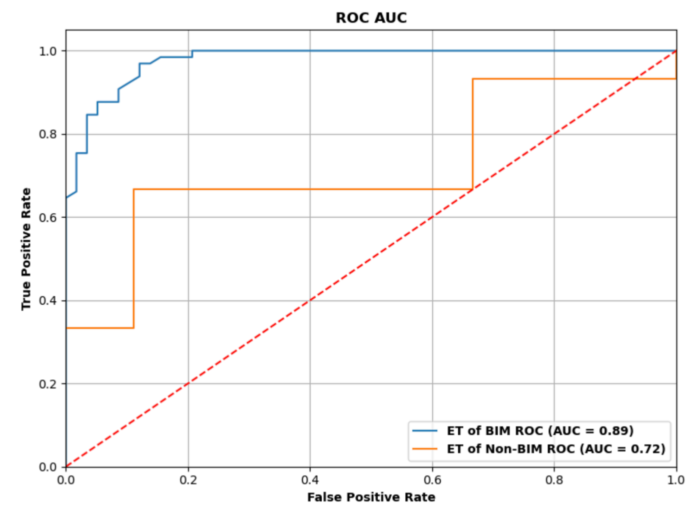

Figure 7.28: ET ROC AUC Plot

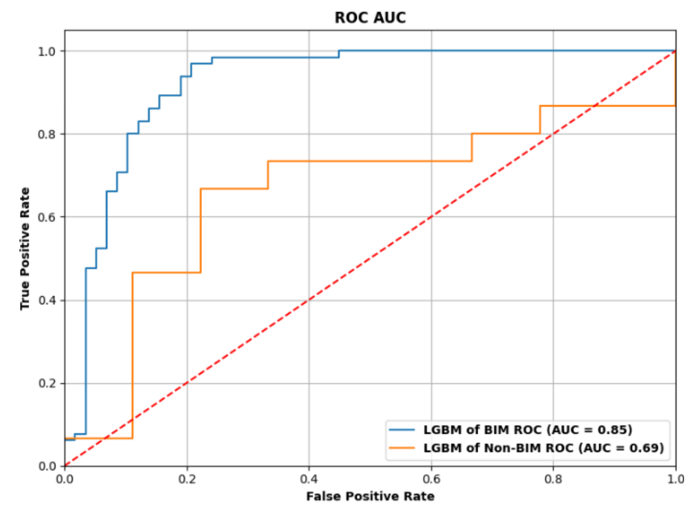

Figure 7.29: LGBM ROC AUC Plot

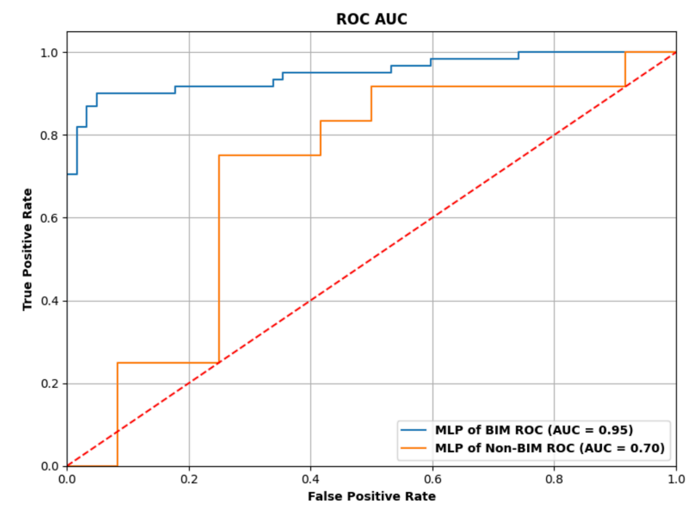

Figure 7.24: MLP ROC AUC Plot

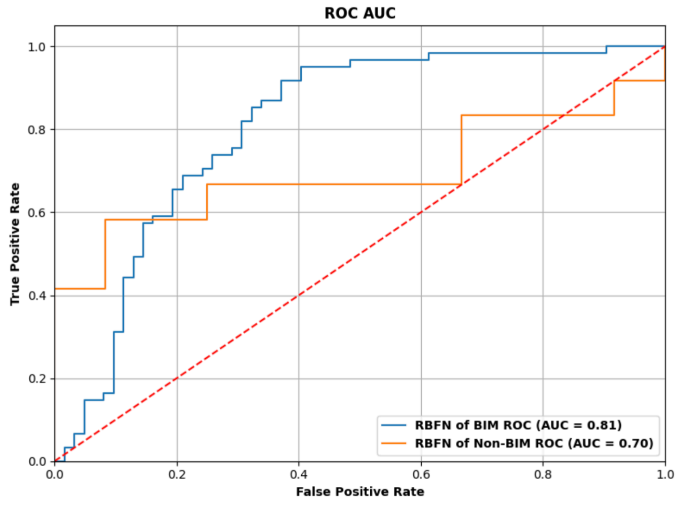

Figure 7.29: RBFN ROC AUC Plot

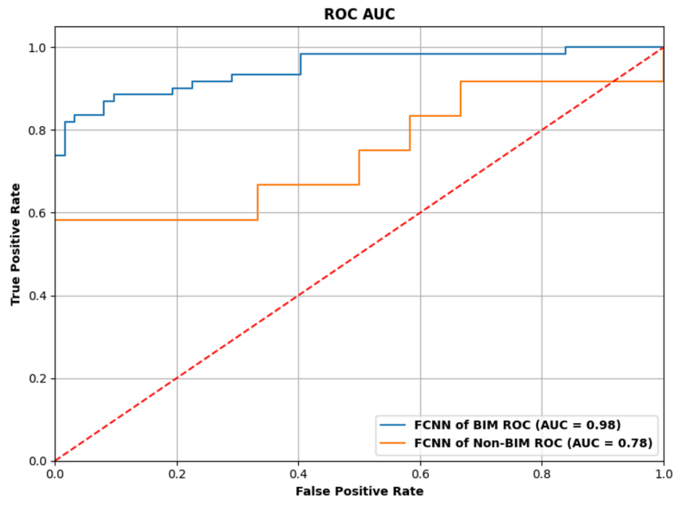

Figure 7.24: FCNN ROC AUC Plot

Table 7.8: Summary of the performance metrics for each supervised learning model developed

AI Technologies	Supervised Learning	BIM Status	Precision (%)	Recall (%)	F_1 Score (%)	Specificity (%)	Accuracy (%)	ROC AUC
Classical Algorithms	DT	1	74%	93%	82%	93%	81%	0.819
		0	65%	73%	69%	33%	58%	0.533
	LR	1	71%	72%	72%	67%	70%	0.698
		0	89%	53%	67%	89%	67%	0.711
	KNN	1	95%	65%	77%	97%	80%	0.806
		0	75%	60%	67%	67%	63%	0.633
	SVM	1	81%	77%	79%	79%	78%	0.781
		0	88%	47%	61%	89%	63%	0.678
	RF	1	91%	82%	86%	91%	86%	0.865
		0	80%	53%	64%	78%	63%	0.656
	GBM	1	92%	85%	88%	91%	88%	0.880
		0	77%	67%	71%	67%	67%	0.667
	AB	1	91%	82%	86%	91%	86%	0.865
		0	71%	67%	69%	56%	63%	0.611
	GNB	1	72%	65%	68%	72%	68%	0.685

AI Technologies	Supervised Learning	BIM Status	Precision (%)	Recall (%)	F_1 Score (%)	Specificity (%)	Accuracy (%)	ROC AUC
Ensemble Methods		**0**	**90%**	**60%**	**72%**	**89%**	**71%**	**0.744**
	XGB	**1**	**93%**	**85%**	**89%**	**93%**	**89%**	**0.889**
		0	**73%**	**73%**	**73%**	**56%**	**67%**	**0.644**
	ET	**1**	**96%**	**82%**	**88%**	**97%**	**87%**	**0.891**
		0	**91%**	**67%**	**77%**	**89%**	**75%**	**0.778**
	LGBM	**1**	**88%**	**82%**	**85%**	**88%**	**85%**	**0.847**
		0	**82%**	**60%**	**69%**	**78%**	**67%**	**0.689**
Artificial Neural Networks	**MLP**	**1**	**95%**	**89%**	**92%**	**96%**	**92%**	**0.952**
		0	**75%**	**75%**	**75%**	**75%**	**75%**	**0.701**
	RBFN	1	**71%**	**90%**	**80%**	**65%**	**77%**	**0.806**
		0	**88%**	**58%**	**70%**	**92%**	**75%**	**0.701**
	FCNN	*1*	***96%***	***90%***	***93%***	***97%***	***93%***	***0.982***
		0	***82%***	***75%***	***78%***	***83%***	***79%***	***0.777***

7.8 Chapter Summary

In this chapter, the development process of AI predictive models for both BIM-based and non-BIM-based construction projects was comprehensively outlined. The journey commenced with a detailed data pre-processing phase, encompassing data profiling, cleaning, and exploratory data analysis. This initial stage unveiled the discrete categorical nature of the dataset, consisting of 52 features and a target variable, with ordinal values ranging from one to five. After data pre-processing, the study examined feature engineering, a pivotal endeavour involving the systematic selection, alteration, and generation of features. Techniques such as one-hot encoding and SMOTE oversampling were employed to address class imbalance and facilitate effective model development. The chapter then transitioned to an in-depth exploration of feature selection methodologies, rigorously examining various approaches, including filter-based, wrapper-based, and embedded techniques. A comprehensive synthesis of these diverse methods culminated in the selection of the top twenty most influential features for model development. The AI technologies and algorithms employed for developing predictive models were then elucidated, encompassing supervised learning algorithms, ensemble methods, and artificial neural networks. These included Decision Trees, Logistic Regression, K-Nearest Neighbours, Support Vector

Machines, Random Forests, Gradient Boosting Machines, AdaBoost, Naïve Bayes, XGBoost, Extra Trees, LightGBM, Multi-Layer Perceptron, Radial Basis Function Networks, and Fully Connected Neural Networks. To evaluate the performance of the developed models, a comprehensive array of evaluation metrics was introduced, including accuracy, precision, recall, F1 score, specificity, and ROC-AUC. The chapter provided a detailed explanation of these metrics and their significance in assessing model performance. Finally, the chapter researched into the intricate process of AI model development, detailing the utilization of Scikit-learn, a Python programming language module, and the implementation of the fourteen algorithms on an Apple MacBook Pro with an M1 chip. The models were developed independently for BIM-based and non-BIM-based construction projects, with data pre-processing, training, and evaluation steps performed separately for each project type. The chapter culminated with a comprehensive comparison of the fourteen developed AI models, categorized as supervised learning models, based on the evaluation metrics. Visualizations, including decision-making processes of decision tree models for both project types, were provided to facilitate a thorough analysis and comparison of the models' performances.

CHAPTER 8

8.0 AI MODEL SELECTION: FINDINGS AND DISCUSSION

8.1 Chapter introduction

This chapter examines the findings of the research and discusses their implications in detail. It presents the results obtained from analysing 14 distinct artificial intelligence prediction models, as well as the critical factors identified by the most effective model. Section 8.2 elaborates on the methodology used to select the optimal model, focusing on three pivotal criteria: accuracy, precision, and recall. Subsection 8.2.1 provides a detailed account of the model selection process based on accuracy, while subsection 8.2.2 explore the selection criteria concerning precision and recall. Furthermore, subsection 8.2.3 introduces the standout model for the respective project type, namely the fully connected neural network (FCNN). Moving forward, Section 8.3 sheds light on the critical factors contributing to delays in both BIM and Non-BIM construction projects, drawing insights from the variables identified by the superior FCNN model via model interpretability/explainability using SHapley Additive exPlanations (SHAP). Subsection 8.3.1 to 8.3.5 investigated the top five factors pertinent to both projects. Moreover, Section 8.4 explores the theoretical implications of the research, particularly in terms of bolstering existing theories surrounding construction project delays. By grounding the findings in theoretical frameworks, this section shows the broader significance of the research outcomes. Lastly, Section 8.5 encapsulates the essence of the chapter by providing a concise summary of the main points discussed throughout.

8.2 Selection of the AI predictive Model

The process of selecting the optimal model from a range of options heavily relies on the goals of the AI engineer, which are intricately intertwined with the needs of the end user. For instance, a contractor's primary focus often centres on the accuracy of an AI model, as it directly influences decisions regarding project timelines. Within the scope of this research, the primary stakeholders are the construction project owners, driven by the overarching aim of mitigating project delays pervasive the construction sector. Consequently, attributes such as accuracy, precision, recall and F1 score assume paramount importance in the curation of an AI predictive model tailored to these exigencies. The selection of an optimal AI predictive model necessitates a rigorous evaluation of the performance metrics across the array of developed models. While accuracy serves as an intuitive gauge of model efficacy, a holistic assessment mandates a deeper examination of precision, recall, and the elusive balance they embody through the F1 score. As a reminder, the outcomes of the developed AI predictive models are succinctly summarized in Table 7.8, encapsulating the comprehensive overview of their performance metrics.

Moreover, the discriminatory prowess of the AI predictive models, as encapsulated by the Receiver Operating Characteristic (ROC) curve and its corresponding Area Under the Curve (AUC), offers pivotal insights. Collectively, these metrics illuminate the multifaceted performance profiles of the respective models, informing a judicious selection tailored to the specific aim of this study and thereby enriching the understanding of their efficacy in real-world applications.

8.2.1 Model Accuracy

Accuracy, though a conspicuous metric, belies the nuanced complexities inherent in evaluating classification models, particularly when grappling with imbalanced datasets. As shown in Table 7.8, the accuracy scores span a considerable range, with the highest performers being the Fully Connected Neural Network (FCNN) at 93% and 79% for BIM-based and non-BIM-based projects, respectively. Notably, the FCNN outperforms all other models, including its neural network counterparts, the Multi-Layer Perceptron (MLP) and the Radial Basis Function Network (RBFN), across both project domains. Among the ensemble methods, the Extra Trees (ET) algorithm emerges as a formidable contender, achieving an accuracy of 87% for BIM-based projects, while the Extreme Gradient Boosting (XGB) model closely trails at 89%. Conversely, in the non-BIM domain, the XGB model exhibits a relatively lower accuracy of 67%, superseded by the Extra Trees (ET) algorithm's 75% and the Gaussian Naive Bayes (GNB) model's 71%. The classical algorithms, though outperformed by their contemporaries, still exhibit reasonable accuracy levels. The Decision Tree (DT) model attains 81% accuracy for BIM-based projects, while its non-BIM counterpart lags at 58%. Similarly, the K-Nearest Neighbour (KNN) algorithm achieves 80% and 63% accuracy for BIM and non-BIM projects, respectively, with the Support Vector Machine (SVM) closely following at 78% and 63%. While these accuracy scores offer a preliminary vantage point, their interpretation necessitates a nuanced understanding of the underlying expert survey dataset characteristics. The inherent class imbalance, wherein instances of project delays outnumbered those without delays (see section 7.2), renders accuracy a potentially misleading metric. Models exhibiting high accuracy may still falter in correctly identifying the minority class, a critical consideration in the scope of this study, where accurate delay prediction is paramount.

8.2.2 Model Precision, Recall

To circumvent the pitfalls of over-reliance on accuracy, a comprehensive evaluation must encompass precision and recall, two complementary metrics that elucidate a model's adeptness in handling imbalanced datasets. Precision quantifies the proportion of true positive predictions among all positive predictions made by the model. In other words, it measures the model's ability to avoid false positive predictions, a trait of paramount importance when assessing the risk of project delays. A high precision score signifies that when the model predicts a delay, that prediction is highly likely to be accurate. On the other hand, recall, also known as sensitivity, measures the proportion of true positive predictions among all actual positive instances in the dataset. It reflects the model's capability to capture and correctly identify instances of project delays, thereby minimizing false negatives, which could prove detrimental in risk assessment endeavours. As illustrated in Table 7.8, the FCNN model exhibits remarkable precision scores of 96% and 82% for BIM-based and non-BIM-based projects, respectively. This exceptional performance is closely trailed by the ET algorithm, with precision scores of 96% and 91% across the two project types. The XGB model also demonstrates high precision, achieving 93% for BIM-based projects and a respectable 73% for non-BIM-based projects. In terms of recall, the FCNN once again emerges as a standout performer, attaining scores of 90% and 75% for BIM-based and non-BIM-based projects, respectively. The MLP model follows closely, with recall scores of 89% and 75%, while the ET algorithm achieves 82% recall for BIM-based projects and a comparatively lower 67% for non-BIM-based endeavours. The synergy between precision and recall is encapsulated by the F1 score, a harmonious mean that balances their respective contributions. Here, the FCNN model exhibits a remarkable F1 score of 93% for BIM-based projects and a commendable 78% for non-BIM-based undertakings. The ET algorithm follows suit, with F1 scores of 88% and 77%, respectively, while the XGB model achieves 89% for BIM-based projects and a lower 73% for non-BIM-based endeavours. The stark contrast in precision and recall scores across the various models underscores the importance of a nuanced evaluation strategy. Models that excel in accuracy may falter when it comes to identifying the minority class, a scenario that could prove catastrophic in the realm of risk assessment for construction projects. The FCNN's exceptional performance in both precision and recall, coupled with its impressive F1 scores, positions it as a compelling choice for the development of robust predictive models capable of accurately predicting project delays, irrespective of the project's adherence to BIM methodologies or traditional non-BIM approaches.

8.2.3 Fully Connected Neural Network (FCNN) as the Chosen AI Predictive Model

The selection of the FCNN as the optimal AI predictive model for this study is buttressed by a multitude of factors that transcend its quantitative performance metrics. While the FCNN's exceptional accuracy, precision, recall, and F1 scores undeniably contribute to its appeal, the inherent characteristics of this neural network architecture further solidify its suitability for the task at hand. Firstly, the FCNN's ability to capture and model complex, non-linear relationships within the data are of paramount importance. This is because the dynamics that underpin project delays are intrinsically multifaceted, with intricate interplays between various factors contributing to their manifestation. Therefore, the FCNN's capacity to disentangle these intricate patterns and discern the underlying mechanisms that govern delay occurrences is an invaluable asset. Moreover, the FCNN's resilience to overfitting, a pervasive challenge in the realm of machine learning, further bolsters its credibility. By virtue of its dense interconnectivity and the judicious application of regularization techniques, the FCNN mitigates the risk of memorizing idiosyncrasies within the training data, thereby enhancing its generalization capabilities. This trait is particularly crucial in construction projects, where unforeseen circumstances and unique project dynamics are the norm rather than the exception. Additionally, the FCNN's interpretability/ explainability contribute to its appeal as the chosen model. While neural networks have often been criticized for their opaque nature, the FCNN's architecture and the ability to visualize its learned weights and biases facilitate a deeper understanding of the model's decision-making process. This explainability(as shown in the next section) is invaluable in the realm of risk assessment, where stakeholders demand not only accurate predictions but also insights into the underlying rationale driving those predictions. Furthermore, the FCNN's scalability and adaptability position it as a future-proof solution. As the construction industry continues to evolve, with new methodologies and technologies emerging, the FCNN's ability to assimilate and adapt to novel data patterns will prove indispensable. Its inherent capacity for continual learning and refinement ensures that the model remains relevant and effective, even as the landscape of construction projects undergoes transformative shifts. Finally, the FCNN's versatility in handling both BIM-based and non-BIM-based projects is a testament to its robustness and applicability across diverse project types. This attribute is particularly significant in the context of this study, which aims to develop predictive models that cater to the nuances of both traditional and cutting-edge construction methodologies.

8.3 Model Interpretability/Explainability using SHapley Additive exPlanations (SHAP)

Model interpretability, often synonymous with explainability, refers to the ability to understand and interpret the decisions made by machine learning models. It plays a pivotal role in fostering trust, enabling stakeholders to comprehend the underlying factors driving model predictions and facilitating informed decision-making. One powerful technique that has gained significant traction in the realm of model interpretability is SHapley Additive exPlanations (SHAP). SHAP provides a framework for explaining the output of any machine learning model by attributing the prediction to its individual feature contributions. SHAP operates based on the principle of Shapley values, a concept borrowed from cooperative game theory. Shapley values assign each feature a proportional contribution to the model prediction, considering all possible combinations of features. This comprehensive approach ensures fairness and consistency in attributing credit to each feature, thereby providing a reliable measure of feature importance. When compared to alternative methods such as Local Interpretable Model-agnostic Explanations (LIME) or Partial Dependence Plots (PDP), SHAP offers several distinct advantages. Firstly, SHAP provides global interpretability by explaining the entire model's behaviour, rather than focusing solely on individual predictions (Molnar, 2022). This holistic view enables stakeholders to gain deeper insights into the model's decision-making process and identify patterns across the dataset. Furthermore, SHAP's ability to handle complex models, including deep neural networks, makes it well-suited for this study. The SHAP formula used in explaining the chosen AI predictive model (FCNN) is as follows:

$$\emptyset_i = \sum_{S \subseteq N \setminus \{i\}} \frac{(N - |S| - 1)!\,|S|!}{N!} \left[f(S \cup \{i\}) - f(S)\right] \ldots\ldots\ldots\ldots\ldots\ldots Equation\ 8.1$$

Where:

$\emptyset_i$represents the Shapley value of feature i ,

N is the total number of features,

S is a subset of features excluding feature i ,

$f(S \cup \{i\})$ is the model's prediction with feature i included,

$f(S)$ is the model's prediction without feature i .

To initiate the SHAP analysis, the FCNN model was subjected to examination to display its predictive mechanisms text dataset. This involved utilizing the 'Kernel Explainer' function to instantiate an explainer object. Subsequently, this explainer object was leveraged to compute SHAP values for each observation within the prospective metrics of the test dataset, where the dataset comprises row records representative of construction projects. Figure 8.1 offers a detailed breakdown of the factors influencing the probability of construction project delays, employing SHAP values to provide insights into feature contributions. Commencing with a baseline value of 0.507, indicative of the expected outcome in the absence of any specific features, the plot progresses towards a final predicted probability of f(x)=0.908, representing the likelihood of experiencing delays. Upon closer examination, certain features are found to exert positive influences on the prediction, thereby increasing the likelihood of delays. Notably, F40 emerges as the most impactful positive factor (solely based on the specified value of f(x) for a given feature), contributing +0.1 to the overall prediction. Additionally, F50, F36, F37, F31, F1, and F4 contribute positively, each enhancing the probability of delay occurrence by +0.08, +0.08, +0.06, +0.06, +0.05, and +0.04, respectively. Conversely, some features exhibit negative impacts, indicating a decreased likelihood of delays. Particularly, F22 stands out as the most influential negative factor, with a contribution of -0.05 to the prediction. The color-coded representation further enhances the interpretability of the plot. Positive impacts are depicted in red, highlighting features that exacerbate delays, while negative impacts appear in blue, indicating factors that mitigate delays. This visual distinction aids in quickly identifying the key drivers behind delay occurrences, enabling stakeholders to focus their attention and resources accordingly. In interpreting these results, it is essential to consider the underlying principle of SHAP theory, which is rooted in game theory and the concept of Shapley values. SHAP values allocate credit for the prediction to each feature, considering all possible combinations of features and their contributions. This approach ensures fair attribution of prediction outcomes to individual features, providing a comprehensive understanding of their respective impacts on the model's output. By leveraging SHAP values within the context of a waterfall plot, stakeholders gain valuable insights into the relative importance of different features in determining the likelihood of project delays. Armed with this knowledge, decision-makers can formulate targeted strategies to address key contributing factors, thereby enhancing project planning, management, and ultimately, success.

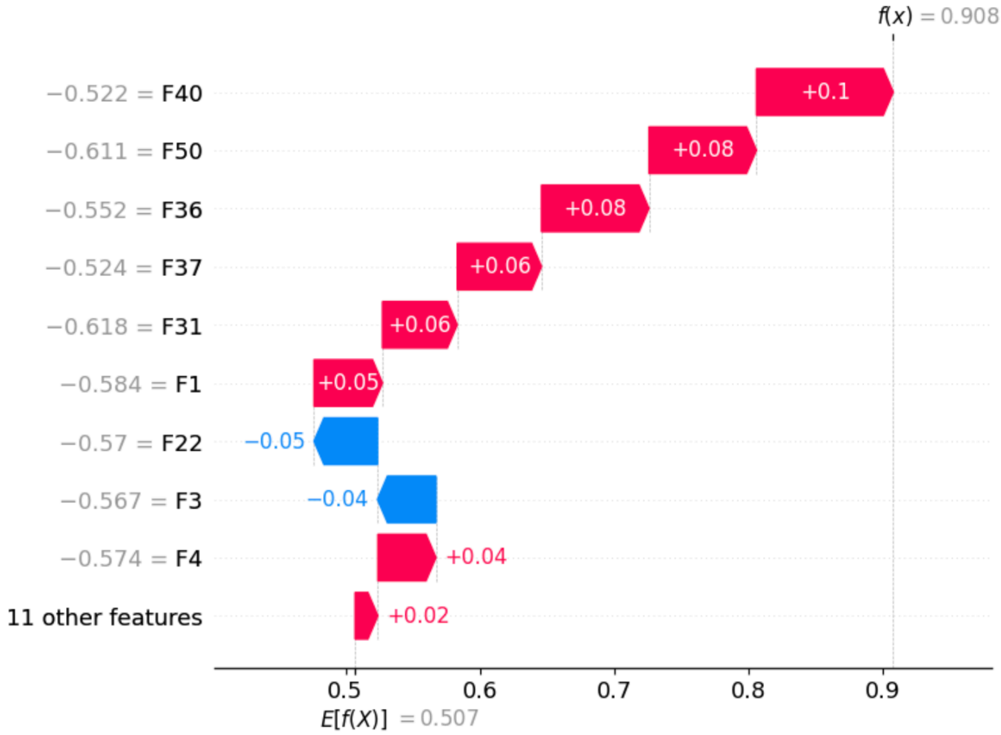

Figure 8.1: SHAP prediction outcome based on all features.

Similarly, we picked an instance of a construction project (a row from the test dataset) to interpret/ explain how the chosen model made its prediction decision. Figure 8.2 offers insights into the factors influencing the prediction of potential delays in the construction project of choice. The plot presents a visualization of the impact of various features (denoted as F36, F12, F22, F37, F14, F8, F1, F27, and F2) on the final prediction outcome, with distinct coloration highlighting different trends in the feature space. A base value of 0.5 suggests a neutral starting point for the prediction process. The final prediction of 0.01 indicates a propensity for minimal delays in that construction project. In the red-coloured region, features such as F36, F12, and F22 exert a notable positive influence on the prediction outcome. Specifically, F36 is associated with a value of 0.93, indicating its substantial contribution to the likelihood of project delays. Similarly, F12 and F22 demonstrate positive effects on the prediction outcome, with values of 1.58 and -0.39, respectively. These features, when observed in the red-coloured region, suggest a higher probability of project delays in scenarios where they exhibit significant values. Conversely, the blue-coloured region is characterized by features including F37, F14, F8, F1, F27, and F2. These features contribute negatively to the prediction outcome, indicating a lower likelihood of project delays. Notably, F8 stands out with a substantial negative effect, denoted by its value of 2.26, followed by F14 with a value of 0.89. Features F37, F1, F27, and F2 also exhibit negative effects on the prediction outcome, albeit to a lesser extent.

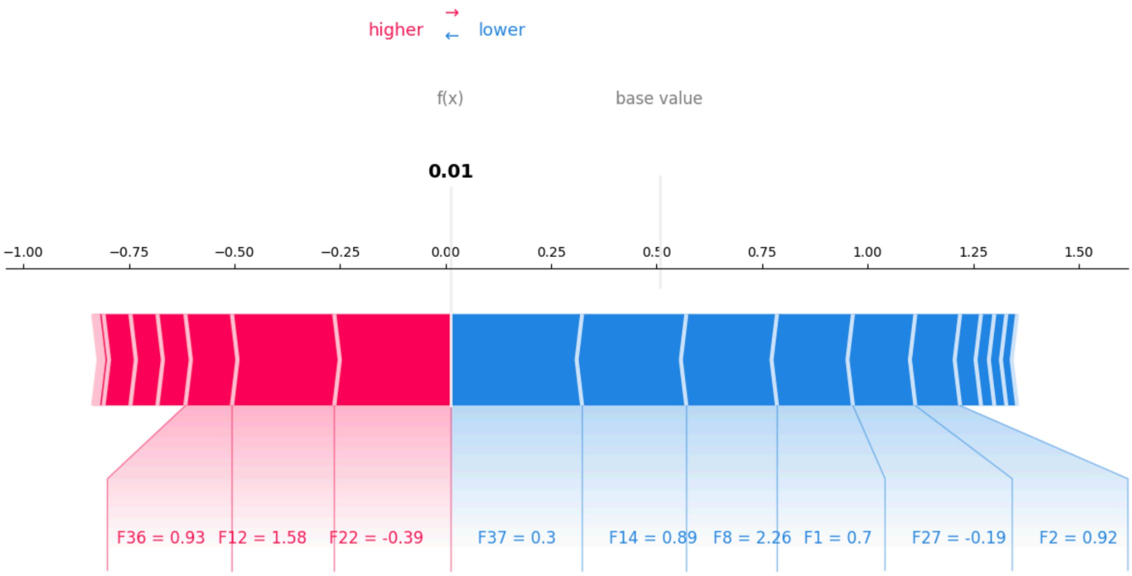

Figure 8.2: SHAP prediction outcome based on selected features.

Furthermore, Figure 8.3 showcases the top five factors (features), namely ***late payment by the owner (F36), inaccurate resource planning (F40), space limitations at site for permanent equipment (F22), reworks due to error in construction (F3), and unskilled labourer (F37),*** arranged in descending order of importance on the left-hand side of the y-axis from the array of the individual twenty factors(features) used to develop the AI chosen AI predictive model (FCNN) for the different project type (see Table 7.6). Each feature's impact on the model output, represented by SHAP values, is depicted along the x-axis, ranging from -0.4 to 0.4. A key aspect of the visualization is the vertical divide at the centre, marked by a value of 0 on the x-axis. This division shows the influence of features on the model output into two distinct regions: negative SHAP values to the left and positive SHAP values to the right. The negative region signifies features that contribute to lower model predictions, while the positive region indicates features associated with higher predictions. The colour gradient from blue to red serves as a legend, symbolizing the magnitude of feature impact. Blue hues represent lower SHAP values, indicating features with minimal influence on model output. In contrast, red hues signify higher SHAP values, indicating features with significant impacts on model predictions. Therefore, it is evident that the top five features possess varying degrees of influence on the model output in comparison to the other fifteen features. These top five features were determined through the application of SHapley Additive exPlanations (SHAP) values, which quantify the contribution of each feature to the model's predictions. These features were chosen based on their overall SHAP values, which indicate the magnitude of their influence on the predictive model's output. F36 stand out as the most impactful feature, as it exhibits the highest SHAP values, signifying a substantial contribution to model predictions. Conversely, from the pool of the other fifteen features such as F10, F28, and F2 demonstrate comparatively lower SHAP values, suggesting a lesser impact on the model output. Through these interpretability tool, construction stakeholders can gain a deeper understanding of the FCNN model's decision-making process, identify influential features, and assess the model's performance comprehensively. Thus, enhances transparency, foster trust, and enable construction companies make informed decisions based on the insights gleaned from their machine learning models.

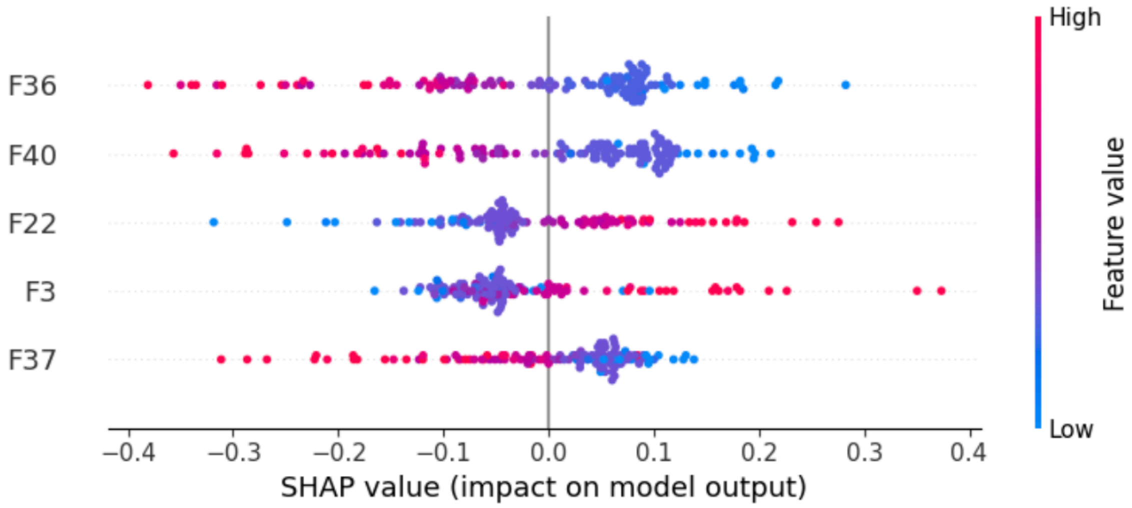

Figure 8.3: SHAP prediction outcome based on top five features.

8.3.1 Late Payment by the Owner

The analysis conducted using the SHAP technique has identified late payment by the owner as the most critical factor contributing to project delays across both BIM-based and non-BIM-based construction projects. This finding aligns with numerous previous studies that have consistently highlighted the detrimental impact of delayed payments on project progress and timelines. Delayed payments from owners disrupt the financial flow essential for construction projects, causing a ripple effect that reverberates throughout the entire supply chain. Contractors and subcontractors heavily rely on timely payments to sustain operations, procure materials, retain skilled labour, and meet their financial obligations. When payments are delayed, it can lead to a cascading series of consequences, including work stoppages, resource shortages, and potentially, legal disputes. Moreover, late payments can strain the relationships between stakeholders, eroding trust and fostering an adversarial environment that hinders collaboration and effective problem-solving. This strain can further exacerbate delays as parties become entrenched in their positions, prioritizing self-preservation over project success. It is crucial to acknowledge that late payments by owners are often symptomatic of deeper underlying issues, such as inadequate project financing, cash flow mismanagement, or contractual disputes. Addressing this critical factor necessitates a multifaceted approach that involves improving communication channels, establishing clear payment schedules, and implementing robust contract management practices. Owners must recognize the far-reaching implications of delayed payments and prioritize prompt and efficient payment processes. This commitment not only fosters a positive working environment but also contributes to the overall success and timely completion of construction projects, thereby minimizing the risk of costly delays. The detrimental impact of late payments by owners on project timelines resonates with findings from various studies in the construction industry. Amoatey et al. (2021) identified delayed payments as a critical factor hindering the delivery of construction projects in Ghana, often leading to work stoppages and project abandonment. Similarly, Hing et al. (2023) investigated the causes of delays in Malaysian construction projects and ranked late payment by the client as the most significant factor, corroborating the results of this study. Prompt payment mechanisms, such as project bank accounts or third-party payment certification systems, have been proposed as potential solutions to mitigate the risk of late payments (Wang et al., 2024). However, these measures require a collaborative effort from all stakeholders, including owners, contractors, and regulatory bodies, to establish industry-wide standards and enforce compliance. It is noteworthy that the impact of late payments may be more pronounced in non-BIM-based projects, as the lack of integrated digital workflows and real-time project visibility can exacerbate communication gaps and delays in dispute resolution. This underscores the potential benefits of adopting BIM to enhance transparency, streamline payment processes, and facilitate more efficient collaboration among project stakeholders.

8.3.2 Inaccurate Resource Planning

Inaccurate resource planning has emerged as the second most influential factor contributing to project delays, according to the SHAP analysis. Effective resource planning is a cornerstone of successful project execution, encompassing the strategic allocation and management of materials, equipment, and labour resources throughout the project lifecycle. Inadequate resource planning can manifest in various forms, such as underestimating material quantities, failing to account for lead times, or miscalculating labour requirements. These shortcomings can lead to critical resource shortages, causing work stoppages, rework, and ultimately, project delays. Moreover, inaccurate resource planning can create a ripple effect that extends beyond the immediate project site. For instance, insufficient planning for material procurement can strain supply chains, leading to delays in material delivery and disrupting the schedules of other projects relying on the same suppliers. Effective resource planning requires a comprehensive understanding of project requirements, accurate forecasting of resource needs, and proactive risk management strategies. It necessitates close collaboration between project managers, contractors, and suppliers to ensure seamless coordination and timely resource availability. Implementing robust resource planning practices, such as detailed scheduling, material tracking systems, and contingency planning, can mitigate the impact of inaccurate resource planning on project timelines. Additionally, leveraging data-driven analytics and predictive modelling can further enhance resource forecasting accuracy, enabling more informed decision-making, and reducing the risk of project delays. The criticality of accurate resource planning in mitigating project delays aligns with findings from numerous studies in the construction domain. Doloi et al. (2022) identified resource shortages and inadequate resource planning as significant contributors to cost overruns and project delays in the Indian construction industry. Similarly, Gunduz et al. (2023) highlighted the importance of effective resource management in reducing delays and improving project performance in the Turkish construction sector. Emerging technologies, such as cloud-based resource planning software and real-time monitoring solutions, offer promising avenues for enhancing resource forecasting accuracy and optimizing resource allocation (Alaloul et al., 2020). By leveraging these tools, project teams can gain greater visibility into resource availability, demand patterns, and potential bottlenecks, enabling proactive decision-making and minimizing the risk of delays. Furthermore, the integration of BIM with resource planning and scheduling tools can facilitate more seamless collaboration and information exchange among project stakeholders (Azhar, 2024). This approach can lead to improved coordination, reduced rework, and more efficient utilization of resources, ultimately contributing to timely project completion.

8.3.3 Space Limitations at Site for Permanent Equipment

The SHAP analysis has identified *space limitations at the construction site for permanent equipment* as the third most critical factor contributing to project delays. This finding underscores the importance of effective site planning and management in ensuring smooth project execution. Construction sites are inherently dynamic environments, with various activities occurring simultaneously, often within confined spaces. Inadequate planning for the placement and movement of permanent equipment, such as cranes, heavy machinery, and storage facilities, can lead to significant bottlenecks and inefficiencies. Space limitations can cause congestion, hampering the flow of materials and personnel, and ultimately impeding progress. Furthermore, improper equipment placement can increase the risk of accidents, potentially leading to work stoppages and further delays. Addressing space limitations at construction sites requires a strategic approach that involves detailed site layout planning, effective communication among stakeholders, and proactive conflict resolution. Project managers must collaborate closely with contractors, equipment suppliers, and site supervisors to develop comprehensive site utilization plans that optimize space allocation while ensuring safety and productivity. Implementing innovative solutions, such as modular construction techniques, off-site prefabrication, and just-in-time material delivery, can help alleviate space constraints and streamline site operations. Additionally, leveraging technologies like BIM and virtual reality simulations can aid in visualizing and analysing site layouts, enabling proactive identification and resolution of potential space conflicts. By proactively managing space limitations at construction sites, project teams can mitigate disruptions, enhance productivity, and ultimately reduce the risk of project delays, ensuring timely completion and adherence to project timelines. The significance of space limitations at construction sites as a critical factor contributing to project delays aligns with findings from several studies across various regions. Gunduz et al. (2013) identified site layout and space constraints as major causes of delays in Turkish construction projects, emphasizing the need for effective site planning and management. Notably, the impact of space limitations may be more pronounced in urban or densely populated areas, where construction sites are often confined and subject to stringent regulations (Onosede, 2019). In such scenarios, innovative approaches such as modular construction and off-site fabrication can provide viable solutions to mitigate space constraints and reduce on-site activities. Additionally, the use of advanced technologies like BIM and virtual reality simulations can significantly aid in optimizing site layouts and identifying potential spatial conflicts before construction commences (Zaki et al., 2021). By visualizing and analyzing the site environment digitally, project teams can make informed decisions regarding equipment placement, material staging areas, and temporary facilities, ultimately minimizing disruptions and delays caused by space limitations.

8.3.4 Reworks Due to Error in Construction

Reworks, which involve the correction or replacement of defective or non-conforming work, can have far-reaching consequences on project timelines and budgets. Reworks can arise from various sources, including design errors, poor workmanship, inadequate quality control measures, or miscommunication among stakeholders. Regardless of the root cause, reworks often necessitate the allocation of additional resources, such as materials, labour, and equipment, thereby disrupting the project schedule and potentially delaying critical path activities. The impact of reworks on project delays is well documented in the construction literature. Love et al. (2018) investigated the causes of rework in Australian construction projects and found that it contributed significantly to cost overruns and schedule delays. Similarly, Han et al. (2019) identified rework as a major factor impeding the timely completion of construction projects in China, highlighting the need for improved quality management practices. Addressing reworks due to errors in construction requires a multifaceted approach that encompasses effective quality control measures, enhanced communication channels, and a culture of continuous improvement. Implementing robust quality assurance programs, with clearly defined procedures and rigorous inspections, can help identify and rectify deficiencies early in the construction process, minimizing the need for extensive reworks and subsequent delays. Furthermore, the adoption of BIM can play a pivotal role in mitigating reworks by facilitating better coordination among stakeholders, enabling clash detection, and ensuring compliance with design specifications (Egwim, 2021). BIM also enables the creation of detailed as-built models, which can aid in future maintenance and renovation activities, reducing the likelihood of reworks in subsequent project phases. Effective communication and collaboration among project stakeholders are crucial in preventing reworks and minimizing their impact on project timelines. Regular meetings, clear lines of communication, and a culture of accountability can foster a shared understanding of project requirements, reduce miscommunication, and facilitate prompt resolution of issues that may lead to reworks. It is also essential to foster a continuous improvement mindset within the construction industry, where lessons learned from previous projects are systematically captured and incorporated into future practices. By analyzing the root causes of reworks and implementing corrective measures, project teams can proactively mitigate the risk of delays and enhance overall project performance.

8.3.5 Unskilled Labourer

The construction industry heavily relies on skilled labour to execute intricate tasks and ensure adherence to quality standards, making the availability and competency of the workforce a crucial determinant of project success. Unskilled or inadequately trained labourers can introduce various challenges to construction projects, including poor workmanship, safety violations, and inefficient use of materials and resources. These issues can lead to reworks, schedule disruptions, and ultimately, project delays, as corrective measures need to be implemented to rectify deficiencies. The impact of unskilled labour on project delays has been extensively documented in construction literature. Gunduz et al. (2013) identified the shortage of skilled labour as a significant factor contributing to delays in Turkish construction projects, emphasizing the need for effective workforce planning and training programs. Similarly, Alaghbari et al. (2019) highlighted the importance of skilled labour in mitigating delays and ensuring timely project completion in the construction industry of Yemen. Addressing the issue of unskilled labour requires a multifaceted approach that encompasses workforce development, training initiatives, and effective human resource management practices. Construction firms and project managers should prioritize the recruitment and retention of skilled workers, offering competitive compensation packages and fostering an environment that values continuous learning and skill development. Collaboration with vocational training institutions and apprenticeship programs can help cultivate a skilled workforce tailored to the specific demands of the construction industry. Additionally, leveraging emerging technologies, such as virtual reality simulations and augmented reality tools, can enhance training effectiveness and accelerate skill acquisition among workers (Eiris & Gheisari, 2018). Furthermore, the adoption of lean construction principles and modular construction techniques can help mitigate the impact of unskilled labour by streamlining processes, reducing on-site activities, and minimizing the reliance on skilled labour for certain tasks (Sheikhkhoshkar et al., 2019). Effective project planning and resource allocation are also crucial in addressing the challenge of unskilled labour. By accurately forecasting labour requirements and allocating tasks based on skill levels, project managers can optimize resource utilization and minimize the risk of delays caused by unskilled workers performing critical tasks beyond their capabilities. It is essential for construction firms and project stakeholders to recognize the significance of investing in a skilled and competent workforce. By prioritizing workforce development and implementing effective strategies to address the issue of unskilled labour, the construction industry can mitigate project delays, enhance quality, and ultimately deliver projects on time and within budget.

8.4 Implication to Theory

The findings of this research, which shed light on the critical factors contributing to project delays in the construction industry, have significant implications for the theoretical underpinnings (established in section 4) that endeavour to explain and address such challenges. The identification of late payment by the owner, inaccurate resource planning, space limitations at the site for permanent equipment, reworks due to errors in construction, and the presence of unskilled labourers as the top influential factors unveils a multifaceted and complex landscape that demands a holistic theoretical perspective. At the outset, the Optimism Bias Theory, proposed by Macdonald (2002) (see section 4), offers a compelling explanation for the underestimation of project costs and durations, which can contribute to delays. However, while optimism bias may play a role in the initial planning stages, the factors identified in this research extend beyond mere underestimation and encompass deeper systemic issues within the construction industry. Conversely, the Innovation Diffusion Theory, as articulated by Rodgers (2003), provides a more comprehensive theoretical framework for understanding and addressing the root causes of project delays. This theory recognizes that the adoption and diffusion of innovative practices, technologies, and processes within a social system, such as the construction industry, are crucial for mitigating challenges and enhancing overall performance. The critical factors identified in this research, such as late payment by the owner, inaccurate resource planning, space limitations, reworks, and the presence of unskilled labourers, are intrinsically linked to the adoption and diffusion of innovative solutions. For instance, the adoption of BIM can facilitate better coordination among stakeholders, enhance transparency in payment processes, optimize site layouts, and reduce the likelihood of reworks. Similarly, addressing issues such as inaccurate resource planning and the presence of unskilled labourers may require the adoption of innovative practices, such as cloud-based resource planning software, real-time monitoring solutions, virtual reality simulations for training, and lean construction principles. The Innovation Diffusion Theory provides a framework for understanding how these innovative solutions are adopted and disseminated within the construction industry. Factors such as relative advantage, compatibility, complexity, trialability, and observability of an innovation can influence its rate of adoption (Rodgers, 2003). By examining these factors and the diffusion process within the construction sector, stakeholders can develop strategies to accelerate the adoption of innovative practices and technologies that can effectively mitigate project delays. Furthermore, the Innovation Diffusion Theory acknowledges the importance of communication channels, social structures, and time in the diffusion process. Addressing project delays may require effective communication and collaboration among stakeholders, as well as a long-term commitment to fostering a culture of continuous improvement and innovation within the construction industry. In light of these considerations, the Innovation Diffusion Theory emerges as the more pertinent theoretical framework for understanding and addressing the critical factors contributing to project delays in the construction industry. By embracing this theory, stakeholders can develop targeted strategies to facilitate the adoption and diffusion of innovative solutions, ultimately enhancing project performance and mitigating the risk of delays.

8.5 Practical Implications of the findings for construction project management

Given the significant role of the identified delay factors in construction projects, it is imperative for project managers and stakeholders to adopt proactive strategies that address these issues. Key practical implications include:

1. **Implementation of Financial Management Tools:** To mitigate the issue of late payments, project managers can employ financial management systems that ensure timely payments from owners, such as the use of project bank accounts or third-party certification systems. These mechanisms can help maintain cash flow and prevent disruptions in the construction process.
2. **Enhanced Resource Planning:** The accuracy of resource planning can be significantly improved by integrating advanced technologies like cloud-based planning software and BIM.

These tools provide real-time data on resource availability and requirements, enabling more precise forecasting and reducing the likelihood of resource shortages that could lead to delays.

3. **Quality Management and Training Programs:** To address the frequent occurrence of reworks due to construction errors, it is essential to establish robust quality assurance programs. These programs should include comprehensive training for workers to enhance their skills and adherence to quality standards. Additionally, employing BIM can facilitate better communication and coordination among project stakeholders, reducing errors and the need for reworks.
4. **Risk Management:** Proactive risk management practices, including regular risk assessments and the development of contingency plans, are critical. By anticipating potential issues related to resource allocation, payment delays, and quality control, project teams can implement preventive measures to minimize their impact on project timelines.

The impact of delay factors can vary significantly depending on contextual factors such as project size and geographical location. For instance, larger projects are often more complex and may require more sophisticated management systems to coordinate multiple teams and resources effectively. In contrast, smaller projects may be more sensitive to issues like resource shortages or delays in material delivery. Geographical location also plays a crucial role. For example, projects in remote areas may face additional challenges such as longer lead times for material delivery, limited availability of skilled labour, and logistical constraints. Furthermore, regulatory and cultural differences can influence project execution and stakeholder interactions, potentially exacerbating or mitigating delay factors. Furthermore, the identified delay factors are not isolated, they often exhibit interdependencies that can compound their effects on project timelines. For example, late payments can lead to resource shortages, as contractors may lack the funds to procure necessary materials or retain skilled labour. This, in turn, can cause further delays as work is halted or slowed down. Similarly, inaccurate resource planning can result in both resource shortages and poor-quality control, leading to reworks. Reworks not only consume additional resources but also extend project timelines, which can increase costs and strain relationships among stakeholders. To effectively manage these interdependencies, project managers should adopt a holistic approach that considers the potential ripple effects of each factor. This includes developing integrated project plans that align financial, resource, and quality management strategies, thereby minimizing the cumulative impact of delays.

8.6 Chapter Summary

This chapter explored the findings of the research, presenting a comprehensive analysis of the critical factors contributing to project delays in the construction industry. Through the application of the SHAP technique, the Fully Connected Neural Network (FCNN) emerged as the optimal AI predictive model, exhibiting exceptional accuracy, precision, recall, and F1 scores across both BIM-based and non-BIM-based projects. The application of SHapley Additive exPlanations (SHAP), a powerful technique for model interpretability, unveiled the top five influential factors driving project delays. These factors included late payment by the owner, inaccurate resource planning, space limitations at the site for permanent equipment, reworks due to errors in construction, and the presence of unskilled laborers. Detailed discussions on each of these critical factors highlighted their far-reaching implications and the urgent need to address them. Late payment by the owner was identified as the most significant factor, disrupting financial flows and straining stakeholder relationships. Inaccurate resource planning, on the other hand, can lead to resource shortages, work stoppages, and supply chain disruptions. Space limitations at construction sites can cause congestion, hampering productivity and increasing safety risks. Reworks due to errors in construction not only consume additional resources but also delay critical path activities. Finally, the presence of unskilled labourers can introduce challenges such as poor workmanship, safety violations, and inefficient use of resources, ultimately impacting project timelines. To mitigate these factors and address project delays effectively, the research underscored the importance of adopting innovative solutions and practices within the construction industry. The Integration of BIM, cloud-based resource planning software, real-time monitoring solutions, virtual reality simulations for training, and lean construction principles emerged as promising avenues for enhancing project performance and reducing delays. Furthermore, the findings of this research have significant implications for theoretical underpinnings. While the Optimism Bias Theory offers insights into the underestimation of project costs and durations, the Innovation Diffusion Theory provides a more comprehensive framework for understanding and addressing the root causes of project delays. This theory recognizes the crucial role of adopting and disseminating innovative practices, technologies, and processes within the construction industry, and highlights the importance of factors such as relative advantage, compatibility, complexity, trialability, and observability in influencing the rate of adoption.

CHAPTER 9

9.0 CONCLUSIONS AND RECOMMENDATION

9.1 Chapter introduction

This chapter marks the end of the research journey, bringing together the main findings and drawing conclusions. Section 9.2 summarizes the findings thoroughly, organized by the study's objectives, along with the conclusions drawn from them. Subsections 9.2.1 through 9.2.4 break down the findings and conclusions related to the four main objectives of the research. It also highlights the significant impact of the research on both academic and industry fronts. Section 9.3.1 explains how the work contributes to academic knowledge, while Section 9.3.2 discusses its practical implications for the construction industry. Acknowledging that every study has its limitations, Section 9.4 honestly discusses the constraints faced. In addition, Section 9.5 identifies areas for future research, opening doors for further exploration in this field. Lastly, Section 9.6 sums up the chapter concisely, reminding us of the key points and emphasizing the importance of our findings and conclusions.

9.2 Findings Overview

9.2.1 First Objective: *Conduct a systematic review toward gathering the most common factors affecting construction project delays and use it to conduct survey of experts to establish the most applicable factors affecting construction project delays in BIM-based construction projects.*

This objective was accomplished through a rigorous and systematic approach, adhering to the guidelines set forth by the PRISMA (see section 2.1). The systematic review process entailed an exhaustive search and analysis of relevant scholarly articles, encompassing a diverse array of construction project types spanning various geographical regions worldwide. This process yielded a comprehensive conceptual framework showing the critical determinants of project delays, thereby serving as an indispensable prerequisite for the subsequent development of AI predictive models aimed at quantifying delay risks across both BIM-based and traditional non-BIM construction projects. Through this analytical process, nine distinct categories of delay risk drivers were identified and systematically examined: owner-related, contractor-related, consultant-related, design-related, labour-related, equipment-related, project-related, supplier-related, and external-related drivers. The systematic review revealed that contractor-related and external-related drivers were cited with the highest frequency as the most critical factors contributing to delays in BIM-based construction projects. These findings underscore the paramount significance of these two categories in the formulation of effective delay mitigation strategies within BIM implementation in construction projects. Contractor-related drivers, encompassing factors such as inadequate management, poor planning and scheduling, and lack of experience or expertise, were repeatedly highlighted as significant contributors to project delays across the reviewed literature. Similarly, external-related drivers, including unfavourable weather conditions, regulatory changes, and economic factors, were identified as crucial elements that can significantly impact project timelines and exacerbate delays. Concurrently, the systematic review process also highlighted the potential positive impact of BIM implementation on mitigating delays in construction projects. Several studies included in the review emphasized the ability of BIM to enhance coordination, communication, and collaboration among project stakeholders, as well as its capacity to facilitate improved planning, visualization, and decision-making processes. These findings suggest that the effective adoption and utilization of BIM can play a pivotal role in reducing the risk of delays and enhancing overall project performance. To further substantiate and contextualize the findings from the systematic review, an expert survey was conducted, engaging industry professionals and subject matter experts in the construction industry. This step ensured that the research was firmly grounded in real-world insights and industry expertise, enabling the development of a robust and industry-relevant framework for assessing delay risks. The expert survey served to validate and refine the identified delay risk drivers, providing valuable insights regarding their applicability and relative importance in construction projects. Through this integration of the systematic review's literature-based framework with the practical insights and expertise derived from the expert survey, an empirically validated framework emerged (see section 7.4). This validated framework not only corroborates the initial conceptual framework formulated (see section 2.3) but also fortifies the theoretical foundations underpinning this study. The systematic identification and integration of the critical delay risk drivers, coupled with the insights gathered from industry experts, provided a robust basis for the subsequent stages of this study.

9.2.2 **Second Objective:** *Conduct a systematic review of AI in the construction industry and use it to establish the most appropriate AI technologies during construction.*

The findings of this systematic review have made significant contributions to the body of knowledge and have informed the subsequent selection and implementation of AI technologies for predicting potential delays in both BIM and non-BIM-based construction projects. Through the analysis of 70 selected studies, the review revealed a surge in research interest and publications concerning the adoption of AI technologies in the construction industry, particularly in recent years. This trend emphasizes the industry's recognition of the transformative potential of AI in addressing longstanding challenges and driving efficiency and productivity gains. Notably, the review identified seven major AI technology types documented in the literature, with supervised learning, and deep learning emerging as the most prominent and influential AI technologies in the construction industry. Crucially, the systematic review provided, insights into the applicability of these AI technologies across the three major stages of the construction project lifecycle: planning/design, construction/execution, and supply/facility management. Furthermore, the review highlighted the potential benefits of implementing AI technologies in the construction industry, with the potential for design expansion emerging as a key advantage cited by most of the selected literature. By harnessing the power of AI, designers and architects can explore previously unimaginable design alternatives, optimizing for various objectives and constraints while fostering creativity and innovation. Additionally, the review underscored the opportunities presented by AI technologies for data augmentation, model generalizability, real-world applicability, and the integration of cutting-edge computer vision and augmented reality techniques, paving the way for technological advancements in the industry. Building upon the comprehensive insights gained from the systematic review, the subsequent stage of this research focused on employing the two most applicable AI technologies identified – supervised learning and deep learning – to develop AI predictive models for assessing delay risks in the project types. The selection of these two AI technologies was driven by their prominence in the literature, their suitability for research aim, and their ability to leverage the available expert survey data to uncover patterns and make informed predictions. Among the various supervised learning and deep learning models evaluated, the FCNN emerged as the optimal choice, exhibiting exceptional performance metrics in terms of accuracy, precision, recall, and F1 scores. The FCNN's superior performance was further bolstered by its inherent architectural characteristics, which lend themselves well to the complexities of the delay prediction task. To enhance model interpretability and foster trust among construction stakeholders, the research employed SHAP, a powerful technique that explains the underlying factors driving model predictions. By illustrating the top contributing features, such as late payment by the owner, inaccurate resource planning, space limitations at the site, reworks due to construction errors, and unskilled labour, the FCNN model's decision-making process becomes more transparent and comprehensible to construction stakeholders, facilitating informed decision-making and enabling targeted interventions to mitigate delay risks.

9.2.3 **Third Objective:** *Utilize the applicable drivers in the first objective as independent features (variables) to develop hyperparameter optimised AI predictive model(s) established in the second objective*.

The third objective of this research was to leverage the applicable delay risk drivers identified in the first objective as independent features and employ the AI technologies established as most suitable through the systematic review in the second objective, to develop hyperparameter-optimized AI predictive models. This objective was successfully accomplished through a comprehensive process that encompassed feature selection, model development, hyperparameter optimization, and rigorous evaluation. The feature selection process was guided by the validated framework of delay factors essential for both BIM and non-BIM construction projects. This framework, derived from the first objective, provided a robust set of twenty independent variables, encompassing critical factors such as late payment by the owner, inaccurate resource planning, space limitations at the site, reworks due to construction errors, and unskilled labour. The inclusion of these factors as input features ensured that the AI models were trained on a comprehensive and relevant set of variables, capturing the intricate interplay of various delay risk drivers and their impact on project timelines. Drawing upon the insights from the second systematic review, which identified supervised learning and deep learning as the most applicable AI technologies for the construction value chain lifecycle, a range of models were developed and evaluated. These included classical algorithms, ensemble methods, and various neural network architectures, such as the Multi-Layer Perceptron (MLP) and the Radial Basis Function Network (RBFN). To achieve optimal performance, a precise hyperparameter optimization process was undertaken for each model. This iterative process involved fine-tuning the models' architectural parameters, learning rates, regularization techniques, and other hyperparameters, ensuring that they were tailored to the specific characteristics of the dataset and the research aim. The evaluation of the models' performance was conducted using a set of performance evaluation metrics, including accuracy, precision, recall, and F1 score. These metrics were essential in assessing the models' ability to handle imbalanced datasets and make accurate predictions, particularly as touching project delay risk assessment. Among the evaluated models, the FCNN emerged as the optimal choice, exhibiting exceptional performance across all metrics for both project types. The FCNN's superior performance can be attributed to its inherent capability to learn complex non-linear relationships, its ability to handle imbalanced datasets effectively, and the successful hyperparameter optimization process. The FCNN model demonstrated remarkable accuracy scores of 93% and 79% for BIM-based and non-BIM-based projects, respectively, outperforming all other models evaluated in this research. Additionally, its precision and recall values were consistently high, indicating its proficiency in correctly identifying instances of potential delays while minimizing false positive and false negative predictions. To further enhance the interpretability and explainability of the FCNN model's predictions, the research employed SHAP. This powerful technique provided valuable insights into the underlying factors driving the model's predictions by visualizing the top contributing features and their respective impacts. For instance, the SHAP analysis revealed that factors such as late payment by the owner, inaccurate resource planning, space limitations at the site, reworks due to construction errors, and unskilled labour were among the most significant contributors to potential project delays. The successful development and validation of the hyperparameter-optimized FCNN model represent a significant achievement in this research. By leveraging the power of AI and the identified critical delay factors, this model has the potential to revolutionize the way project delays are anticipated and mitigated in the construction industry. Furthermore, the integration of SHAP has significantly enhanced the transparency and interpretability of the model's predictions, fostering trust amongst construction stakeholders, and enabling targeted interventions to address the most influential delay risk drivers.

9.3 Contributions of study

9.3.1 Contribution of Study to Academic Knowledge

This research has made significant contributions to academic knowledge in several ways. Firstly, the systematic review conducted on identifying the most common drivers affecting construction project delays represents a comprehensive and up-to-date synthesis of existing literature on this topic. By rigorously adhering to the PRISMA guidelines, the review process ensured a thorough and unbiased exploration of delay factors across diverse construction project types and geographical regions. The resulting conceptual framework, encompassing nine distinct categories of delay risk drivers, serves as a valuable foundation for future research in this domain, providing a robust starting point for further investigation and validation. Moreover, the integration of industry expertise through the expert survey adds a crucial practical dimension to the study, bridging the gap between academic research and real-world industry practices. This approach not only validates the findings from the systematic review but also ensures that the resulting framework is grounded in the realities and nuances of the construction industry. By incorporating the perspectives of subject matter experts, the study enhances the relevance and applicability of its findings, fostering a stronger connection between academic research and practical implementation. The systematic review on the application of AI technologies in the construction industry represents another significant contribution to academic knowledge. By rigorously analyzing vast body of academic literature, the review provides a comprehensive and up-to-date understanding of the current state of AI adoption in the construction value chain lifecycle. The identification of seven major AI technology types, with supervised learning and deep learning emerging as the most prominent, offers valuable insights for researchers and practitioners alike. Furthermore, the review's exploration of the applicability of these AI technologies across the three major stages of the construction project lifecycle (planning/design, construction/execution, and supply/facility management) provides a roadmap for future research and technological advancements in the industry. The development of hyperparameter-optimized AI predictive models for assessing delay risks in both BIM-based and non-BIM-based construction projects represents a ground-breaking contribution to academic knowledge such that has never been done before in one study. By leveraging the identified critical delay factors as input features and employing the most suitable AI technologies (supervised learning and deep learning), this study has successfully created a robust and accurate predictive model. The FCNN model, which emerged as the optimal choice, exhibits exceptional performance metrics. This model not only demonstrates the potential of AI in addressing complex challenges in the construction industry but also serves as a blueprint for future research in developing AI-driven solutions for risk assessment and project management. The integration of SHAP into the predictive model further enhances the study's contribution to academic knowledge. By providing a transparent and interpretable explanation of the model's predictions, SHAP addresses the long-standing issue of "black box" models in AI, fostering trust and understanding among researchers and construction stakeholders Consequently, this study has made significant strides in advancing academic knowledge in the fields of construction project management, delay risk assessment, and AI applications in the construction industry.

9.3.2 Contribution of Study to Practice

This research makes substantial contributions to the practice of construction project management, particularly in the realm of delay risk assessment and mitigation. By identifying and validating the critical drivers of construction project delays through a systematic review and expert survey, the study provides construction professionals with a comprehensive and industry-relevant framework for understanding and addressing these challenges. The developed hyperparameter-optimized AI predictive model, specifically the FCNN, represents a significant practical contribution to the construction industry. This model offers construction stakeholders a powerful tool for assessing the risk of potential delays in both

BIM-based and non-BIM-based construction projects. The model's exceptional performance metrics, including accuracy scores for BIM-based and non-BIM-based projects, respectively, demonstrate its reliability and effectiveness in predicting potential delays. The integration of SHAP into the predictive model further enhances its practical value by providing transparency and interpretability to the model's predictions. Construction professionals can gain valuable insights into the specific factors contributing to potential delays, such as late payment by the owner, inaccurate resource planning, space limitations at the site, reworks due to construction errors, and unskilled labour. This information empowers construction stakeholders to develop targeted strategies and interventions to mitigate these critical delay risk drivers proactively. Also, the predictive model's ability to assess delay risks early in the project lifecycle offers construction professionals a significant advantage in project planning and resource allocation. By identifying potential bottlenecks and high-risk areas before they manifest, construction teams can implement preventive measures, optimize resource allocation, and develop contingency plans to minimize the impact of delays on project timelines and budgets. Furthermore, the study's contribution extends beyond individual construction projects. The validated framework of delay risk drivers and the predictive model's capabilities can be leveraged by construction firms and organizations to enhance their overall project management practices. By integrating these tools into their project management processes, companies can improve their risk management strategies, enhance decision-making processes, and foster a culture of proactive delay mitigation. The practical implications of this research also extend to the broader construction industry ecosystem. Government agencies, regulatory bodies, and policymakers can utilize the findings to develop guidelines, best practices, and industry standards for delay risk assessment and mitigation. By incorporating these insights into industry regulations and certification processes, the construction industry can collectively promote transparency, accountability, and a proactive approach to addressing project delays. Moreover, the study's findings can inform the development of training programs and educational curricula for construction professionals. By incorporating the identified delay risk drivers and the predictive model's capabilities into professional development programs, construction companies can better equip their workforce with the knowledge and tools necessary to tackle project delays effectively.

9.4 Limitations of study

While this study has made significant contributions to both academic knowledge and industry practice, it is essential to acknowledge and address its inherent limitations. One of the primary limitations lies in the scope and representativeness of the data used for model development and validation. The expert survey data, although valuable, may not fully capture the diverse range of construction projects, geographic regions, and industry practices. Consequently, the generalizability of the findings and the predictive model's performance might be limited when applied to contexts outside the scope of the data. Another potential limitation is the dynamic nature of the construction industry itself. The factors influencing project delays are subject to continuous evolution, as new technologies, regulations, and industry practices emerge. This study's findings, while comprehensive at the time of research, may require periodic updates and refinements to remain relevant and accurately reflect the changing landscape of the construction industry. Furthermore, the study's reliance on expert judgements and systematic reviews, although rigorous and systematic, may inadvertently introduce biases or overlook emerging trends or factors that have not yet been widely documented or recognized within the industry. This limitation highlights the need for continuous monitoring and incorporation of new insights as they become available. It is important to note that the FCNN predictive model, despite its impressive performance, is not infallible. Like any machine learning model, it is subject to the limitations of the training data and the assumptions inherent in its architecture and optimization process. Unforeseen edge cases or outliers in real-world scenarios may challenge the model's predictive capabilities, necessitating ongoing refinement and adaptation.

9.5 Future research opportunities

This study's contributions and limitations also pave the way for numerous future research opportunities that can further advance the field of construction project delay prediction and mitigation. One promising avenue for future research is the exploration of more diverse and extensive datasets, encompassing a wider range of construction projects, geographical regions, and industry practices. By leveraging larger and more representative datasets, researchers can develop more generalizable predictive models and uncover patterns and insights that may have been overlooked in the current study. Additionally, as new technologies and industry practices emerge, there is a need for continuous monitoring and incorporation of these developments into the predictive models. Future research could focus on developing adaptive and self-learning models that can dynamically update their parameters and decision-making processes based on new data and industry trends, ensuring that the models remain relevant and accurate over time. The integration of alternative or complementary AI techniques, such as unsupervised learning, reinforcement learning, or hybrid approaches, may yield further insights and improve the predictive capabilities of the models. Exploring these techniques could unlock new avenues for understanding and mitigating project delays, potentially leading to more robust and accurate predictions. Another promising research direction involves the integration of predictive models with decision support systems and risk management frameworks. By combining the predictive power of AI with expert knowledge and industry best practices, researchers can develop comprehensive solutions that not only identify potential delays but also provide actionable recommendations and mitigation strategies tailored to specific project contexts. Finally, future research could explore the synergies between AI-based predictive models and other emerging technologies in the construction industry, such as IoT, and blockchain. By integrating these technologies, researchers may uncover novel approaches to data collection, real-time monitoring, and collaborative decision-making, further enhancing the accuracy and applicability of predictive models in the construction domain.

9.6 Reflection on the Research Process, Challenges Faced and Lessons Learned.

The research process for this project was extensive and multifaceted, encompassing a variety of methodologies and sources to ensure a comprehensive understanding of the subject matter. It began with a thorough literature review to establish a theoretical foundation, followed by the collection and analysis of primary data through surveys with industry professionals. The systematic approach facilitated the identification of key trends and patterns in the construction market, ensuring that the findings were both robust and relevant. Several challenges were encountered during the research process. Firstly, obtaining reliable and recent data was difficult due to the rapidly changing nature of the construction industry. Moreover, gathering primary data through surveys presented logistical difficulties. Coordinating with industry professionals across various geographical locations required diligent planning and scheduling. Furthermore, the interdisciplinary nature of the research required synthesizing information from various fields such as social science and engineering, which was both time-consuming and complex. A significant lesson learned from this research is the importance of flexibility and adaptability in research design. Given the dynamic environment of the construction industry, it was crucial to remain open to adjusting methodologies and frameworks to accommodate new information and insights. Another key lesson is the value of cross-disciplinary collaboration. Engaging experts from various fields enriched the research, providing a more holistic understanding of the issues at hand. Finally, the necessity of robust data management practices was highlighted, ensuring that data collection, storage, and analysis were conducted efficiently and accurately.

9.7 Application or Variation of Findings in Different Global Perspectives

The findings of this research have broad implications that can vary significantly across different global contexts and construction markets. One of the primary factors influencing the applicability of these findings is the economic environment of the region in question. For instance, in developed markets such as North America and Europe, advanced construction technologies and sustainable practices are more likely to be adopted due to higher levels of investment and regulatory support. These regions often have robust frameworks for green building standards and incentives for innovation, making the implementation of the research findings more feasible and impactful. Conversely, in emerging markets, such as those in parts of Asia and Africa, the focus might be more on cost-effective construction solutions that address basic infrastructure needs. Here, the findings related to low-cost building materials and efficient construction techniques would be particularly relevant. However, challenges such as limited access to advanced technology, financial constraints, and less stringent regulatory environments could hinder the direct application of some of the more sophisticated aspects of the research. Cultural factors also play a significant role in how the findings might be adopted. In regions where there is a strong emphasis on traditional construction methods and materials, there might be resistance to change or a slower uptake of new practices. Educational initiatives and pilot projects could be effective strategies in these contexts to demonstrate the benefits of modern construction techniques and gradually build acceptance. Furthermore, the regulatory landscape is a crucial determinant of how the findings can be implemented globally. In countries with stringent building codes and environmental regulations, the adoption of sustainable practices and innovative construction methods is likely to be more straightforward. In contrast, in regions with less developed regulatory frameworks, advocacy and policy development efforts would be necessary to create an enabling environment for the adoption of these findings.

9.8 Impact of Research Findings on Project Delay Reduction

The potential impact of the research findings on project delay reduction is significant, with the implementation of identified best practices and innovative technologies promising measurable improvements in project timelines. Quantitative data from various studies and industry reports highlight the substantial benefits of these strategies. For instance, the adoption of Building Information Modelling has been shown to reduce project delays by up to 15% due to enhanced coordination and real-time problem-solving capabilities (Alnaser, Alsanabani and Al-Gahtani, 2023). In a study conducted by the McKinsey Global Institute, it was found that digitization of the construction sector, including the use of advanced analytics and automation, could lead to a reduction in project delays by approximately 20-30%(McKinsey, 2022). Furthermore, Lean Construction principles, which emphasize waste reduction and process efficiency, have demonstrated a reduction in project delays ranging from 10-15% in various case studies across North America and Europe (Aziz and Hafez, 2013; Bajjou and Chafi, 2020). These principles focus on improving workflow reliability and optimizing resource utilization, which are critical factors in minimizing delays. Another key finding from the research is the impact of improved project management practices. The integration of advanced project management software and tools can lead to a reduction in delays by up to 25%, as these tools facilitate better scheduling, resource allocation, and risk management (Conforto *et al.*, 2014; Demirkesen and Ozorhon, 2017). Additionally, the emphasis on early contractor involvement and collaborative planning has been associated with a 15-20% reduction in delays due to improved stakeholder alignment and decision-making efficiency (Mosey, 2014). In emerging markets, where the construction sector faces unique challenges, the implementation of cost-effective construction techniques and local material utilization has shown potential delay reductions of around 10-12% (Vasista and Jakhanwal, 2023). These strategies address common issues such as supply chain disruptions and financial constraints, thereby enhancing project timelines.

9.9 Specific Implications for Stakeholders in the Construction Industry

The implications of the findings in this study extend across various stakeholders within the construction industry, each playing a distinct yet interconnected role. Project managers, for instance, are positioned at the forefront of implementing changes to mitigate delay factors. They must prioritize the adoption of financial management systems that ensure timely payments, thus preventing the cash flow disruptions that often lead to project delays. Moreover, the integration of advanced project management software, such as cloud-based planning tools and BIM, can enhance resource planning and coordination. This technological adoption not only streamlines workflows but also improves communication among stakeholders, thereby reducing the incidence of errors and the subsequent need for reworks. Policymakers also bear significant responsibility in shaping an environment conducive to innovation and efficiency in construction. They can encourage the use of AI technologies by providing regulatory support and incentives. Such measures may include tax breaks for companies that adopt AI-driven solutions for safety monitoring and project management, or subsidies for projects demonstrating sustainable practices. Policymakers should also address the systemic issue of late payments by establishing frameworks that protect contractors' financial stability, thereby ensuring smoother project execution. For technology developers, the construction industry's increasing reliance on advanced technologies presents an opportunity to develop specialized solutions. There is a growing demand for AI tools that can predict maintenance needs, optimize resource use, and enhance safety on-site. Developers should focus on creating user-friendly applications that integrate seamlessly with existing systems, such as BIM, to maximize their adoption and effectiveness. Furthermore, the protection of data privacy and security is paramount, as AI applications often handle sensitive information. Compliance with data protection regulations is essential to maintaining stakeholder trust and protecting the integrity of construction data. Construction firms, meanwhile, must recognize the value of investing in new technologies and fostering innovation. Allocating resources towards AI-powered tools and advanced analytics can lead to significant efficiency gains and cost reductions. Firms should also consider forming partnerships with academic institutions and other industry players to stay abreast of the latest research and developments. Such collaborations can facilitate the exchange of knowledge and best practices, driving continuous improvement in project delivery. Clients and project owners, as the ultimate beneficiaries of construction projects, also have a pivotal role. Their engagement throughout the project lifecycle is crucial. Active involvement in planning and monitoring ensures that projects align with their expectations and that any potential issues are addressed proactively. Clients should demand transparency and accountability from contractors and project managers, fostering a culture of open communication that can mitigate risks and prevent delays.

9.10 Chapter Summary

This chapter represents the apex of the research journey, drawing together the key findings, conclusions, contributions, and future perspectives. It summarises the study's approach to addressing the challenges of construction project delays and the innovative AI techniques to mitigate this pervasive issue. The chapter begins by providing an overview of the findings, organized according to the study's objectives. It highlights the systematic review process that identified the most common drivers affecting construction project delays, reaching a peak in the development of a validated framework through expert surveys. Furthermore, the chapter highlights the study's systematic exploration of AI technologies in the construction industry, pinpointing supervised learning and deep learning as the most prominent and influential techniques. This insight paved the way for the development of a hyperparameter-optimized AI predictive model. The chapter then explores into the study's significant contributions to academic knowledge and industry practice. Acknowledging the limitations of the study, the chapter openly discussed the constraints faced, such as the scope and representativeness of the data, the dynamic nature of the construction industry, potential biases, and the inherent limitations of machine learning models. However, these limitations also pave the way for future research opportunities, including exploring more diverse datasets, incorporating new AI techniques, enhancing interpretability, integrating with decision support systems, and leveraging synergies with emerging technologies like IoT, and blockchain.

Appendix A:

Construction Delay Questionnaire

Dear sir/ma,

My name is Christian Nnaemeka Egwim a doctoral researcher of the University of Hertfordshire, Hatfield, Hertfordshire, United Kingdom.

This questionnaire is intended to elicit responses for research on the causes of construction project delays. Because of your construction industry background, you have been listed as a possible respondent. It would be extremely helpful if you could help me complete this questionnaire.

Thank you.

Please provide answers to the questions in this questionnaire based on ANY single construction project you have worked on in the past. Questions on the frequency of an event should be answered based on how frequent during the life of the project.

Note: All your answers should be related to this specific project you have in mind.

Responder's Detail:

Your highest qualification: ☐ HND ☐ B.Sc./B.Tech ☐ M.Sc./M.Tech ☐ PhD ☐ Others ______________
Position you have held in construction industry: ☐ Architect ☐ Quantity Surveyor ☐ Consultant ☐ Contractor
Approximate size of project you have in mind: ☐ £50k-999K ☐ £1-50mil ☐ £51-100mil ☐ £101-150mil ☐ Above £150mil.
Years of experience in construction industry: ☐ 0–5 years ☐ 6–10 years ☐ 11–15 years ☐ 16–20 years ☐ Above 20 years.
Type of construction project you have in mind: ☐ Residential ☐ Commercial (e.g., hospitals, stadium, etc.) ☐ Infrastructure (e.g., roads, railways, etc.) ☐ Others ______________
Location (e.g., country/region/continent) of the project you have in mind: United Kingdom European Union Africa North America Asia South America

☐	☐	☐	☐	☐	☐
☐ Oceania	☐ Antarctica	☐ Others ______________________________			

Was Building Information Model (BIM) used on the project?

☐ Yes ☐ No

To what level was BIM implemented in the design stage?

☐ No implementation ☐ 2D ☐ 3D ☐ 4D ☐ 5D

To what level was BIM implemented in the construction stage?

☐ No implementation ☐ consultant checks only ☐ for full collaboration between consultant and contractors

☐ for full collaboration between all involved parties ☐ for full collaboration between all involved parties with drawings and updates accessed directly on cloud

Section A:

For each of the questions below, tick the response that best describes the frequency for the following statement, where **1** = Zero, **2** = Scarcely, **3** = Medium level **4** = Frequently and **5** = Very frequently.

Zero (None), **Scarcely** (1 -10% of project duration e.g. 10% if it happened on 10 different days on a 100 days duration project), **Medium level** (30% of project duration), **Frequently** or **Very frequently**.

	What frequency would you give for the following statement?	1	2	3	4	5
1	How frequently did staff use Building information Model (BIM)?	☐	☐	☐	☐	☐
2	How frequently did quality control issues occur in the project?	☐	☐	☐	☐	☐
3	How frequently did conflict(s) occur between contractor and sub-contractor(s)?	☐	☐	☐	☐	☐
4	How frequently did rework(s) due to error in construction occur on site?	☐	☐	☐	☐	☐
5	How frequently did conflict(s) occur between consultant & contractor	☐	☐	☐	☐	☐
6	How frequently did equipment break down on site?	☐	☐	☐	☐	☐
7	How frequently did Inflation or sudden increase in commodities occur?	☐	☐	☐	☐	☐
8	How frequently did labour dispute or strikes occur?	☐	☐	☐	☐	☐
9	How frequently did unfavourable weather conditions occur?	☐	☐	☐	☐	☐
10	How frequently did natural disasters like floods, earthquakes etc. occur?	☐	☐	☐	☐	☐

11	How frequently did contractor experience financial difficulties?	☐	☐	☐	☐	☐
12	How frequently did variation in structural design variation occur?	☐	☐	☐	☐	☐
13	How frequently did late delivery of construction materials to the site occur?	☐	☐	☐	☐	☐
14	How frequently did change orders occur?	☐	☐	☐	☐	☐
15	How frequently did the project experience issues with discrepancies in contract documents?	☐	☐	☐	☐	☐
16	How frequently did significant fluctuation in material prices occur in the project life?	☐	☐	☐	☐	☐
17	How frequently did the contractor experience cash flow issues during the project life?	☐	☐	☐	☐	☐
18	How frequently did the client experience cash flow issues during the project life?	☐	☐	☐	☐	☐
19	How frequently did the consultant experience cash flow issues during the project life?	☐	☐	☐	☐	☐
20	How frequently did unforeseen site conditions (e.g., unanticipated groundwater, quicksand, mud, rock formations etc) occur?	☐	☐	☐	☐	☐
21	How frequently did the project experience political Influence?	☐	☐	☐	☐	☐
22	How frequently did you experience shortage of resources (human resources, machinery, equipment)?	☐	☐	☐	☐	☐
23	How frequently did the project experience space limitations at site for temporary equipment?	☐	☐	☐	☐	☐
24	How frequently did the project experience space limitations at site for permanent equipment?					
25	How frequently did the project experience changes in site topography after design?	☐	☐	☐	☐	☐
26	How frequently did the project experience poor site investigation or management?	☐	☐	☐	☐	☐
27	How frequently did you experience shortage of materials on market?	☐	☐	☐	☐	☐
28	How frequently did client requirements change?	☐	☐	☐	☐	☐
29	How frequently did corruption issues occur?	☐	☐	☐	☐	☐
30	How frequently did staff use outdated construction methods (e.g., precast twin wall technology, 2D drawing etc.)?	☐	☐	☐	☐	☐
31	How frequently did staff use new technologies?	☐	☐	☐	☐	☐
32	How frequently did accident occur on the site?	☐	☐	☐	☐	☐
33	How frequently did specifications change during construction change?	☐	☐	☐	☐	☐
34	How frequently did worker worked below average productivity level?	☐	☐	☐	☐	☐
35	How frequently was poor decision made on the project?	☐	☐	☐	☐	☐
36	How frequently did change in economic conditions affect the project?	☐	☐	☐	☐	☐

Section B:

For each of the questions below, circle the response that best characterises the percentage you will give each statement, where **1** = 0 - 20%, **2** = 21 - 40%, **3** = 41 - 60%, **4** = 61 - 80% and **5** = 81 - 100%.

	What percentage would you give for the following statement?	1	2	3	4	5
1	What proportion of the payment made by the owner was late?	☐	☐	☐	☐	☐
2	What proportion of workers were unskilled labourers?	☐	☐	☐	☐	☐
3	What percentage of ordered materials was delivered late by suppliers?	☐	☐	☐	☐	☐
4	What percentage of budgeting was inaccurate?	☐	☐	☐	☐	☐
5	What percentage of resource planning was inaccurate?	☐	☐	☐	☐	☐
6	What proportion of communication among stakeholders was poor?	☐	☐	☐	☐	☐
7	What proportion of project planning was ineffective?	☐	☐	☐	☐	☐
8	What percentage of contract management was poor?	☐	☐	☐	☐	☐
9	What percentage of the project was affected by ineffective government regulations?	☐	☐	☐	☐	☐
10	What proportion of materials in storage was damaged?	☐	☐	☐	☐	☐
11	What proportion of equipment used was brand new or near new?	☐	☐	☐	☐	☐
12	What percentage of the project was affected by ineffective project management?	☐	☐	☐	☐	☐
13	What percentage of the project was affected by ineffective project supervision?	☐	☐	☐	☐	☐
14	What percentage of project activity was affected by pandemic?	☐	☐	☐	☐	☐
15	What proportion of project activity was affected by traffic restrictions and planning issues outside of the project?	☐	☐	☐	☐	☐
16	What proportion of project activity was affected before obtaining building permit from authorities?	☐	☐	☐	☐	☐

Section C:

For each of the questions below, tick the response that best describes how detailed the schedule of the project was in the following statement, where **1** = **No Schedule**, **2** = **Basic Schedule**, **3** = **Good Schedule** **4** = **Detailed Schedule** and **5** = **Detailed and Frequently Updated Schedule**.

		1	2	3	4	5
1	To what level of detail was the schedule/program of work of this project?	☐	☐	☐	☐	☐

Section D:

For the question below, tick the response that best characterises how the following statement affected the project

Where:

1 = 0 – 20% extra of initial schedule duration, **2** = 21 – 40% extra of initial schedule duration,
3 = 41 – 60% extra of initial schedule duration, **4** = 61- 80% extra of initial schedule duration and
5 = Above 80% extra of initial schedule duration (i.e., twice initial schedule duration and more).

		1	2	3	4	5
1	How long was the entire project delayed for?	☐	☐	☐	☐	☐

References:

Abbasnejad, B. and Moud, H.I. (2013) ‘BIM and Basic Challenges Associated with its Definitions , Interpretations and Expectations’, *International Journal of Engineering Research and Application*, 3(2), pp. 287–294. Available at: www.ijera.com (Accessed: 28 April 2020).

Abdelbary, M., Edkins, A. and Dorra, E.M. (2020) ‘Reducing crr in fast-track projects through bim’, *Journal of Information Technology in Construction*, 25(December 2018), pp. 140–160. doi:10.36680/j.itcon.2020.009.

Abdul-Rahman, H. *et al.* (2011) ‘Project schedule influenced by financial issues: Evidence in construction industry’, *Scientific Research and Essays*, 6(1), pp. 205–212. doi:10.5897/SRE10.989.

Abdul-Rahman, H., Takim, R. and Min, W.S. (2009) ‘Financial-related causes contributing to project delays’, *Journal of Retail and Leisure Property*, 8(3), pp. 225–238. doi:10.1057/rlp.2009.11.

Abeynayake, H.I.M.M. *et al.* (2023) ‘Efficacy of information extraction from bar, line, circular, bubble and radar graphs’, *Applied Ergonomics*, 109, p. 103996. doi:10.1016/J.APERGO.2023.103996.

Abioye, S.O. *et al.* (2021) ‘Artificial intelligence in the construction industry: A review of present status, opportunities and future challenges’, *Journal of Building Engineering*, 44, p. 103299. doi:10.1016/J.JOBE.2021.103299.

Aghalari, A. *et al.* (2021) ‘Inverse reinforcement learning to assess safety of a workplace under an active shooter incident’, *IISE Transactions*, 53(12), pp. 1337–1350. doi:10.1080/24725854.2021.1922785.

Agyekum-Mensah, G. and Knight, A.D. (2017) ‘The professionals’ perspective on the causes of project delay in the construction industry’, *Engineering, Construction and Architectural Management*, 24(5), pp. 828–841. doi:10.1108/ECAM-03-2016-0085.

Ajayi, A. *et al.* (2020) ‘Deep Learning Models for Health and Safety Risk Prediction in Power Infrastructure Projects’, *Risk Analysis*, 40(10), pp. 2019–2039. doi:10.1111/RISA.13425.

Akhund, M.A. *et al.* (2017) ‘Time overrun in construction projects of developing countries’, *Imperial journal of interdisciplinary research (IJIR)*, 3(5), pp. 124–129.

Al-Hazim, N., Salem, Z.A. and Ahmad, H. (2017) ‘Delay and Cost Overrun in Infrastructure Projects in Jordan’, *Procedia Engineering*, 182, pp. 18–24. doi:10.1016/j.proeng.2017.03.105.

Al-Mohammad, M.S. *et al.* (2021) ‘Factors affecting BIM implementation in post-conflict low-income economies: the case of Afghanistan’, *Journal of Engineering, Design and Technology* [Preprint]. doi:10.1108/JEDT-04-2021-0205.

Alaghbari, W. and Sultan, B. (2018) ‘Delay Factors Impacting Construction Projects in Sana ’ a - Yemen 1’, *PM World Journal*, VII(December), pp. 1–28. Available at: https://www.researchgate.net/publication/329656460%0ADelay.

Alnaser, A.A., Alsanabani, N.M. and Al-Gahtani, K.S. (2023) ‘BIM Impact on Construction Project

Time Using System Dynamics in Saudi Arabia's Construction', *Buildings*, 13(9), p. 2267. doi:10.3390/buildings13092267.

Amany, A., Taghizade, K. and Noorzai, E. (2020) 'Investigating conflicts of expert contractors using the last planner system in building information modeling process', *Journal of Engineering, Design and Technology*, 18(6), pp. 1381–1402. doi:10.1108/JEDT-09-2019-0223.

Ameziane, F. (2000) 'Information system for building production management', *International Journal of Production Economics*, 64(1), pp. 345–358. doi:10.1016/S0925-5273(99)00071-7.

Amin, M.N. *et al.* (2021) 'Comparison of Machine Learning Approaches with Traditional Methods for Predicting the Compressive Strength of Rice Husk Ash Concrete', *Crystals 2021, Vol. 11, Page 779*, 11(7), p. 779. doi:10.3390/CRYST11070779.

Amini Toosi, H. *et al.* (2022) 'A novel LCSA-Machine learning based optimization model for sustainable building design-A case study of energy storage systems', *Building and Environment*, 209, p. 108656. doi:10.1016/J.BUILDENV.2021.108656.

Amiri, M., Ardeshir, A. and Fazel Zarandi, M.H. (2017) 'Fuzzy probabilistic expert system for occupational hazard assessment in construction', *Safety Science*, 93, pp. 16–28. doi:10.1016/J.SSCI.2016.11.008.

Andrew, R. (2013) *UK Construction: An economic analysis of the sector*, *Department for Business Information & Skills*. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/210060/bis-13-958-uk-construction-an-economic-analysis-of-sector.pdf (Accessed: 2 May 2020).

Ansah, R.H. *et al.* (2016) 'Advancing towards delay-free construction project: A review', *Proceedings of the International Conference on Industrial Engineering and Operations Management*, pp. 744–751.

Arditi, D. and Pulket, T. (2005) 'Predicting the outcome of construction litigation using boosted decision trees', *Journal of Computing in Civil Engineering*, 19(4), pp. 387–393. doi:10.1061/(ASCE)0887-3801(2005)19:4(387).

Armeshi, H., Sahebi, M.R. and Aghababaei, H. (2024) 'A deterministic descriptive regularization-based method for SAR tomography in urban areas', *International Journal of Remote Sensing*, 45(6), pp. 1884–1903. doi:10.1080/01431161.2024.2321466.

Asadi, A., Alsubaey, M. and Makatsoris, C. (2015) 'A machine learning approach for predicting delays in construction logistics', *International Journal of Advanced Logistics*, 4(2), pp. 115–130. doi:10.1080/2287108x.2015.1059920.

ASCE (2010) 'About asce', *ASCE News Online*, p. 1. Available at: https://www.asce.org/about_asce/ (Accessed: 22 April 2021).

Assaf, S.A. and Al-Hejji, S. (2006) 'Causes of delay in large construction projects', *International Journal of Project Management*, 24(4), pp. 349–357. doi:10.1016/j.ijproman.2005.11.010.

Assaf, S.A., Al-Khalil, M. and Al-Hazmi, M. (1995) 'Causes of delay in large building construction projects', *Journal of Management in Engineering*, 11(2), pp. 45–50. doi:10.1061/(ASCE)0742-597X(1995)11:2(45).

Austin, P.C. (2008) 'The large-sample performance of backwards variable elimination', *Journal of Applied Statistics*, 35(12), pp. 1355–1370. doi:10.1080/02664760802382434.

Ayadi, M.I. *et al.* (2019) 'Deep learning in building management systems over ndn: Use case of forwarding and hvac control', *Proceedings - 2019 IEEE International Congress on Cybermatics: 12th IEEE International Conference on Internet of Things, 15th IEEE International Conference on Green Computing and Communications, 12th IEEE International Conference on Cyber, Physical and So*, pp. 1192–1198. doi:10.1109/iThings/GreenCom/CPSCom/SmartData.2019.00200.

Ayhan, B.U. and Tokdemir, O.B. (2019) 'Safety assessment in megaprojects using artificial intelligence', *Safety Science*, 118, pp. 273–287. doi:10.1016/J.SSCI.2019.05.027.

Ayhan, M., Dikmen, I. and Birgonul, M.T. (2021) 'Predicting the Occurrence of Construction Disputes Using Machine Learning Techniques', *Journal of Construction Engineering and Management*, 147(4), p. 04021022. doi:10.1061/(ASCE)CO.1943-7862.0002027.

Azhar, S. (2011) 'Building information modeling (BIM): Trends, benefits, risks, and challenges for the AEC industry', *Leadership and Management in Engineering*, 11(3), pp. 241–252. doi:10.1061/(ASCE)LM.1943-5630.0000127.

Azhar, S., Khalfan, M. and Maqsood, T. (2012) 'Building information modeling (BIM): Now and beyond', *Australasian Journal of Construction Economics and Building*, 12(4), pp. 15–28. doi:10.5130/ajceb.v12i4.3032.

Aziz, R.F. and Hafez, S.M. (2013) 'Applying lean thinking in construction and performance improvement', *Alexandria Engineering Journal*, 52(4), pp. 679–695. doi:10.1016/J.AEJ.2013.04.008.

Bagheri, A., Nazari, A. and Sanjayan, J. (2019) 'The use of machine learning in boron-based geopolymers: Function approximation of compressive strength by ANN and GP', *Measurement*, 141, pp. 241–249. doi:10.1016/J.MEASUREMENT.2019.03.001.

Bajaj, R. and Sharma, V. (2018) 'Smart Education with artificial intelligence based determination of learning styles', *Procedia Computer Science*, 132, pp. 834–842. doi:10.1016/J.PROCS.2018.05.095.

Bajjou, M.S. and Chafi, A. (2020) 'Identifying and Managing Critical Waste Factors for Lean Construction Projects', *Engineering Management Journal*, 32(1), pp. 2–13. doi:10.1080/10429247.2019.1656479.

Bajpai, A. and Misra, S.C. (2020) 'Identifying Critical Risk Factors for Use of Digitalization in Construction Industry: A Case Study', *Proceedings - 2020 IEEE India Council International Subsections Conference, INDISCON 2020*, pp. 124–128. doi:10.1109/INDISCON50162.2020.00036.

Balaguer, C. *et al.* (2002) 'FutureHome: An integrated construction automation approach', *IEEE Robotics and Automation Magazine*, 9(1), pp. 55–66. doi:10.1109/100.993155.

Baldwin, J.R., Manthei, J.M., Rothbart, H. and Harris, R.B. (1971) 'Causes of delay in the construction industry', *Journal of the Construction Division*, pp. 177–187. Available at: https://cedb.asce.org/CEDBsearch/record.jsp?dockey=0018302 (Accessed: 23 April 2020).

Bao, Y., Hilary, G. and Ke, B. (2022) 'Artificial Intelligence and Fraud Detection', in. Springer, Cham, pp. 223–247. doi:10.1007/978-3-030-75729-8_8.

Barai, S. V and Nair, R.S. (2004) 'NEURO-FUZZY MODELS FOR CONSTRUCTABILITY ANALYSIS', 9. Available at: http://www.itcon.org/2004/4/ (Accessed: 22 March 2022).

Bassier, M. and Vergauwen, M. (2020) 'Unsupervised reconstruction of Building Information Modeling wall objects from point cloud data', *Automation in Construction*, 120, p. 103338. doi:10.1016/J.AUTCON.2020.103338.

Bayer, J. and Melone, N. (1989) 'A critique of diffusion theory as a managerial framework for understanding adoption of software engineering innovations', *Journal of Systems and Software*, 9(2), pp. 161–166. doi:10.1016/0164-1212(89)90018-6.

Bell, E., Bryman, A. and Harley, B. (2018) 'Business research methods'.

Bernstein, H. (2010) 'McGraw-Hill Construction: The Business Value of BIM in Europe, Smart Market Report'.

Bertschinger, D. *et al.* (2023) 'Training Fully Connected Neural Networks is $\exists\mathbb{R}$-Complete', *Advances in Neural Information Processing Systems*, 36, pp. 36222–36237.

Blanco, J.L. *et al.* (2018) 'Artificial intelligence : Construction technology ' s next frontier', *Mckinsey & Company*, (April), pp. 1–8. Available at: https://www.mckinsey.com/industries/capital-projects-and-infrastructure/our-insights/artificial-intelligence-construction-technologys-next-frontier.

Bless, C., Higson-Smith, C. and Sithole, S.L. (2013) *Fundamentals of social research: An African perspective*, *Kenwyn: Juta*. Juta and Co.

Bolón-Canedo, V., Sánchez-Maroño, N. and Alonso-Betanzos, A. (2015) 'Recent advances and emerging challenges of feature selection in the context of big data', *Knowledge-Based Systems*, 86, pp. 33–45. doi:10.1016/J.KNOSYS.2015.05.014.

Bramble, B. and Callahan, M. (2004) 'Construction delay claims'.

Bruckmann, T., Reichert, C., *et al.* (2018) 'Concept Studies of Automated Construction Using Cable-Driven Parallel Robots', *Mechanisms and Machine Science*, 53, pp. 364–375. doi:10.1007/978-3-319-61431-1_31.

Bruckmann, T., Spengler, A.J., *et al.* (2018) 'Process Analysis of Cable-Driven Parallel Robots for Automated Construction', *Intelligent Systems, Control and Automation: Science and Engineering*, 92, pp. 63–83. doi:10.1007/978-3-319-68646-2_3.

Bryman, A. (2012) 'Social Research Methods', *Oxford: Oxford University Press*, pp. 3–25. doi:10.4135/9781849209939.

Btoush, M. and Harun, A.T. (2017) 'Minimizing delays in the Jordanian construction industry by adopting BIM technology', *IOP Conference Series: Materials Science and Engineering*, 271(1). doi:10.1088/1757-899X/271/1/012041.

Caffieri, J.J. *et al.* (2018) 'Planning for production in construction: controlling costs in major capital projects', *Production Planning and Control*, 29(1), pp. 41–50. doi:10.1080/09537287.2017.1376258.

Cantú-Ortiz, F.J. and Fangmeyer, J. (2017) 'University performance in the age of research analytics', in *Research Analytics: Boosting University Productivity and Competitiveness through Scientometrics*. CRC Press, pp. 237–258. doi:10.1201/9781315155890.

De Caro, F. *et al.* (2023) 'Review of Data-Driven Techniques for On-Line Static and Dynamic Security Assessment of Modern Power Systems', *IEEE Access*, 11, pp. 130644–130673. doi:10.1109/ACCESS.2023.3334394.

Cha, G.W. *et al.* (2020a) 'Development of a prediction model for demolition waste generation using a random forest algorithm based on small datasets', *International Journal of Environmental Research and Public Health*, 17(19), pp. 1–15. doi:10.3390/IJERPH17196997.

Cha, G.W. *et al.* (2020b) 'Development of a Prediction Model for Demolition Waste Generation Using a Random Forest Algorithm Based on Small DataSets', *International Journal of Environmental Research and Public Health 2020, Vol. 17, Page 6997*, 17(19), p. 6997. doi:10.3390/IJERPH17196997.

Cha, G.W., Moon, H.J. and Kim, Y.C. (2021a) 'Comparison of random forest and gradient boosting machine models for predicting demolition waste based on small datasets and categorical variables', *International Journal of Environmental Research and Public Health*, 18(16). doi:10.3390/IJERPH18168530.

Cha, G.W., Moon, H.J. and Kim, Y.C. (2021b) 'Comparison of Random Forest and Gradient Boosting Machine Models for Predicting Demolition Waste Based on Small Datasets and Categorical Variables', *International Journal of Environmental Research and Public Health 2021, Vol. 18, Page 8530*, 18(16), p. 8530. doi:10.3390/IJERPH18168530.

Charizanos, G., Demirhan, H. and İçen, D. (2024) 'A Monte Carlo fuzzy logistic regression framework against imbalance and separation', *Information Sciences*, 655, p. 119893. doi:10.1016/j.ins.2023.119893.

Charoenkwan, P. and Homkong, N. (2017) 'CSDeep: A crushed stone image predictor based on deep learning and intelligently selected features', *Proceeding of 2017 2nd International Conference on Information Technology, INCIT 2017*, 2018-January, pp. 1–6. doi:10.1109/INCIT.2017.8257857.

Chen, C. *et al.* (2022) 'Comprehensive performance evaluation of intrusion detection model based on radar chart method', *Proceedings - 2022 2nd International Conference on Electronic Information Engineering and Computer Technology, EIECT 2022*, pp. 44–47. doi:10.1109/EIECT58010.2022.00014.

Chen, C. and Tang, L. (2019) 'BIM-based integrated management workflow design for schedule and cost planning of building fabric maintenance', *Automation in Construction*, 107(August), p. 102944. doi:10.1016/j.autcon.2019.102944.

Chen, G.-X. *et al.* (2019) 'Investigating the causes of delay in grain bin construction projects: the case of China', *International Journal of Construction Management*, 19(1), pp. 1–14. doi:10.1080/15623599.2017.1354514.

Chen, G.X. *et al.* (2019) 'Investigating the causes of delay in grain bin construction projects: the case of China', *International Journal of Construction Management*, 19(1), pp. 1–14. doi:10.1080/15623599.2017.1354514.

Chen, J. *et al.* (2021) 'Artificial intelligence-based human-centric decision support framework: an application to predictive maintenance in asset management under pandemic environments', *Annals of Operations Research*, pp. 1–24. doi:10.1007/S10479-021-04373-W/TABLES/11.

Chen, N.-C. *et al.* (2018) *Using Machine Learning to Support Qualitative Coding in Social Science: Shifting The Focus to Ambiguity*. Available at: https://doi.org/0000001.0000001 (Accessed: 16 June 2020).

Chiponde, D.B. *et al.* (2017) 'Assessing the feasibility of using building information modelling (BIM) to improve collaboration on public sector projects in the Zambian construction industry', *WIT Transactions on the Built Environment*, 169, pp. 191–199. doi:10.2495/BIM170181.

Chou, H.Y. and Chen, P.Y. (2017) 'Benefit Evaluation of Implementing BIM in Construction Projects', *IOP Conference Series: Materials Science and Engineering*, 245(6). doi:10.1088/1757-899X/245/6/062049.

Chu, B. *et al.* (2013) 'Robot-based construction automation: An application to steel beam assembly (Part I)', *Automation in Construction*, 32, pp. 46–61. doi:10.1016/J.AUTCON.2012.12.016.

CIOB (2011) 'Guide to Good Practice in the Management of Time in Complex Projects', *Structural Survey*, 29(3), pp. 73–76. doi:10.1108/ss.2011.11029caa.017.

Cohen, L., Manion, L. and Morrison, K. (2002) 'Research Methods in Education', *Research Methods in Education* [Preprint]. doi:10.4324/9780203224342.

Collis, J. and Hussey, R. (2009) 'Business research : a practical guide for undergraduate & postgraduate students', p. 358.

Commission, E. (2015) 'User guide to the SME Definition', *Luxembourg: Publications Office of the European Union*, pp. 10–60. doi:10.2873/782201.

Conforto, E.C. *et al.* (2014) 'Can Agile Project Management Be Adopted by Industries Other than Software Development?', *Project Management Journal*, 45(3), pp. 21–34. doi:10.1002/PMJ.21410.

Cooney, J.P., Oloke, D. and Gyoh, L. (2021) 'A novel heritage BIM (HBIM) framework development for heritage buildings refurbishment based on an investigative study of microorganisms', *Journal of Engineering, Design and Technology* [Preprint]. doi:10.1108/JEDT-07-2021-0370/FULL/PDF.

Couto, R. and Ericson, D. (2017) ‘A Case Study of Logan International Airport’s Terminal C to e Connector: Structural Challenges and Risks for Existing Building Projects’, in *Structures Congress 2017: Business, Professional Practice, Education, Research, and Disaster Management - Selected Papers from the Structures Congress 2017*. American Society of Civil Engineers (ASCE), pp. 1–13. doi:10.1061/9780784480427.001.

Crotty, M. (1998) ‘The Foundations of Social Research BT - Introduction: The Research Process’, *Introduction: The Research Process*, pp. 1–17. Available at: http://books.google.com/books?id=Op_SDKrf1ZQC&printsec=frontcover&dq=inauthor:crotty+(1998+book)&hl=&cd=1&source=gbs_api (Accessed: 10 October 2022).

Dai, Y., Wu, Q. and Zhang, Y. (2024) ‘Generalized sparse radial basis function networks for multi-classification problems’, *Applied Soft Computing*, 154, p. 111361. doi:10.1016/J.ASOC.2024.111361.

Dallasega, P. *et al.* (2019) ‘BIM-based construction progress measurement of non-repetitive HVAC installation works’, *27th Annual Conference of the International Group for Lean Construction, IGLC 2019*, pp. 819–830. doi:10.24928/2019/0152.

Davila Delgado, J.M. *et al.* (2019) ‘Robotics and automated systems in construction: Understanding industry-specific challenges for adoption’, *Journal of Building Engineering*, 26. doi:10.1016/J.JOBE.2019.100868.

Davila Delgado, J.M. and Oyedele, L. (2021) ‘Deep learning with small datasets: using autoencoders to address limited datasets in construction management’, *Applied Soft Computing*, 112, p. 107836. doi:10.1016/J.ASOC.2021.107836.

Davis, D. (2007) ‘National Building Information Model Standard NBIMS • BIM GIS Integration • Why NBIMS • What is NBIMS • Issues of Interoperability Chair-National BIM Standards Scoping Chair’, *Infosystems* [Preprint].

Debrah, C., Chan, A.P.C. and Darko, A. (2022) ‘Artificial intelligence in green building’, *Automation in Construction*, 137, p. 104192. doi:10.1016/J.AUTCON.2022.104192.

Deeba, F. *et al.* (2018) ‘Performance assessment of a bleeding detection algorithm for endoscopic video based on classifier fusion method and exhaustive feature selection’, *Biomedical Signal Processing and Control*, 40, pp. 415–424. doi:10.1016/J.BSPC.2017.10.011.

Delgado, F. *et al.* (2015) ‘Towards a client-oriented integration of construction processes and building GIS systems’, *Computers in Industry*, 73, pp. 51–68. doi:10.1016/j.compind.2015.07.012.

Demirkesen, S. and Ozorhon, B. (2017) ‘Impact of integration management on construction project management performance’, *International Journal of Project Management*, 35(8), pp. 1639–1654. doi:10.1016/J.IJPROMAN.2017.09.008.

Derakhshanfar, H. *et al.* (2019) ‘Construction delay risk taxonomy, associations and regional contexts A systematic review and meta-analysis’, 26(10), pp. 2364–2388. doi:10.1108/ECAM-07-2018-0307.

Dey, S.K. *et al.* (2022) ‘Chi2-MI: A hybrid feature selection based machine learning approach in diagnosis of chronic kidney disease', *Intelligent Systems with Applications*, 16, p. 200144. doi:10.1016/J.ISWA.2022.200144.

Diaz, B. *et al.* (2019) 'Time and cost optimization for road projects through BIM', in *2019 Congreso Internacional de Innovacion y Tendencias en Ingenieria, CONIITI 2019 - Conference Proceedings*. Institute of Electrical and Electronics Engineers Inc. doi:10.1109/CONIITI48476.2019.8960767.

Duan, K.K. and Cao, S.Y. (2020) 'Emerging RFID technology in structural engineering – A review', *Structures*, 28, pp. 2404–2414. doi:10.1016/J.ISTRUC.2020.10.036.

Egan, S.J. (2018) 'Rethinking the Report of the Construction Task Force', *Construction Task Force* [Preprint]. Available at: http://constructingexcellence.org.uk/wp-content/uploads/2014/10/rethinking_construction_report.pdf.

Egwim, C.N., Egunjobi, O.O., *et al.* (2021) 'A Comparative Study on Machine Learning Algorithms for Assessing Energy Efficiency of Buildings', in *Communications in Computer and Information Science*. Springer, Cham, pp. 546–566. doi:10.1007/978-3-030-93733-1_41.

Egwim, C.N., Alaka, H., Toriola-Coker, L.O., Balogun, H. and Sunmola, F. (2021a) 'Applied artificial intelligence for predicting construction projects delay', *Machine Learning with Applications*, 6, p. 100166. doi:10.1016/j.mlwa.2021.100166.

Egwim, C.N., Alaka, H., Toriola-Coker, L.O., Balogun, H. and Sunmola, F. (2021b) 'Applied artificial intelligence for predicting construction projects delay', *Machine Learning with Applications*, 6, p. 100166. doi:10.1016/j.mlwa.2021.100166.

Egwim, C.N., Alaka, H., Toriola-Coker, L.O., Balogun, H., Ajayi, S., *et al.* (2021) 'Extraction of underlying factors causing construction projects delay in Nigeria', *Journal of Engineering, Design and Technology*, ahead-of-p(ahead-of-print). doi:10.1108/jedt-04-2021-0211.

Egwim, C.N. *et al.* (2022a) 'Comparison of machine learning algorithms for evaluating building energy efficiency using big data analytics', *Journal of Engineering, Design and Technology*, ahead-of-p(ahead-of-print). doi:10.1108/JEDT-05-2022-0238.

Egwim, C.N. *et al.* (2022b) 'Comparison of machine learning algorithms for evaluating building energy efficiency using big data analytics', *Journal of Engineering, Design and Technology*, ahead-of-p(ahead-of-print). doi:10.1108/JEDT-05-2022-0238.

Egwim, C.N., Alaka, H., Pan, Y., *et al.* (2023a) 'Ensemble of ensembles for fine particulate matter pollution prediction using big data analytics and IoT emission sensors', *Journal of Engineering, Design and Technology*, ahead-of-p(ahead-of-print). doi:10.1108/JEDT-07-2022-0379.

Egwim, C.N., Alaka, H., Pan, Y., *et al.* (2023b) 'Ensemble of ensembles for fine particulate matter pollution prediction using big data analytics and IoT emission sensors', *Journal of Engineering, Design and Technology* [Preprint]. doi:10.1108/JEDT-07-2022-0379.

Egwim, C.N., Alaka, H., Demir, E., *et al.* (2023) 'Systematic review of critical drivers for delay risk prediction: towards a conceptual framework for BIM-based construction projects', *Frontiers in*

Engineering and Built Environment, 3(1), pp. 16–31. doi:10.1108/febe-05-2022-0017.

Egwim, C.N. *et al.* (2024) ‘Artificial Intelligence in the Construction Industry: A Systematic Review of the Entire Construction Value Chain Lifecycle’, *Energies*. Multidisciplinary Digital Publishing Institute, p. 182. doi:10.3390/en17010182.

Egwim, C.N. and Alaka, H. (2021) ‘A Comparative Study on Machine Learning Algorithms for Predicting Construction Projects Delay’, in *Environmental Design and Management International Conference, Bristol, United Kingdom.*

Elawi, G.S.A., Algahtany, M. and Kashiwagi, D. (2016) ‘Owners’ Perspective of Factors Contributing to Project Delay: Case Studies of Road and Bridge Projects in Saudi Arabia’, *Procedia Engineering*, 145(480), pp. 1402–1409. doi:10.1016/j.proeng.2016.04.176.

Elghaish, F. *et al.* (2020) ‘Revolutionising cost structure for integrated project delivery: a BIM-based solution’, *Engineering, Construction and Architectural Management*, 28(4), pp. 1214–1240. doi:10.1108/ECAM-04-2019-0222.

Elhusseiny, H.O., Nosair, I. and Ezeldin, A.S. (2021) ‘Systematic processing framework for analyzing the factors of construction projects’ delays in Egypt’, *Ain Shams Engineering Journal* [Preprint], (xxxx). doi:10.1016/j.asej.2020.10.016.

Elsheikh, A. *et al.* (2021) ‘Research on Energy-efficiency Building Design Based on BIM and Artificial Intelligence’, *IOP Conference Series: Earth and Environmental Science*, 825(1), p. 012003. doi:10.1088/1755-1315/825/1/012003.

Enshassi, A., Al□Najjar, J. and Kumaraswamy, M. (2009) ‘Delays and cost overruns in the construction projects in the Gaza Strip’, *Journal of Financial Management of Property and Construction*, 14(2), pp. 126–151. doi:10.1108/13664380910977592.

Ern, P.A.S., Ooi, Y.Y. and Al-Ashmori, Y.Y. (2020) ‘Comparative study on the perspective towards the benefits and hindrances of implementing building information modelling (Bim)’, *International Journal of Sustainable Construction Engineering and Technology*, 11(1), pp. 194–205. doi:10.30880/ijscet.2020.11.01.019.

Eskandari, H. *et al.* (2024) ‘Innovative framework for accurate and transparent forecasting of energy consumption: A fusion of feature selection and interpretable machine learning’, *Applied Energy*, 366, p. 123314. doi:10.1016/J.APENERGY.2024.123314.

European Comission (2017) ‘European Construction Sector Observatory’, (June), p. 27.

Evans, M. *et al.* (2021) ‘Critical success factors for adopting building information modelling (BIM) and lean construction practices on construction mega-projects: a Delphi survey’, *Journal of Engineering, Design and Technology*, 19(2), pp. 537–556. doi:10.1108/JEDT-04-2020-0146/FULL/PDF.

Falagas, M.E. *et al.* (2008) ‘Comparison of PubMed, Scopus, Web of Science, and Google Scholar: strengths and weaknesses’, *The FASEB Journal*, 22(2), pp. 338–342. doi:10.1096/FJ.07-9492LSF.

Fan, Y. *et al.* (2024) ‘Learning correlation information for multi-label feature selection’, *Pattern*

Recognition, 145, p. 109899. doi:10.1016/J.PATCOG.2023.109899.

Faraji, A., Rashidi, M. and Perera, S. (2021) 'Text Mining Risk Assessment–Based Model to Conduct Uncertainty Analysis of the General Conditions of Contract in Housing Construction Projects: Case Study of the NSW GC21', *Journal of Architectural Engineering*, 27(3), p. 04021025. doi:10.1061/(ASCE)AE.1943-5568.0000489.

Filipe Barbosa, Jonathan Woetzel, Jan Mischke, Maria João Ribeirinho, Mukund Sridhar, Matthew Parsons, Nick Bertram, and S.B. (2017) 'Re-inventing Construction In Brief'.

Firth, C. *et al.* (2020) 'Development of an Anthropomorphic End-effector for Collaborative use on Construction Sites', in *RE: Anthropocene, Design in the Age of Humans - Proceedings of the 25th International Conference on Computer-Aided Architectural Design Research in Asia, CAADRIA 2020*, pp. 363–372.

Fisher-Gewirtzman, D. and Polak, N. (2019) 'A learning automated 3D architecture synthesis model: demonstrating a computer governed design of minimal apartment units based on human perceptual and physical needs', *Architectural Science Review*, 62(4), pp. 301–312. doi:10.1080/00038628.2019.1611537.

Flick, U. (2011) 'Gathering Data: Quantitaive and Qualitative Approaches', *Introducing Research Methodology: A Beginner's Guide to Doing a Research Project*, pp. 106–111. Available at: https://books.google.co.uk/books?id=-EeXiLAi4TgC (Accessed: 20 October 2022).

Flyvbjerg, B. (2004) 'Procedures for Dealing with Optimism Bias in Transport Planning', *The British Department for Transport*, (June), p. 61.

Flyvbjerg, B. (2008) 'Curbing optimism bias and strategic misrepresentation in planning: Reference class forecasting in practice', *European Planning Studies*, 16(1), pp. 3–21. doi:10.1080/09654310701747936.

Flyvbjerg, B. (2014) 'What you should know about megaprojects and why: An overview', *Project Management Journal*, 45(2), pp. 6–19. doi:10.1002/pmj.21409.

Foithong, S., Pinngern, O. and Attachoo, B. (2012) 'Feature subset selection wrapper based on mutual information and rough sets', *Expert Systems with Applications*, 39(1), pp. 574–584. doi:10.1016/J.ESWA.2011.07.048.

Frimpong, Y., Oluwoye, J. and Crawford, L. (2003) 'Causes of delay and cost overruns in construction of groundwater projects in a developing countries; Ghana as a case study', *International Journal of Project Management*, 21(5), pp. 321–326. doi:10.1016/S0263-7863(02)00055-8.

Fuller, S. *et al.* (2001) 'Positivism , History of Positivism : Sociological Positivism', *International Encyclopedia of the Social & Behavioral Sciences* [Preprint].

Gambao, E., Balaguer, C. and Gebhart, F. (2000) 'Robot assembly system for computer-integrated construction', *Automation in Construction*, 9(5–6), pp. 479–487. doi:10.1016/S0926-5805(00)00059-5.

García de Soto, B. *et al.* (2018) 'Productivity of digital fabrication in construction: Cost and time analysis of a robotically built wall', *Automation in Construction*, 92, pp. 297–311. doi:10.1016/j.autcon.2018.04.004.

Gardezi, S.S.S., Manarvi, I.A. and Gardezi, S.J.S. (2014) 'Time extension factors in construction industry of Pakistan', *Procedia Engineering*, 77, pp. 196–204. doi:10.1016/j.proeng.2014.07.022.

Gharbia, M. *et al.* (2020) 'Robotic technologies for on-site building construction: A systematic review', *Journal of Building Engineering*, 32, p. 101584. doi:10.1016/J.JOBE.2020.101584.

Gharouni Jafari, K., Ghazi Sharyatpanahi, N.S. and Noorzai, E. (2021) 'BIM-based integrated solution for analysis and management of mismatches during construction', *Journal of Engineering, Design and Technology*, 19(1), pp. 81–102. doi:10.1108/JEDT-02-2020-0044/FULL/PDF.

Gibbs, D.-J. *et al.* (2013) 'An Investigation into whether Building Information Modelling (BIM) can Assist with Construction Delay Claims', *International Journal of 3-D Information Modeling*, 2(1), pp. 45–52. doi:10.4018/ij3dim.2013010105.

Giel, B.K. and Issa, R.R.A. (2013) 'Return on Investment Analysis of Using Building Information Modeling in Construction', *Journal of Computing in Civil Engineering*, 27(5), pp. 511–521. doi:10.1061/(asce)cp.1943-5487.0000164.

Given, L. (2012) 'Postpositivism', in *The SAGE Encyclopedia of Qualitative Research Methods*. doi:10.4135/9781412963909.n332.

Głuszak, M. and Lešniak, A. (2015) 'Construction Delays in Clients Opinion - Multivariate Statistical Analysis', in *Procedia Engineering*, pp. 182–189. doi:10.1016/j.proeng.2015.10.075.

Goddard, W. and Melville, S. (2005) 'Research Methodology: An Introduction', pp. 1–122.

Gondia, A. *et al.* (2019) 'Machine Learning Algorithms for Construction Projects Delay Risk Prediction', *Journal of Construction Engineering and Management*, 146(1), p. 04019085. doi:10.1061/(ASCE)CO.1943-7862.0001736.

Gondia, A. *et al.* (2020) 'Machine Learning Algorithms for Construction Projects Delay Risk Prediction', *Journal of Construction Engineering and Management*, 146(1), p. 04019085. doi:10.1061/(ASCE)CO.1943-7862.0001736.

Gong, H. *et al.* (2024) 'A new filter feature selection algorithm for classification task by ensembling pearson correlation coefficient and mutual information', *Engineering Applications of Artificial Intelligence*, 131, p. 107865. doi:10.1016/J.ENGAPPAI.2024.107865.

Goyal, M. (2019) 'ARTIFICIAL INTELLIGENCE : A TOOL FOR HYPER PERSONALIZATION', *International Journal of 360 Management Review*, 07, pp. 2320–7132.

Group, S. (2020) *North Bridge Refurbishment*. Available at: https://www.scapegroup.co.uk/case-studies/north-bridge-refurbishment (Accessed: 12 May 2020).

Grünbaum, N.N. (2007) 'Identification of ambiguity in the case study research typology: What is a unit of analysis?', *Qualitative Market Research*, 10(1), pp. 78–97. doi:10.1108/13522750710720413.

Hadipriono, F.C. (1992) ‘Expert System for Construction Safety. I: FaultTree Models’, *Journal of Performance of Constructed Facilities*, 6(4), pp. 246–260. doi:10.1061/(ASCE)0887-3828(1992)6:4(246).

Handayani, T.N., Likhitruangsilp, V. and Yabuki, N. (2019) ‘A building information modeling (BIM)-integrated system for evaluating the impact of change orders’, *Engineering Journal*, 23(4), pp. 67–90. doi:10.4186/ej.2019.23.4.67.

Haq, S. *et al.* (2017) ‘Effects of Delay in construction Projects of Punjab-Pakistan : An Empirical Study Effects of Delay in construction Projects of Punjab-Pakistan ’:, *Journal of Basic and Applied Scientific Research.*, 3(January 2014), pp. 87–96.

Harmon, J. *et al.* (2021) ‘Use of artificial intelligence and virtual reality within clinical simulation for nursing pain education: A scoping review’, *Nurse Education Today*, 97, p. 104700. doi:10.1016/J.NEDT.2020.104700.

Hartmann, T. *et al.* (2012) ‘Aligning building information model tools and construction management methods’, in *Automation in Construction*. Elsevier, pp. 605–613. doi:10.1016/j.autcon.2011.12.011.

Haseeb, M., Bibi, A. and Rabbani, W. (2011) ‘PROBLEMS OF PROJECTS AND EFFECTS OF DELAYS IN THE CONSTRUCTION INDUSTRY OF PAKISTAN’, *Australian Journal of Business and Management Research*, 1(5).

Haylen, A. (2019) ‘Crossrail (Elizabeth Line)’, *House of Commons Library* [Preprint], (January).

Hong, Y. *et al.* (2021) ‘Comparing Natural Language Processing Methods to Cluster Construction Schedules’, *Journal of Construction Engineering and Management*, 147(10), p. 04021136. doi:10.1061/(ASCE)CO.1943-7862.0002165.

Hossain, M.A. *et al.* (2018) ‘Design-for-Safety knowledge library for BIM-integrated safety risk reviews’, *Automation in Construction*, 94(June), pp. 290–302. doi:10.1016/j.autcon.2018.07.010.

Hosseini, M.R. *et al.* (2016) ‘BIM adoption within Australian Small and Medium-sized Enterprises (SMEs): an innovation diffusion model’, *Construction Economics and Building*, 16(3), pp. 71–86. doi:10.5130/AJCEB.v16i3.5159.

House, S.O. *et al.* (2007) *Adopting BIM for facilities management: Solutions for managing the Sydney Opera House*, *CRC for construction Innovation participants*. CRC for Construction Innovation. doi:10.1061/(ASCE)CO.1943-7862.0000433.

Hu, D. *et al.* (2020) ‘Segmenting areas of potential contamination for adaptive robotic disinfection in built environments’, *Building and Environment*, 184, p. 107226. doi:10.1016/J.BUILDENV.2020.107226.

Huang, S. *et al.* (2021) ‘Artificial intelligence in the diagnosis of covid-19: Challenges and perspectives’, *International Journal of Biological Sciences*, 17(6), pp. 1581–1587. doi:10.7150/IJBS.58855.

Husin, A.E. (2019) ‘Time performance upgrade by critical chain project management and BIM 4D integration on top structural work of a high rise building construction project', *ARPN Journal of Engineering and Applied Sciences*, 14(17), pp. 3063–3072. doi:10.13140/RG.2.2.26024.11523.

Ibrahim, H.S., Hashim, N. and Ahmad Jamal, K.A. (2019) 'The Potential Benefits of Building Information Modelling (BIM) in Construction Industry', *IOP Conference Series: Earth and Environmental Science*, 385(1). doi:10.1088/1755-1315/385/1/012047.

Ifinedo, P. (2011) 'An empirical analysis of factors influencing internet/e-business technologies adoption by smes in Canada', *International Journal of Information Technology and Decision Making*, 10(4), pp. 731–766. doi:10.1142/S0219622011004543.

Ilozor, B.D. and Kelly, D.J. (2012) 'Building Information Modeling and Integrated Project Delivery in the Commercial Construction Industry: A Conceptual Study', *Journal of Engineering, Project, and Production Management*, 2(1), pp. 23–36. doi:10.32738/jeppm.201201.0004.

Imani, M.B., Keyvanpour, M.R. and Azmi, R. (2013) 'A NOVEL EMBEDDED FEATURE SELECTION METHOD: A COMPARATIVE STUDY IN THE APPLICATION OF TEXT CATEGORIZATION', *Applied Artificial Intelligence*, 27(5), pp. 408–427. doi:10.1080/08839514.2013.774211.

Imriyas, K. (2009) 'An expert system for strategic control of accidents and insurers' risks in building construction projects', *Expert Systems with Applications*, 36(2), pp. 4021–4034. doi:10.1016/J.ESWA.2008.02.029.

Iong-Zong Chen, J. and Lai, K.-L. (2021) 'Deep Convolution Neural Network Model for Credit-Card Fraud Detection and Alert', *Journal of Artificial Intelligence and Capsule Networks*, 03(02), pp. 101–112. doi:10.36548/jaicn.2021.2.003.

Ishak, S.S.M. and Newton, S. (2016) 'An innovation resistance factor model', *Construction Economics and Building*, 16(3), pp. 87–103. doi:10.5130/AJCEB.v16i3.5164.

Jallan, Y. *et al.* (2019) 'Application of Natural Language Processing and Text Mining to Identify Patterns in Construction-Defect Litigation Cases', *Journal of Legal Affairs and Dispute Resolution in Engineering and Construction*, 11(4), p. 04519024. doi:10.1061/(ASCE)LA.1943-4170.0000308.

Jang, S. and Lee, G. (2018) 'Impact of organizational factors on delays in bim-based coordination from a decision-making view: A case study', *Journal of Civil Engineering and Management*, 24(1), pp. 19–30. doi:10.3846/jcem.2018.296.

Ji, Y. *et al.* (2018) 'Assessing and prioritising delay factors of prefabricated concrete building projects in China', *Applied Sciences (Switzerland)*, 8(11), p. 2324. doi:10.3390/app8112324.

Jin, R., Tang, L. and Fang, K. (2015) 'Investigation into the current stage of BIM application in China's AEC industries'. doi:10.2495/BIM150401.

Johansen, T.F. (2015) 'How does BIM contributes to LEAN?', *KEA-Copenhagen School of Design and Technology*, p. 57. Available at: https://buildingsmart.no/sites/buildingsmart.no/files/2015_kea_thomas_felipe_johansen_how_does_bim_contributes_to_lean.pdf (Accessed: 27 April 2020).

Johnson, M. *et al.* (2021) ‘Impact of Big Data and Artificial Intelligence on Industry: Developing a Workforce Roadmap for a Data Driven Economy’, *Global Journal of Flexible Systems Management*, 22(3), pp. 197–217. doi:10.1007/S40171-021-00272-Y/METRICS.

Johnston, R.P.D. *et al.* (2018) ‘Sustainability of Cold-formed Steel Portal Frames in Developing Countries in the Context of Life Cycle Assessment and Life Cycle Costs’, *Structures*, 13, pp. 79–87. doi:10.1016/J.ISTRUC.2017.11.003.

Jones, D. and Dewberry, E. (2012) *Building Information Modelling design ecologies-a new model?*

Jones, S.A., Young Jr., N.W. and Bernstein, H.M. (2008) ‘Building Information Modeling (BIM): Transforming Design and Construction to Achieve Greater Industry Productivity’, *McGraw Hill Construction - SmartMarket Report*, p. 45.

Jung, K., Chu, B. and Hong, D. (2013) ‘Robot-based construction automation: An application to steel beam assembly (Part II)’, *Automation in Construction*, 32, pp. 62–79. doi:10.1016/J.AUTCON.2012.12.011.

Kassotakis, N. and Sarhosis, V. (2021) ‘Employing non-contact sensing techniques for improving efficiency and automation in numerical modelling of existing masonry structures: A critical literature review’, *Structures*, 32, pp. 1777–1797. doi:10.1016/J.ISTRUC.2021.03.111.

Kermanshahi, E.K. *et al.* (2020) ‘Implementation of Building Information Modeling for Construction Clash Detection Process in the Design Stage: A Case Study of Malaysian Police Headquarter Building’, *IOP Conference Series: Earth and Environmental Science*, 476(1). doi:10.1088/1755-1315/476/1/012009.

Keshavarzi, M. *et al.* (2020) ‘GenScan: A Generative Method for Populating Parametric 3D Scan Datasets’, *Projections - Proceedings of the 26th International Conference of the Association for Computer-Aided Architectural Design Research in Asia, CAADRIA 2021*, 1, pp. 91–100. doi:10.48550/arxiv.2012.03998.

Khalesi, H. *et al.* (2020) ‘Application of hybrid swara–bim in reducing reworks of building construction projects from the perspective of time’, *Sustainability (Switzerland)*, 12(21), pp. 1–20. doi:10.3390/su12218927.

Khattab, M.M. *et al.* (2020) ‘Regularization-based multi-frame super-resolution: A systematic review’, *Journal of King Saud University - Computer and Information Sciences*, 32(7), pp. 755–762. doi:10.1016/J.JKSUCI.2018.11.010.

Khodadadi, N. *et al.* (2023) ‘BAOA: Binary Arithmetic Optimization Algorithm with K-Nearest Neighbor Classifier for Feature Selection’, *IEEE Access*, 11, pp. 94094–94115. doi:10.1109/ACCESS.2023.3310429.

Kim, J.M. *et al.* (2021) ‘Development of Model to Predict Natural Disaster-Induced Financial Losses for Construction Projects Using Deep Learning Techniques’, *Sustainability 2021, Vol. 13, Page 5304*, 13(9), p. 5304. doi:10.3390/SU13095304.

Kim, T. and Chi, S. (2019) ‘Accident Case Retrieval and Analyses: Using Natural Language Processing in the Construction Industry’, *Journal of Construction Engineering and Management*, 145(3), p. 04019004. doi:10.1061/(ASCE)CO.1943-7862.0001625.

Kılkış, Ş. (2021) ‘Better security and protection for people and ecological systems: integrated approaches for decoupling urban growth from emission pressures in megacities’, *Sustainable Mega City Communities*, pp. 73–93. doi:10.1016/B978-0-12-818793-7.00004-4.

Ko, C.H., Cheng, M.Y. and Wu, T.K. (2007) ‘Evaluating sub-contractors performance using EFNIM’, *Automation in Construction*, 16(4), pp. 525–530. doi:10.1016/J.AUTCON.2006.09.005.

Koc, K., Ekmekcioğlu, Ö. and Gurgun, A.P. (2021) ‘Integrating feature engineering, genetic algorithm and tree-based machine learning methods to predict the post-accident disability status of construction workers’, *Automation in Construction*, 131, p. 103896. doi:10.1016/J.AUTCON.2021.103896.

Kong, X. *et al.* (2022) ‘Latent variable models in the era of industrial big data: Extension and beyond’, *Annual Reviews in Control*, 54, pp. 167–199. doi:10.1016/J.ARCONTROL.2022.09.005.

Kontovourkis, O. and Konatzii, P. (2021) ‘Environmental and cost assessment of customized modular wall components production based on an adaptive formwork casting mechanism: An experimental study’, *Journal of Cleaner Production*, 286, p. 125380. doi:10.1016/J.JCLEPRO.2020.125380.

Koo, T.K. and Tiong, R. (1993) ‘An expert system for assessing the performance of RC beams and slabs’, *Construction Management and Economics*, 11(5), pp. 347–357. doi:10.1080/01446199300000039.

Kothari, C. (2012) ‘Research Methods: Methods and Techniques’, pp. 1–414.

Krieg, O.D. and Lang, O. (2019) ‘Adaptive automation strategies for robotic prefabrication of parametrized mass timber building components’, *Proceedings of the 36th International Symposium on Automation and Robotics in Construction, ISARC 2019*, pp. 521–528. doi:10.22260/ISARC2019/0070.

Kruachottikul, P. *et al.* (2021) ‘Deep learning-based visual defect-inspection system for reinforced concrete bridge substructure: a case of Thailand’s department of highways’, *Journal of Civil Structural Health Monitoring 2021 11:4*, 11(4), pp. 949–965. doi:10.1007/S13349-021-00490-Z.

Kumar R, D. (2016) ‘Causes and Effects of Delays in Construction Industry’, *International Research Journal of Engineering and Technology*, 3(4), pp. 1831–1837. Available at: www.irjet.net.

Latiffi, A.A. *et al.* (2013) ‘Building Information Modeling (BIM) Application in Malaysian Construction Industry’, 2, pp. 1–6. doi:10.5923/s.ijcem.201309.01.

Lee, J., Yi, J.-S. and Son, J. (2019) ‘Development of Automatic-Extraction Model of Poisonous Clauses in International Construction Contracts Using Rule-Based NLP’, *Journal of Computing in Civil Engineering*, 33(3), p. 04019003. doi:10.1061/(ASCE)CP.1943-5487.0000807.

Lee, N. and Lings, I. (2008) ‘Concepts, constructs and measurement’, *Doing business research: a guide to theory and practice*. Edited by L.& Lings, pp. 134–155.

Lee, Y.-C., Scarpiniti, M. and Uncini, A. (2020) ‘Advanced Sound Classifiers and Performance Analyses for Accurate Audio-Based Construction Project Monitoring’, *Journal of Computing in Civil Engineering*, 34(5), p. 04020030. doi:10.1061/(ASCE)CP.1943-5487.0000911.

Lee, Y.C. *et al.* (2020) ‘Evidence-driven sound detection for prenotification and identification of construction safety hazards and accidents’, *Automation in Construction*, 113, p. 103127. doi:10.1016/J.AUTCON.2020.103127.

Li, C.Z. *et al.* (2017) ‘Integrating RFID and BIM technologies for mitigating risks and improving schedule performance of prefabricated house construction’, *Journal of Cleaner Production*, 165, pp. 1048–1062. doi:10.1016/j.jclepro.2017.07.156.

Li, H. (1996) ‘Case-based reasoning for intelligent support of construction negotiation’, *Information and Management*, 30(5), pp. 231–238. doi:10.1016/S0378-7206(96)01058-0.

Li, H., Luo, X. and Skitmore, M. (2020) ‘Intelligent Hoisting with Car-Like Mobile Robots’, *Journal of Construction Engineering and Management*, 146(12), p. 04020136. doi:10.1061/(ASCE)CO.1943-7862.0001931.

Liang, J. *et al.* (2019) ‘Weighted Graph Embedding-Based Metric Learning for Kinship Verification’, *IEEE Transactions on Image Processing*, 28(3), pp. 1149–1162. doi:10.1109/TIP.2018.2875346.

Lim, T.Y. *et al.* (2022) ‘An information entropy-based evolutionary computation for multi-factorial optimization’, *Applied Soft Computing*, 114, p. 108071. doi:10.1016/J.ASOC.2021.108071.

Ling, J. *et al.* (2023) ‘Data Modeling Techniques for Pipeline Integrity Assessment: A State-of-the-Art Survey’, *IEEE Transactions on Instrumentation and Measurement*, 72. doi:10.1109/TIM.2023.3279910.

Liu, D. *et al.* (2020) ‘Real-Time Optimization of Precast Concrete Component Transportation and Storage’, *Advances in Civil Engineering*, 2020. doi:10.1155/2020/5714910.

Liu, H., Zhou, M. and Liu, Q. (2019) ‘An embedded feature selection method for imbalanced data classification’, *IEEE/CAA Journal of Automatica Sinica*, 6(3), pp. 703–715. doi:10.1109/JAS.2019.1911447.

Liu, R. and Liu, F. (2020) *Research on bim and mobile equipment in substation construction schedule management*, *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Springer International Publishing. doi:10.1007/978-3-030-49183-3_5.

Liu, W. and Wang, J. (2021) ‘Recursive elimination–election algorithms for wrapper feature selection’, *Applied Soft Computing*, 113, p. 107956. doi:10.1016/J.ASOC.2021.107956.

Lomio, F. *et al.* (2018) ‘Classification of Building Information Model (BIM) Structures with Deep Learning’, in *2018 7th European Workshop on Visual Information Processing (EUVIP)*. IEEE, pp. 1–6. doi:10.1109/EUVIP.2018.8611701.

Love, P.E.D. *et al.* (2017) ‘Light rail transit cost performance: Opportunities for future-proofing’, *Transportation Research Part A: Policy and Practice*, 100, pp. 27–39. doi:10.1016/J.TRA.2017.04.002.

Lyytinen, K. and Damsgaard, J. (no date) 'What's Wrong with the diffusion of innovation theory?'

Ma, Y. *et al.* (2020) 'Artificial intelligence applications in the development of autonomous vehicles: A survey', *IEEE/CAA Journal of Automatica Sinica*, 7(2), pp. 315–329. doi:10.1109/JAS.2020.1003021.

Macdonald, M. (2002) 'Review of Large Public Procurement in the UK', *HM Treasury*, 44(April), pp. 1–83.

Mahfouz, T. and Kandil, A. (2012) 'Litigation outcome prediction of differing site condition disputes through machine learning models', *Journal of Computing in Civil Engineering*, 26(3), pp. 298–308. doi:10.1061/(ASCE)CP.1943-5487.0000148.

Mahjoubi, S. *et al.* (2021) 'Prediction and multi-objective optimization of mechanical, economical, and environmental properties for strain-hardening cementitious composites (SHCC) based on automated machine learning and metaheuristic algorithms', *Journal of Cleaner Production*, 329, p. 129665. doi:10.1016/J.JCLEPRO.2021.129665.

Malacarne, G. *et al.* (2018) 'Investigating benefits and criticisms of bim for construction scheduling in SMEs: An Italian case study', *International Journal of Sustainable Development and Planning*, 13(1), pp. 139–150. doi:10.2495/SDP-V13-N1-139-150.

Maldonado, J., Riff, M.C. and Neveu, B. (2022) 'A review of recent approaches on wrapper feature selection for intrusion detection', *Expert Systems with Applications*, 198, p. 116822. doi:10.1016/J.ESWA.2022.116822.

Malik, M., Khatana, R. and Kaushik, A. (2021) 'Machine learning with health care: A perspective', in Jain, V. and Chatterjee, J.M. (eds) *Journal of Physics: Conference Series*. Cham: Springer International Publishing (Learning and Analytics in Intelligent Systems). doi:10.1088/1742-6596/2040/1/012022.

Manoharan, S. (2019) 'AN IMPROVED SAFETY ALGORITHM FOR ARTIFICIAL INTELLIGENCE ENABLED PROCESSORS IN SELF DRIVING CARS', *Journal of Artificial Intelligence and Capsule Networks*, 01, pp. 95–104. doi:10.36548/jaicn.2019.2.005.

Mansfield, N., Ugwu, O. and Doran, T. (1994) 'Causes of delay and cost overruns in Nigerian construction projects', *International Journal of Project Management*, 12(4), pp. 254–260. doi:10.1016/0263-7863(94)90050-7.

Mark, E.-S. and Lyles, M.A. (2005) *The Blackwell Handbook of Organizational Learning and Knowledge Management*. Available at: https://www.wiley.com/en-us/9780631226727 (Accessed: 8 October 2022).

Marks, M. (2017) *Construction: The next great tech transformation Voices Michael Marks*.

May, T. (2011) *Social Research: Issues, Methods and Process (4th edition): Interviewing: methods and process.* Open University Press: McGraw. Available at: http://www.mcgraw-hill.co.uk/html/0335235670.html (Accessed: 8 October 2022).

McKinsey, I. (2022) *Imagining construction's digital future*. Available at: https://www.mckinsey.com/capabilities/operations/our-insights/imagining-constructions-digital-future#/ (Accessed: 30 July 2024).

Mei, S. *et al.* (2024) 'A Novel Center-Boundary Metric Loss to Learn Discriminative Features for Hyperspectral Image Classification', *IEEE Transactions on Geoscience and Remote Sensing*, 62, pp. 1–16. doi:10.1109/TGRS.2024.3362391.

Mei Yee, J.N. *et al.* (2019) 'OPTIMISING IMAGE CLASSIFICATION Implementation of Convolutional Neural Network Algorithms to Distinguish Between Plans and Sections within the Architectural, Engineering and Construction (AEC) Industry'.

Miao, J. and Zhu, W. (2022) 'Precision–recall curve (PRC) classification trees', *Evolutionary Intelligence*, 15(3), pp. 1545–1569. doi:10.1007/S12065-021-00565-2/TABLES/32.

Milošević, I., Kovačević, M. and Petronijević, P. (2021) 'Estimating Residual Value of Heavy Construction Equipment Using Ensemble Learning', *Journal of Construction Engineering and Management*, 147(7), p. 04021073. doi:10.1061/(ASCE)CO.1943-7862.0002088.

Mnich, K. and Rudnicki, W.R. (2020) 'All-relevant feature selection using multidimensional filters with exhaustive search', *Information Sciences*, 524, pp. 277–297. doi:10.1016/J.INS.2020.03.024.

Mohd, S. and Latiffi, A.A. (2013) *Building Information Modeling (BIM): What?, When? and Why?*

Moher, D. *et al.* (2009) 'Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement', *PLoS Medicine*, 6(7). doi:10.1371/journal.pmed.1000097.

Molnar, C. (2022) *Interpretable Machine Learning,A Guide for Making Black Box Models Explainable*. Available at: https://christophm.github.io/interpretable-ml-book.

Moselhi, O., Bardareh, H. and Zhu, Z. (2020) 'Automated data acquisition in construction with remote sensing technologies', *Applied Sciences (Switzerland)*, 10(8). doi:10.3390/APP10082846.

Mosey, D. (2014) 'PROJECT PROCUREMENT AND DELIVERY GUIDANCE', *Kings College London* [Preprint]. Available at: https://www.gov.uk/government/collections/government-construction. (Accessed: 30 July 2024).

Motaleb, O. and Kishk, M. (2010) 'An investigation into causes and effects of construction delays in UAE', in *Association of Researchers in Construction Management, ARCOM 2010 - Proceedings of the 26th Annual Conference*, pp. 1149–1157. Available at: https://www.researchgate.net/publication/266174953 (Accessed: 24 April 2020).

Muqeem, S. *et al.* (2012) 'Application of Fuzzy expert systems for construction labor productivity estimation', *2012 International Conference on Computer and Information Science, ICCIS 2012 - A Conference of World Engineering, Science and Technology Congress, ESTCON 2012 - Conference Proceedings*, 1, pp. 506–511. doi:10.1109/ICCISCI.2012.6297298.

Mushava, J. and Murray, M. (2024) ‘Flexible loss functions for binary classification in gradient-boosted decision trees: An application to credit scoring’, *Expert Systems with Applications*, 238, p. 121876. doi:10.1016/J.ESWA.2023.121876.

Narlawar G.S., Chaphalkar N.B., S.S. (2017) ‘Strategies to improve cost and time control using building information model (BIM); Conceptual paper’, *International Conference on Innovation and Management (ICIM) 2016*, 2(2), p. 270. Available at: https://www.cream.my/main/index.php/publication/malaysian-construction-research-journal-mcrj?download=46:mcrj-special-issue-volume-1-no-1-2017.

Narlawar, G.S., Chaphalkar, N.B. and Sandbhor, S. (2019) ‘Time and resource management of residential apartment construction using building information modeling’, *International Journal of Innovative Technology and Exploring Engineering*, 8(10), pp. 4238–4246. doi:10.35940/ijitee.J9951.0881019.

Nawi, M.N.M. *et al.* (2014) ‘Improving integrated practice through Building Information Modeling-Integrated Project Delivery (BIM-IPD) for Malaysian Industrialised Building System (IBS) construction projects’, *Malaysian Construction Research Journal*, 15(2), pp. 29–38.

Neuman, W.L. (2014) ‘Social Research Methods: Qualitative and Quantitative Approaches W. Lawrence Neuman Seventh Edition’. Available at: www.pearsoned.co.uk (Accessed: 27 October 2022).

Niwas, S.I. *et al.* (2015) ‘Reliable Feature Selection for Automated Angle Closure Glaucoma Mechanism Detection’, *Journal of Medical Systems*, 39(3), pp. 1–10. doi:10.1007/S10916-015-0199-1/TABLES/4.

Nokhwal, S. and Kumar, N. (2023) ‘RTRA: Rapid Training of Regularization-based Approaches in Continual Learning’, *2023 10th International Conference on Soft Computing and Machine Intelligence, ISCMI 2023*, pp. 188–192. doi:10.1109/ISCMI59957.2023.10458644.

Norrdine, A. and Motzko, C. (2020) ‘An internet of things based transportation cart for smart construction site’, *Proceedings - IEEE Congress on Cybermatics: 2020 IEEE International Conferences on Internet of Things, iThings 2020, IEEE Green Computing and Communications, GreenCom 2020, IEEE Cyber, Physical and Social Computing, CPSCom 2020 and IEEE Smart Data, SmartD*, pp. 160–167. doi:10.1109/iThings-GreenCom-CPSCom-SmartData-Cybermatics50389.2020.00042.

Ocheoha, I.A. and Moselhi, O. (2013) ‘Impact of Building Information Modeling on Just-In-Time material delivery’, in *ISARC 2013 - 30th International Symposium on Automation and Robotics in Construction and Mining, Held in Conjunction with the 23rd World Mining Congress*. Canadian Institute of Mining, Metallurgy and Petroleum, pp. 793–801. doi:10.22260/isarc2013/0086.

Odeyinka, H. and Adebayo, Y. (1997) ‘“The Causes and Effects of Construction Delays on Completion Cost of Housing Projects in Nigeria”’, *Journal of Financial Management of Property and Construction*, 2(3), pp. 31-44. Available at: https://www.researchgate.net/publication/249643683

(Accessed: 24 April 2020).

Oladinrin, T., Ogunsemi, D. and Aje, I. (2012) ‘Role of Construction Sector in Economic Growth: Empirical Evidence from Nigeria’, *FUTY Journal of the Environment*, 7(1), pp. 50–60. doi:10.4314/fje.v7i1.4.

Orangi, A., Palaneeswaran, E. and Wilson, J. (2011) ‘Exploring delays in Victoria-based Astralian pipeline projects’, *Procedia Engineering*, 14, pp. 874–881. doi:10.1016/j.proeng.2011.07.111.

Owalabi, J.D. *et al.* (2014) ‘Causes and effects of delay on project construction delivery time’, *International Journal of Education and Research*, 2(4), pp. 197–208. Available at: www.ijern.com (Accessed: 24 April 2020).

Owolabi, H.O. *et al.* (2018) ‘Predicting Completion Risk in PPP Projects using Big Data Analytics’, *IEEE Transactions on Engineering Management* [Preprint]. doi:10.1109/TEM.2018.2876321.

P.S.Kulkarni1, S.N.L.M.C.D. (2017) ‘Artificial Neural Networks for Construction Management: A Review’, *Journal of Soft Computing in Civil Engineering1-2(2017)*, pp. 70–88. doi:10.22115/SCCE.2017.49580.

Padilla, R. *et al.* (2021) ‘A Comparative Analysis of Object Detection Metrics with a Companion Open-Source Toolkit’, *Electronics 2021, Vol. 10, Page 279*, 10(3), p. 279. doi:10.3390/ELECTRONICS10030279.

Padilla, R., Netto, S.L. and Da Silva, E.A.B. (2020) ‘A Survey on Performance Metrics for Object-Detection Algorithms’, *International Conference on Systems, Signals, and Image Processing*, 2020-July, pp. 237–242. doi:10.1109/IWSSIP48289.2020.9145130.

Pahlevan-Sharif, S., Mura, P. and Wijesinghe, S.N.R. (2019) ‘A systematic review of systematic reviews in tourism’, *Journal of Hospitality and Tourism Management*, 39(November 2018), pp. 158–165. doi:10.1016/j.jhtm.2019.04.001.

Pahlevan Sharif, S., Mura, P. and Wijesinghe, S.N.R. (2019a) ‘Systematic Reviews in Asia: Introducing the “PRISMA” Protocol to Tourism and Hospitality Scholars’, pp. 13–33. doi:10.1007/978-981-13-2463-5_2.

Pahlevan Sharif, S., Mura, P. and Wijesinghe, S.N.R. (2019b) ‘Systematic Reviews in Asia: Introducing the “PRISMA” Protocol to Tourism and Hospitality Scholars’, in. Springer, Singapore, pp. 13–33. doi:10.1007/978-981-13-2463-5_2.

Palma, V. (2019) ‘TOWARDS DEEP LEARNING for ARCHITECTURE: A MONUMENT RECOGNITION MOBILE APP’, *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 42(2/W9), pp. 551–556. doi:10.5194/ISPRS-ARCHIVES-XLII-2-W9-551-2019.

Park, J. *et al.* (2017) ‘Database-Supported and Web-Based Visualization for Daily 4D BIM’, *Journal of Construction Engineering and Management*, p. 04017078. doi:10.1061/(asce)co.1943-7862.0001392.

Park, J.H. and Lee, G. (2017) ‘Design coordination strategies in a 2D and BIM mixed-project environment: social dynamics and productivity’, *Building Research and Information*, 45(6), pp. 631–

648. doi:10.1080/09613218.2017.1288998.

Parveen, R. (2018) ‘Legal Issues and Regulatory Challenges’, *International Journal of Civil Engineering and Technology*, 9(13), pp. 957–962. Available at: http://iaeme.com/Home/issue/IJCIET?Volume=9&Issue=13http://iaeme.comhttp://iaeme.com/Home/journal/IJCIET958 (Accessed: 31 January 2022).

Pedregosa, F. *et al.* (2011) ‘Scikit-learn: Machine Learning in Python’, *Journal of Machine Learning Research*, 12(85), pp. 2825–2830. Available at: http://scikit-learn.sourceforge.net. (Accessed: 7 January 2021).

Pereira, P.F., Ramos, N.M.M. and Simões, M.L. (2019) ‘Data-driven occupant actions prediction to achieve an intelligent building’, *https://doi.org/10.1080/09613218.2019.1692648*, 48(5), pp. 485–500. doi:10.1080/09613218.2019.1692648.

Pham, A.D. *et al.* (2020) ‘Hybrid machine learning for predicting strength of sustainable concrete’, *Soft Computing 2020 24:19*, 24(19), pp. 14965–14980. doi:10.1007/S00500-020-04848-1.

Piroozfar, P. *et al.* (2019) ‘Facilitating Building Information Modelling (BIM) using Integrated Project Delivery (IPD): A UK perspective’, *Journal of Building Engineering*, 26, p. 100907. doi:10.1016/J.JOBE.2019.100907.

Poh, C.Q.X., Ubeynarayana, C.U. and Goh, Y.M. (2018) ‘Safety leading indicators for construction sites: A machine learning approach’, *Automation in Construction*, 93, pp. 375–386. doi:10.1016/j.autcon.2018.03.022.

Puggini, L. and McLoone, S. (2017) ‘Forward Selection Component Analysis: Algorithms and Applications’, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(12), pp. 2395–2408. doi:10.1109/TPAMI.2017.2648792.

PwC (2021) ‘The Potential Impact of Artificial Intelligence on UK Employment and the Demand for Skills. A report by PwC for the Department for Business, Energy and Industrial Strategy’, *BEIS Research Report Number: 2021/042* [Preprint], (August). Available at: www.pwc.com/structure (Accessed: 22 December 2023).

Rahkovsky, I. *et al.* (2021) ‘AI Research Funding Portfolios and Extreme Growth’, *Frontiers in Research Metrics and Analytics*, 0, p. 11. doi:10.3389/FRMA.2021.630124.

Raisbeck, P., Duffield, C. and Xu, M. (2010) ‘Comparative performance of PPPs and traditional procurement in australia’, *Construction Management and Economics*, 28(4), pp. 345–359. doi:10.1080/01446190903582731.

Randolph, D.A., Rajendra, K. and Campfield, J.J. (1987) ‘Using risk management techniques to control contract costs’, *Journal of Management in Engineering*, 3(4), pp. 314–324. doi:10.1061/(ASCE)9742-597X(1987)3:4(314).

Raymaekers, J. and Rousseeuw, P.J. (2021) ‘Transforming variables to central normality’, *Machine Learning*, pp. 1–23. doi:10.1007/S10994-021-05960-5/FIGURES/13.

Rhodes, C. (2019) ‘Construction Industry: Statistics and policy’, *House of Commons Library*, (01432), pp. 1–13.

Richards, K. (2003) ‘Qualitative inquiry in TESOL’, *Qualitative Inquiry in TESOL*, pp. 1–323. doi:10.1057/9780230505056/COVER.

Richhariya, B., Tanveer, M. and Rashid, A.H. (2020) ‘Diagnosis of Alzheimer’s disease using universum support vector machine based recursive feature elimination (USVM-RFE)’, *Biomedical Signal Processing and Control*, 59, p. 101903. doi:10.1016/J.BSPC.2020.101903.

Rijo George, M., Nalluri, M.R. and Anand, K.B. (2021) ‘Severity Prediction of Construction Site Accidents Using Simple and Ensemble Decision Trees’, *Lecture Notes in Civil Engineering*, 171, pp. 599–608. doi:10.1007/978-3-030-80312-4_50.

Ritchie, J. (2003) ‘Qualitative research practice: a guide for social science students and researchers’, *Choice Reviews Online*, 41(03), pp. 41-1319-41–1319. doi:10.5860/choice.41-1319.

Ritchie, J. and Lewis, J. (2003) ‘QUALITATIVE RESEARCH PRACTICE A Guide for Social Science Students and Researchers Edited by’.

Rodgers, E.M. (2003) *Diffusion of Innovations, 5th Edition Everett M. Rogers*. Simon & Schuster. Available at: https://ebookcentral.proquest.com/lib/ethz/reader.action?docID=4935198&ppg=1 (Accessed: 31 July 2022).

Rose, S., Spinks, N. and Canhoto, A. (2014) *Management Research*, *Management Research*. SAGE. doi:10.4324/9781315819198.

Rouse, J. (1997) ‘Business Research: A Practical Guide for Undergraduate and Postgraduate Students’, *International Small Business Journal*, 15(4), pp. 103–105. Available at: https://go.gale.com/ps/i.do?p=AONE&sw=w&issn=02662426&v=2.1&it=r&id=GALE%7CA19737914&sid=googleScholar&linkaccess=fulltext (Accessed: 8 October 2022).

Roy, A. and Chakraborty, S. (2023) ‘Support vector machine in structural reliability analysis: A review’, *Reliability Engineering and System Safety*. Elsevier, p. 109126. doi:10.1016/j.ress.2023.109126.

Royal Society, T. (2017) *MACHINE LEARNING: THE POWER AND PROMISE OF COMPUTERS THAT LEARN BY EXAMPLE*.

Russell, K. (2010) ‘The Art of Being a Scientist: A Guide for Graduate Students and their Mentors by Roel Snieder, Ken Larner’, *International Statistical Review*, 78(1), pp. 159–159. doi:10.1111/j.1751-5823.2010.00109_28.x.

Ryu, D.W. *et al.* (2015) ‘Evaluating risks using simulated annealing and Building Information Modeling’, *Applied Mathematical Modelling*, 39(19), pp. 5925–5935. doi:10.1016/j.apm.2015.04.024.

Saha, P., Patikar, S. and Neogy, S. (2020) ‘A correlation - Sequential forward selection based feature selection method for healthcare data analysis’, *2020 IEEE International Conference on Computing, Power and Communication Technologies, GUCON 2020*, pp. 69–72. doi:10.1109/GUCON48875.2020.9231205.

Sahebi, G. *et al.* (2020) 'GeFeS: A generalized wrapper feature selection approach for optimizing classification performance', *Computers in Biology and Medicine*, 125, p. 103974. doi:10.1016/J.COMPBIOMED.2020.103974.

Saka, A. *et al.* (2024) 'GPT models in construction industry: Opportunities, limitations, and a use case validation', *Developments in the Built Environment*, 17, p. 100300. doi:10.1016/J.DIBE.2023.100300.

Saka, A.B. *et al.* (2023) 'Conversational artificial intelligence in the AEC industry: A review of present status, challenges and opportunities', *Advanced Engineering Informatics*, 55, p. 101869. doi:10.1016/J.AEI.2022.101869.

Saka, A.B. and Chan, D.W.M. (2021) 'BIM divide: an international comparative analysis of perceived barriers to implementation of BIM in the construction industry', *Journal of Engineering, Design and Technology*, ahead-of-p(ahead-of-print). doi:10.1108/JEDT-07-2021-0348.

Sami Ur Rehman, M. *et al.* (2020) 'Project schedule risk management through building information modelling', *International Journal of Construction Management*, 0(0), pp. 1–11. doi:10.1080/15623599.2020.1728606.

Sanni-Anibire, M.O., Mohamad Zin, R. and Olatunji, S.O. (2020) 'Causes of delay in the global construction industry: a meta analytical review', *International Journal of Construction Management* [Preprint]. doi:10.1080/15623599.2020.1716132.

Sanni-Anibire, M.O., Zin, R.M. and Olatunji, S.O. (2021) 'Machine learning - based framework for construction delay mitigation', *ITcon Vol. 26, pg. 303-318, http://www.itcon.org/2021/17*, 26(17), pp. 303–318. doi:10.36680/J.ITCON.2021.017.

Šatrevičs, V. (no date) 'Commercialization Potential for Deep Machine Learning Technology Using Line Scan Camera'.

Saunders, M.A., Lewis, P. and Thornhill, A. (2019) *RESEARCH METHODS FOR BUSINESS STUDENTS SIXTH EDITION*, *Research Methods for Business Students*. Available at: www.pearsoned.co.uk (Accessed: 26 September 2021).

Schia, M H *et al.* (no date) 'THE INTRODUCTION OF AI IN THE CONSTRUCTION INDUSTRY AND ITS IMPACT ON HUMAN BEHAVIOR', pp. 903–914. doi:10.24928/2019/0191.

Sepasgozar, S.M.E. *et al.* (2019) *Delay causes and emerging digital tools: A novel model of delay analysis, including integrated project delivery and PMBOK*, *Buildings*. doi:10.3390/buildings9090191.

Shah, R.K. (2016) 'An Exploration of Causes for Delay and Cost Overruns In Construction Projects: Case Study of Australia, Malaysia & Ghana', *Journal of Advanced College of Engineering and Management*, 2, p. 41. doi:10.3126/jacem.v2i0.16097.

Sharot, T. (2011) 'The optimism bias', *Current Biology*. Cell Press. doi:10.1016/j.cub.2011.10.030.

Shebob, A. *et al.* (2012) 'Comparative study of delay factors in Libyan and the UK construction industry', *Engineering, Construction and Architectural Management*, 19(6), pp. 688–712. doi:10.1108/09699981211277577.

Shehadeh, A. *et al.* (2021) ‘Machine learning models for predicting the residual value of heavy construction equipment: An evaluation of modified decision tree, LightGBM, and XGBoost regression’, *Automation in Construction*, 129, p. 103827. doi:10.1016/J.AUTCON.2021.103827.

Shin, M.H., Lee, H.K. and Kim, H.Y. (2018) ‘Benefit-Cost analysis of Building Information Modeling (BIM) in a Railway Site’, *Sustainability (Switzerland)*, 10(11), pp. 1–10. doi:10.3390/su10114303.

Siddiq, A.A. (2018) ‘BIM – Evolution and Emerging Research Trends’, *International Journal of Research* [Preprint].

Sierra, E.A. *et al.* (2007) ‘Building automation by intelligent control of its environment’, *Renewable Energy and Power Quality Journal*, 1(5), pp. 21–24. doi:10.24084/repqj05.204.

Sigalov, K. and König, M. (2018) ‘4d bim model adaptation based on construction progress monitoring’, in *eWork and eBusiness in Architecture, Engineering and Construction - Proceedings of the 12th European Conference on Product and Process Modelling, ECPPM 2018*. CRC Press/Balkema, pp. 337–344. doi:10.1201/9780429506215-42.

Silverio, A.K. and Suresh, S. (2021) ‘Status of BIM implementation in the Dominican Republic construction industry – an empirical study’, *Journal of Engineering, Design and Technology* [Preprint]. doi:10.1108/JEDT-05-2021-0253/FULL/PDF.

Silverman, D. (2000) ‘Doing qualitative research: A handbook’, *SAGE, London* [Preprint].

Silverman, D. (2013) ‘Doing Qualitative Research A Practical HandbookFourth Edition’, *SAGE Publications*, p. 488.

Snape, D. and Spencer, L. (2003) ‘The foundations of qualitative research’, in *Qualitative research practice*, pp. 1–23. Available at: https://books.google.co.uk/books?hl=en&lr=&id=EQSIAwAAQBAJ&oi=fnd&pg=PA1&dq=Ormston,+R.,+Spencer,+L.,+Barnard,+M.,+%26+Snape,+D.+(2014).+The+foundations+of+qualitative+research.&ots=l-VRjqVw5P&sig=fGuWfy8h8-AmiRynv63i-Gvc8Fo (Accessed: 10 October 2022).

Soman, R.K. and Molina-Solana, M. (2022) ‘Automating look-ahead schedule generation for construction using linked-data based constraint checking and reinforcement learning’, *Automation in Construction*, 134, p. 104069. doi:10.1016/J.AUTCON.2021.104069.

Sonetti, G., Naboni, E. and Brown, M. (2018) ‘Exploring the Potentials of ICT Tools for Human-Centric Regenerative Design’, *Sustainability 2018, Vol. 10, Page 1217*, 10(4), p. 1217. doi:10.3390/SU10041217.

Spallone, R. and Palma, V. (2021) ‘Artificial intelligence and augmented reality: A possible continuum for the enhancement of built heritage’, *DISEGNARECON*, 14(26). doi:10.20365/DISEGNARECON.26.2021.16.

Srao, B.K., Rai, H.S. and Mann, K.S. (2018) ‘Why india should make it compulsory to go for BIM’, *Analyzing the Role of Risk Mitigation and Monitoring in Software Development*, pp. 266–277. doi:10.4018/978-1-5225-6029-6.ch017.

Stegnar, G. and Cerovšek, T. (2019) ‘Information needs for progressive BIM methodology supporting the holistic energy renovation of office buildings’, *Energy*, 173, pp. 317–331. doi:10.1016/j.energy.2019.02.087.

Studebaker, P. (2014) ‘BIM on the Run: Building information systems are revolutionizing how industry manages equipment data’, *Control*, 27(5), pp. 53–57.

Su, L. chu *et al.* (2021) ‘Smart performance-based design for building fire safety: Prediction of smoke motion via AI’, *Journal of Building Engineering*, 43, p. 102529. doi:10.1016/J.JOBE.2021.102529.

Subramani, T. and Ammai, A. (2018) ‘Maturing construction management up the BIM model & scheduling using Primavera’, *International Journal of Engineering and Technology(UAE)*, 7(3), pp. 1–4. doi:10.14419/ijet.v7i3.10.15617.

Suermann, P.C. and Issa, R.R.A. (2009) ‘Dynamic prototyping: The united states air force building information modeling initiative’, *Proceedings of the 2009 ASCE International Workshop on Computing in Civil Engineering*, pp. 485–494. doi:10.1061/41052(346)48.

Sullivan, A. and Harris, F.C. (1986) ‘Delays on Large Construction Projects’, *International Journal of Operations & Production Management*, 6(1), pp. 25–33. doi:10.1108/eb054752.

Sun, J. *et al.* (2024) ‘A Memristive Fully Connect Neural Network and Application of Medical Image Encryption Based on Central Diffusion Algorithm’, *IEEE Transactions on Industrial Informatics*, 20(3), pp. 3778–3788. doi:10.1109/TII.2023.3312405.

Sun, J. and Xia, Y. (2024) ‘Pretreating and normalizing metabolomics data for statistical analysis’, *Genes & Diseases*, 11(3), p. 100979. doi:10.1016/J.GENDIS.2023.04.018.

Sun, L. *et al.* (2021) ‘Feature selection using Fisher score and multilabel neighborhood rough sets for multilabel classification’, *Information Sciences*, 578, pp. 887–912. doi:10.1016/J.INS.2021.08.032.

Surendhra Babu, P.R. and Hayath Babu, N. (2018) ‘Using technology to achieve lean objectives’, in *IGLC 2018 - Proceedings of the 26th Annual Conference of the International Group for Lean Construction: Evolving Lean Construction Towards Mature Production Management Across Cultures and Frontiers*. The International Group for Lean Construction, pp. 1069–1078. doi:10.24928/2018/0543.

TAFESSE, S. (2021) ‘A Review on the Critical Factors Causing Delay of Delivery Time in Construction Projects’, *International Journal of Engineering Technologies IJET*, 6(4), pp. 69–81. doi:10.19072/ijet.815025.

Tahir, M.M. *et al.* (2018) ‘Improving cost and time control in construction using building information model (Bim): A review’, *Pertanika Journal of Science and Technology*, 26(1), pp. 21–36.

Tahir Muhammad, M. *et al.* (2019) ‘The impact of BIM application on construction delays and cost overrun in developing countries’, in *IOP Conference Series: Earth and Environmental Science*. Institute of Physics Publishing. doi:10.1088/1755-1315/357/1/012027.

Tai, S., Zhang, Y. and Li, T. (2021) ‘Factors affecting BIM application in China: a social network model', *Journal of Engineering, Design and Technology*, 19(2), pp. 373–384. doi:10.1108/JEDT-12-2019-0330/FULL/PDF.

Tainton B.E. (1990) 'The unit of analysis "problem" in educational research', *Queensland Researcher*, 6(1), pp. 4–19. Available at: https://www.iier.org.au/qjer/qr6/tainton.html (Accessed: 29 October 2022).

Talat Birgonul, M. *et al.* (2016) 'An expert system for the quantification of fault rates in construction fall accidents', *http://dx.doi.org/10.1080/10803548.2015.1123516*, 22(1), pp. 20–31. doi:10.1080/10803548.2015.1123516.

Talukder, M.A. *et al.* (2024) 'Machine learning-based network intrusion detection for big and imbalanced data using oversampling, stacking feature embedding and feature extraction', *Journal of Big Data*, 11(1), pp. 1–44. doi:10.1186/S40537-024-00886-W/TABLES/16.

Teng, J.Y. *et al.* (2013) 'Study on integrated project delivery construction project collaborative application based on building information model', in *Advanced Materials Research*, pp. 370–374. doi:10.4028/www.scientific.net/AMR.621.370.

Tharwat, A. (2018) 'Classification assessment methods', *Applied Computing and Informatics*, 17(1), pp. 168–192. doi:10.1016/J.ACI.2018.08.003/FULL/PDF.

Theerthagiri, P. (2022) 'Predictive analysis of cardiovascular disease using gradient boosting based learning and recursive feature elimination technique', *Intelligent Systems with Applications*, 16, p. 200121. doi:10.1016/J.ISWA.2022.200121.

Theng, D. and Bhoyar, K.K. (2024a) 'Feature selection techniques for machine learning: a survey of more than two decades of research', *Knowledge and Information Systems*, 66(3), pp. 1575–1637. doi:10.1007/S10115-023-02010-5/TABLES/6.

Theng, D. and Bhoyar, K.K. (2024b) 'Feature selection techniques for machine learning: a survey of more than two decades of research', *Knowledge and Information Systems*, 66(3), pp. 1575–1637. doi:10.1007/S10115-023-02010-5/TABLES/6.

Tidhar, N. *et al.* (2021) 'A Slack approach to optimised modularisation of prefabricated structures using a multi-variable modularisation index', *Structures*, 33, pp. 1235–1251. doi:10.1016/J.ISTRUC.2021.04.064.

Tserng, H.P., Ho, S.P. and Jan, S.H. (2014) 'Developing BIM-assisted as-built schedule management system for general contractors', *Journal of Civil Engineering and Management*, pp. 47–58. doi:10.3846/13923730.2013.851112.

Uncu, Ö. and Türkşen, I.B. (2007) 'A novel feature selection approach: Combining feature wrappers and filters', *Information Sciences*, 177(2), pp. 449–466. doi:10.1016/J.INS.2006.03.022.

Vacanas, Y. *et al.* (2015) 'Building Information Modelling (BIM) and Unmanned Aerial Vehicle (UAV) technologies in infrastructure construction project management and delay and disruption analysis', *Third International Conference on Remote Sensing and Geoinformation of the Environment (RSCy2015)*, 9535(September), p. 95350C. doi:10.1117/12.2192723.

Vahdani, B. *et al.* (2014) ‘A New Hybrid Model Based on Least Squares Support Vector Machine for Project Selection Problem in Construction Industry’, *Arabian Journal for Science and Engineering 2014 39:5*, 39(5), pp. 4301–4314. doi:10.1007/S13369-014-1032-8.

Vahdatikhaki, F. and Mawlana, M. (2017) ‘A framework for augmenting 4D visualization of construction projects with scheduling uncertainties’, *6th CSCE-CRC International Construction Specialty Conference 2017 - Held as Part of the Canadian Society for Civil Engineering Annual Conference and General Meeting 2017*, 1, pp. 70–78.

Van, L. T., Sang, N. M., & Viet, N.T. (2015) ‘A conceptual model of delay factors affecting government construction projects’, *ARPN Journal of Science and Technology*, (5(2)), pp. 92–100.

Varouqa, I.F. (2021) ‘Using Artificial Intelligence and computation Enhanced apply in neural network’, *Journal of Applied Science and Engineering*, 24(5), pp. 763–770. doi:10.6180/JASE.202110_24(5).0011.

Vasista, T.G. and Jakhanwal, M.P. (2023) ‘Cost Effective Methods in Construction Engineering’, *Civil Engineering and Urban Planning: An International Journal (CiVEJ)*, 10(2/3), pp. 01–08. doi:10.5121/civej.2023.10301.

Vázquez-Canteli, J.R. *et al.* (2019) ‘Fusing TensorFlow with building energy simulation for intelligent energy management in smart cities’, *Sustainable Cities and Society*, 45, pp. 243–257. doi:10.1016/J.SCS.2018.11.021.

Vilventhan, A., Razin, S. and Rajadurai, R. (2020) ‘4D BIM models for smart utility relocation management in urban infrastructure projects’, *Facilities*, 39(1–2), pp. 50–63. doi:10.1108/F-08-2019-0091.

Wagner, H.J. *et al.* (2020) ‘Flexible and transportable robotic timber construction platform – TIM’, *Automation in Construction*, 120, p. 103400. doi:10.1016/J.AUTCON.2020.103400.

Wang, D. and Li, J. (2021) ‘Artificial Intelligence Aided Prediction of Building Structure Anti-seismic’, *Proceedings - 5th International Conference on Computing Methodologies and Communication, ICCMC 2021*, pp. 1403–1407. doi:10.1109/ICCMC51019.2021.9418404.

Wang, J. *et al.* (2015) ‘An integrated approach for progress tracking in liquefied natural gas construction’, *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 9320, pp. 259–267. doi:10.1007/978-3-319-24132-6_33.

Wang, P. *et al.* (2019) ‘On Defining Artificial Intelligence’, *Journal of Artificial General Intelligence*, 10(2), pp. 2019–2022. doi:10.2478/jagi-2019-0002.

Wang, R., Wu, X.J. and Kittler, J. (2021) ‘Graph Embedding Multi-Kernel Metric Learning for Image Set Classification with Grassmannian Manifold-Valued Features’, *IEEE Transactions on Multimedia*, 23, pp. 228–242. doi:10.1109/TMM.2020.2981189.

Wang, T. (2018) • *China: construction industry’s contribution share to GDP 2018-2021* | *Statista*,

Statista. Available at: https://www.statista.com/statistics/1068213/china-construction-industry-gdp-contribution-share/ (Accessed: 18 April 2020).

Wang, T. (2019) *Value added of U.S. construction as a percentage of GDP 2018*, *Statista*. Available at: https://www.statista.com/statistics/192049/value-added-by-us-construction-as-a-percentage-of-gdp-since-2007/ (Accessed: 18 April 2020).

Wang, Z., Zhang, N. and Wang, S. (2022) 'Geometry control of special-shaped arch pylon considering seasonal temperature changes during construction', *Structures*, 36, pp. 416–427. doi:10.1016/J.ISTRUC.2021.12.022.

White, I. and Sidhu, I. (2005) 'Building the Scottish Parliament, The Holyrood Project', (January), p. 24.

White, M.J., Cunningham, L.C. and Titchener, K. (2011) 'Young drivers' optimism bias for accident risk and driving skill: Accountability and insight experience manipulations', *Accident Analysis and Prevention*, 43(4), pp. 1309–1315. doi:10.1016/j.aap.2011.01.013.

Wiles, R., Crow, G. and Pain, H. (2011) 'Innovation in qualitative research methods: A narrative review', *Qualitative Research*, 11(5), pp. 587–604. doi:10.1177/1468794111413227/ASSET/IMAGES/LARGE/10.1177_1468794111413227-FIG2.JPEG.

Woetzel, J. *et al.* (2017) 'Bridging Infrastructure Gaps Has the World Made Progress?', *McKinsey Global Institute*, (October), p. 10. Available at: https://www.mckinsey.com/~/media/McKinsey/Industries/Capital Projects and Infrastructure/Our Insights/Bridging infrastructure gaps has the world made progress/Bridging infrastructure gaps How has the world made progress v2/MGI-Bridging-infrastructure-gaps.

Wong, J.K.W., Zhou, J.X. and Chan, A.P.C. (2018) 'Exploring the linkages between the adoption of bim and design error reduction', *International Journal of Sustainable Development and Planning*, 13(1), pp. 108–120. doi:10.2495/SDP-V13-N1-108-120.

Wu, C. *et al.* (2022) 'Natural language processing for smart construction: Current status and future directions', *Automation in Construction*, 134, p. 104059. doi:10.1016/J.AUTCON.2021.104059.

Wu, Y. *et al.* (2018) 'CU-brick cable-driven robot for automated construction of complex brick structures: From simulation to hardware realisation', *2018 IEEE International Conference on Simulation, Modeling, and Programming for Autonomous Robots, SIMPAR 2018*, pp. 166–173. doi:10.1109/SIMPAR.2018.8376287.

Xiao, W. *et al.* (2019) 'A robust classification algorithm for separation of construction waste using NIR hyperspectral system', *Waste Management*, 90, pp. 1–9. doi:10.1016/J.WASMAN.2019.04.036.

Xie, Y. *et al.* (2019) 'Historical Accident and Injury Database-Driven Audio-Based Autonomous Construction Safety Surveillance', *Computing in Civil Engineering 2019: Data, Sensing, and Analytics - Selected Papers from the ASCE International Conference on Computing in Civil Engineering 2019*, pp. 105–113. doi:10.1061/9780784482438.014.

Xu, S. *et al.* (2020) 'Computer Vision Techniques in Construction: A Critical Review', *Archives of Computational Methods in Engineering 2020 28:5*, 28(5), pp. 3383–3397. doi:10.1007/S11831-020-09504-3.

Xue, F. and Yao, E. (2022) 'Adopting a random forest approach to model household residential relocation behavior', *Cities*, 125, p. 103625. doi:10.1016/J.CITIES.2022.103625.

Yan, L. *et al.* (2008) 'PP2A regulates the pro-apoptotic activity of FOXO1', *Journal of Biological Chemistry*, 283(12), pp. 7411–7420. doi:10.1074/jbc.M708083200.

Yan, L. (2016) *Expatriate manager's adaption and knowledge acquisition: Personal development in multi-national companies in China*, *Expatriate Manager's Adaption and Knowledge Acquisition: Personal Development in Multi-National Companies in China*. doi:10.1007/978-981-10-0053-9.

Yang, K. and Miller, G.J. (2017) *Handbook of Research Methods in Public Administration, Second Edition*, *Handbook of Research Methods in Public Administration, Second Edition*. doi:10.4324/9780203908303.

Yang, L.R., O'Connor, J.T. and Wang, C.C. (2006) 'Technology utilization on different sizes of projects and associated impacts on composite project success', *International Journal of Project Management*, 24(2), pp. 96–105. doi:10.1016/j.ijproman.2005.06.008.

Yangxuan and Zhaoqianjing (2021) 'Influence of virtual reality and 3D printing on architectural innovation evaluation based on quality of experience evaluation using fuzzy logic', *Journal of Intelligent & Fuzzy Systems*, 40(4), pp. 8501–8509. doi:10.3233/JIFS-189670.

Yaseen, Z.M. *et al.* (2020a) 'Prediction of risk delay in construction projects using a hybrid artificial intelligence model', *Sustainability (Switzerland)*, 12(4), pp. 1–14. doi:10.3390/su12041514.

Yaseen, Z.M. *et al.* (2020b) 'Prediction of Risk Delay in Construction Projects Using a Hybrid Artificial Intelligence Model', *Sustainability*, 12(4), p. 1514. doi:10.3390/su12041514.

Yaseen, Z.M. *et al.* (2020c) 'Prediction of Risk Delay in Construction Projects Using a Hybrid Artificial Intelligence Model', *Sustainability 2020, Vol. 12, Page 1514*, 12(4), p. 1514. doi:10.3390/SU12041514.

You, Z. and Feng, L. (2020) 'Integration of Industry 4.0 Related Technologies in Construction Industry: A Framework of Cyber-Physical System', *IEEE Access*, 8, pp. 122908–122922. doi:10.1109/ACCESS.2020.3007206.

Young, N.W. *et al.* (2008) *BIM Handbook A Guide to Building Information Modeling for Owners, Managers, Designers, Engineers, and Contractors BIM Handbook: A Guide to Building Informationn n Modeling for Owners, Managers*. Available at: www.wiley.com/go/permissions. (Accessed: 27 April 2020).

Yu, Y. *et al.* (2018) 'Estimating construction workers' physical workload by fusing computer vision and smart insole technologies', *ISARC 2018 - 35th International Symposium on Automation and Robotics in Construction and International AEC/FM Hackathon: The Future of Building Things* [Preprint].

doi:10.22260/ISARC2018/0168.

Yuan, C. and Yang, L. (2023) ‘An efficient multi-metric learning method by partitioning the metric space’, *Neurocomputing*, 529, pp. 56–79. doi:10.1016/J.NEUCOM.2023.01.074.

Yuk Carrie Lin, K. (2024) ‘Optimizing variable selection and neighbourhood size in the K-nearest neighbour algorithm’, *Computers & Industrial Engineering*, 191, p. 110142. doi:10.1016/J.CIE.2024.110142.

Zhang, F. *et al.* (2019) ‘Construction site accident analysis using text mining and natural language processing techniques’, *Automation in Construction*, 99, pp. 238–248. doi:10.1016/J.AUTCON.2018.12.016.

Zhang, H. *et al.* (2015) ‘A new method for nondestructive quality evaluation of the resistance spot welding based on the radar chart method and the decision tree classifier’, *International Journal of Advanced Manufacturing Technology*, 78(5–8), pp. 841–851. doi:10.1007/S00170-014-6654-1/METRICS.

Zhang, J. *et al.* (2018) ‘A RMM based word segmentation method for Chinese design specifications of building stairs’, *Proceedings - 14th International Conference on Computational Intelligence and Security, CIS 2018*, pp. 277–280. doi:10.1109/CIS2018.2018.00068.

Zhang, J. (2021) ‘Potential energy saving estimation for retrofit building with ASHRAE-Great Energy Predictor III using machine learning’, *ACM International Conference Proceeding Series*, pp. 425–429. doi:10.1145/3473714.3473788.

Zhang, J.P., Liu, L.H. and Coble, R.J. (2002) ‘Hybrid intelligence utilization for construction site layout’, *Automation in Construction*, 11(5), pp. 511–519. doi:10.1016/S0926-5805(01)00071-1.

Zhang, L. and Liu, H. (2021) ‘Research on Application Strategy of BIM Technology in Construction Project Management’, in *E3S Web of Conferences*. EDP Sciences. doi:10.1051/e3sconf/202123605002.

Zhang, R. and Li, D. (2011) ‘Development of risk assessment model in construction project using fuzzy expert system’, *ICEMMS 2011 - Proceedings: 2011 2nd IEEE International Conference on Emergency Management and Management Sciences*, pp. 866–869. doi:10.1109/ICEMMS.2011.6015820.

Zhang, Y. and Yuen, K.V. (2022) ‘Applications of Deep Learning in Intelligent Construction’, *Structural Integrity*, 21, pp. 227–245. doi:10.1007/978-3-030-81716-9_11.

Zhang, Z. *et al.* (2019) ‘Whole building energy model for HVAC optimal control: A practical framework based on deep reinforcement learning’, *Energy and Buildings*, 199, pp. 472–490. doi:10.1016/J.ENBUILD.2019.07.029.

Zhou, P. *et al.* (2024) ‘Explainable feature selection and ensemble classification via feature polarity’, *Information Sciences*, 676, p. 120818. doi:10.1016/J.INS.2024.120818.

Zhou, P. and Chang, Y. (2021) ‘Automated classification of building structures for urban built environment identification using machine learning', *Journal of Building Engineering*, 43, p. 103008. doi:10.1016/J.JOBE.2021.103008.

Zou, Z. and Ergan, S. (2019) 'Leveraging Data Driven Approaches to Quantify the Impact of Construction Projects on Urban Quality of Life', *arXiv preprint arXiv:1901.09084*, pp. 1–31. Available at: https://arxiv.org/abs/1901.09084 (Accessed: 8 May 2020).