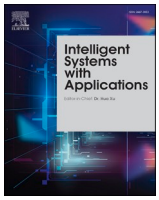
Intelligent Systems with Applications 24 (2024) 200433

Contents lists available at ScienceDirect

Intelligent Systems with Applications

journal homepage: www.journals.elsevier.com/intelligent-systems-with-applications



Deep learning for automated encrustation detection in sewer inspection

Wasiu Yusuf [a], Hafiz Alaka [a,*], Mubashir Ahmad [a], Wusu Godoyon [a], Saheed Ajayi [b], Luqman Olalekan Toriola-Coker [c], Abdullahi Ahmed [d]

[a] Big Data Technologies and Innovation Laboratory, University of Hertfordshire, Hatfield, AL10 9AB, United Kingdom
[b] School of Built Environment, Engineering and Computing, Leeds Beckett University, Leeds, LS2 8AG, United Kingdom
[c] School of Built Environment, Engineering and Computing, University of Salford, Manchester, M5 4WT, United Kingdom
[d] School of Energy, Construction, and Environment, Coventry University, Coventry, CV1 5FB, United Kingdom

ARTICLE INFO

Keywords:
Deep-learning
Encrustation
CNN
Sewer systems
CCTV

ABSTRACT

Rapid urbanization and population growth in recent decades have placed significant pressure on urban cities to rely heavily on underground infrastructure, such as sewers and tunnels, to maintain the provision of essential services. These sewers, typically having a limited lifespan of 50 to 100 years, are prone to various forms of defects. While prior research has primarily addressed common sewer defect like crack, root intrusion, and infiltration among others, the challenge of encrustation—the formation of hard deposits within sewer systems—has received less attention. This study presents a pioneering deep-learning approach to detect encrustation in sewers by leveraging survey videos from 14 different sewers in the United Kingdom. Our work marks the first effort to develop models specifically for detecting encrustation using deep learning techniques, as previous studies have focused on other types of deposits such as settled and attached deposits. By converting the videos into sequential image frames, we subjected them to thorough analysis and several image pre-processing techniques. Our contributions include the development and comparison of different classification models using backbone CNN networks such as AlexNet, VGG16, EfficientNet, and VGG19 to classify encrustation. Notably, this study provides the first metric-based comparison of these backbone networks to identify the most effective model for encrustation detection. The results demonstrate an impressive 96 % accuracy using the deep architecture of VGG19. Beyond accuracy, this research explores the impact of data augmentation and network dropout on reducing overfitting and enhancing model performance. Additionally, we analyze the time complexities associated with training models with and without data augmentation, providing valuable insights into the efficiency of our approach.

1. Introduction

Sewers, one of the most important underground infrastructures (Wang et al., 2021a), have seen more frequent usage in the United Kingdom (UK) due to the increase in population and urbanization (Statista, 2021), with approximately 8 % increase in population in the last 2 decades (Office for National Statistics, 2021). This has led to more subscriptions to the current sewers and put more pressure on the current sewerage system, without any substantial additional sewers in recent times. Covering 347,000 km in distance, sewers in the UK convey 11 billion litres of wastewater daily (DEFRA, 2014) and about 39 million tonnes of sewage into the Thames annually (Ofwat, 2023; Varghese, 2023). Failure to properly manage these sewers can have devastating environmental, social, and financial impacts on society and water companies, such as in-house and environmental flooding, episodes of land, water, and air pollution, reduction in the lifespan of household drainage systems, huge penalty fines on sewer management companies, and increased maintenance cost among others. Avoiding these impacts prevents waterborne illnesses, provides safe drinking water, and enhances general sanitation (Decor et al., 2019; Moskalenko et al., 2020).

As a durable structure, sewers have a lifespan between 50 to 100 years (Wang et al., 2021b), and many of the ones in use today are either old or ageing gradually (Moradi & Zayed, 2017; Myrans et al., 2018a; Wang et al., 2021b). However, the long lifespan catalyses some internal and external factors that expose them to significant defects over this period. These factors include deposits, wear and tear from sewage, high vehicular traffic and train vibrations on roads and lines above the sewer (Aşchilean et al., 2018; Xu et al., 2021), surrounding construction

* Corresponding author.
E-mail address: hafizalaka@outlook.com (H. Alaka).

https://doi.org/10.1016/j.iswa.2024.200433
Received 13 January 2024; Received in revised form 20 July 2024; Accepted 29 August 2024
Available online 28 September 2024

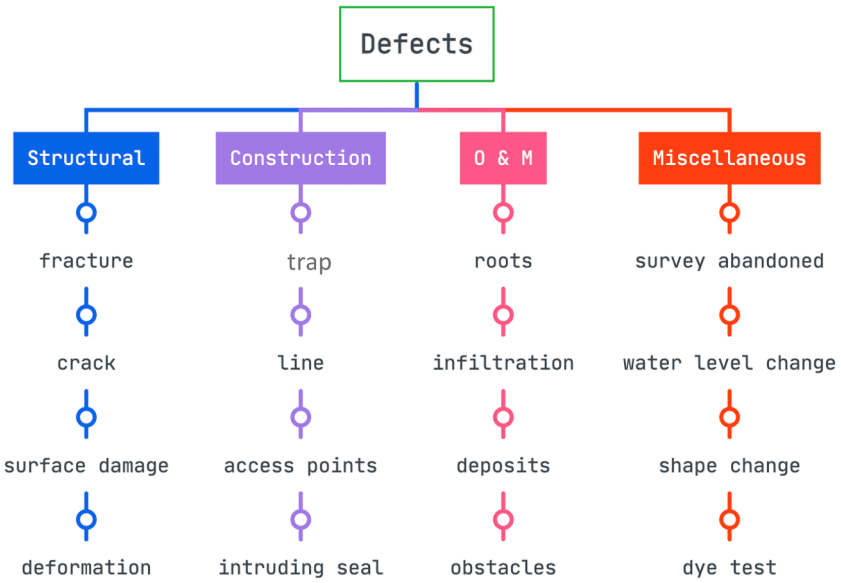


Fig. 1. Defect categorisation (Guo et al., 2009).

activities and poor weather conditions (Hughes et al., 2021) among others. Defects resulting from these vary based on their origin and location within the sewer. According to the Water Research Centre (WRc), these defects are grouped into structural, construction, operation & maintenance, and miscellaneous. WRc provides the coding standard for sewer condition assessment and classification using Closed Circuit Television (CCTV).

Assessment of these defects through a survey is a tedious, time-consuming, and resource-centric task. Traditional defect detection mechanism involves sewer inspectors either going into the sewer to conduct physical and manual inspections or watching many hours-long survey videos to manually spot defects (Pan et al., 2020a; Savino & Tondolo, 2021). This approach is a subjective way of defect detection since it majorly depends on the inspector's judgement (Li et al., 2021; Wang & Cheng, 2019). Additionally, it takes a lot of time, cost, workforce, and other meaningful resources to manually analyse these survey videos. As a result, there is an increased need for automatic defect detection in sewers (Liu et al., 2019) where defects are identified automatically from survey videos or still images. Several techniques have been applied to automatic defect classification using machine learning and computer vision (Ren et al., 2022; Westphal & Seitz, 2021), with the CCTV becoming a major data source and the most popular tool today to carry out a non-invasive inspection and monitoring of operational sewers (Guo et al., 2009; Hawari et al., 2018). In recent years, images or videos obtained from CCTV have been analysed and used in training deep-learning models that can classify or identify defects in sewers (Myrans et al., 2018b; Thiyagarajan et al., 2016).

Most machine learning-based studies in defect classification of sewers have focused on other types of defects such as cracks, tree roots, and blockage (Chen et al., 2018a; Pan et al., 2020b; Kumar et al., 2018), with little or no attention specifically on encrustation, which is common in UK sewers (ClearView Survey, 2023; RapidResponse Drain Care, 2023). This study focuses on the automatic classification of encrustation (i.e., scaled or hard deposits) - an operational and maintenance defect in concrete sewers. Encrustation originates primarily from the accumulation of periodic residuals or deposits from infiltration, seeping, and groundwater leakages. The residuals form a hard coating on the sewer lining and reduce its cross-sectional area. Accumulation of this defect over time can result in reduced hydraulic capacity, structural damage, and partial or complete blockage of the sewer, thereby leading to problems such as environmental flooding, sewer collapse, economic loss, and even a threat to human lives (Xu et al., 2021). Detecting encrustation earlier, just like other defects, provides experts with ample time to plan and strategize on sewer repair or maintenance.

Consequently, this study developed a deep-learning model that can automatically recognize encrustation in sewers due to its importance in sewer maintenance and the lack of studies on this defect. Hence, we employed a deep-learning approach to developing an encrustation classification model for conditional assessment while following a supervised learning approach. We used different CNN models such as AlexNet, VGG16, VGG19, and EfficientNet, however, VGG19 provides the best result. We also employed data augmentation and network dropout to avoid overfitting, and the result shows a substantial performance compared to most existing classification models in sewer defect detection.

2. Related work

The maintenance and assessment of urban underground infrastructure, particularly concrete sewers, is of vital importance to ensure the continuous flow of wastewater and prevent structural degradation(Hu et al., 2019; White et al., 2013). Among the numerous challenges faced in sewer maintenance, the identification of defects, such as encrustation and others, plays a crucial role (Fang et al., 2020; Li et al., 2023). Several studies have been conducted in the classification of sewer defects using CCTV images or other data acquisition technologies. These defects are in various forms and shapes, and Fig. 1 pictorially shows defect categorisation adapted according to Water Research Centre (WRc).

WRc categorisation is briefly summarised below.

- **Structural:** Defects that affect the structure of the sewer pipes.
- **Operation and maintenance (O&M)** are internal and external objects found in pipes that may obstruct the sewage system's regular operation.
- **Construction:** defects and conditions associated with the methodology employed in constructing and connecting pipes.
- **Miscellaneous:** Any defect or features that are not part of the other categories or general items concerning the sewer.

Consequently, issues such as encrustation are categorized as O&M since it's a type of deposit, and evaluating such defect category through traditional methods proves to be a laborious, time-intensive, and resource-demanding endeavour. Therefore, numerous investigations have sought to address these challenges by applying various

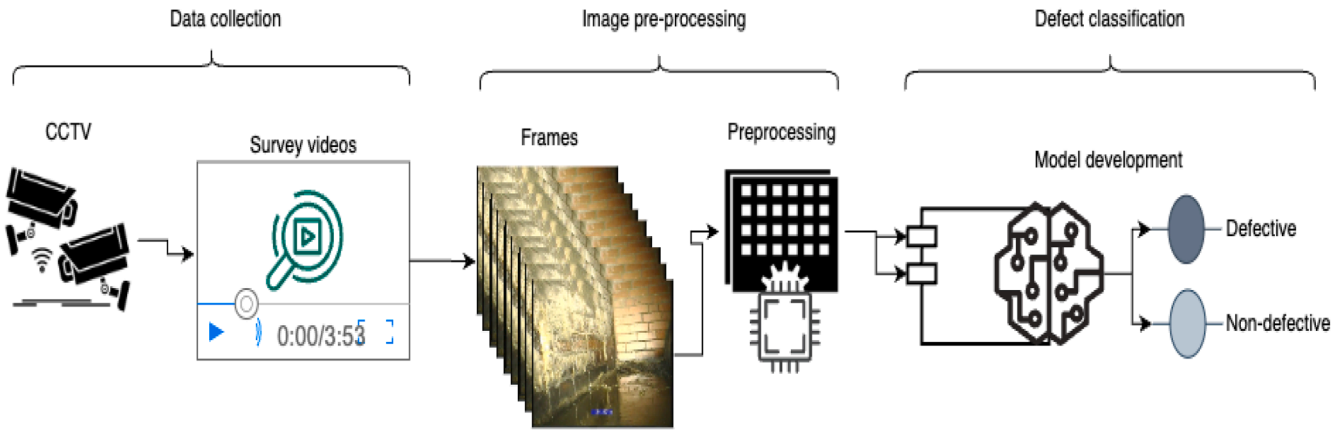


Fig. 2. Architecture of Proposed Deep-learning Approach.

methodologies, utilizing different defects as case studies, with little mention of encrustation. Recently, machine learning (Myrans et al., 2018b) has been leading the revolution in the field of computer vision (Smith et al., 2021; Michael, 2017; Sutton, 2013), especially its application in defect detection using variations of deep-learning algorithms like CNN (Gao et al., 2019; Gutierrez-Mondragon et al., 2020) which is often used in analysing survey videos and images. Before that, conventional Machine Learning (ML) techniques used in linear regression and algorithms such as Support Vector Machine (SVM) and Random Forest (RF) have been employed in defect classification problems (Shang et al., 2019) and productivity analysis. While they had commendable success in defect classification and can adapt to new types of defects, they cannot be used as a standalone method for complete end-to-end defect classification (Li et al., 2022; Xu et al., 2020) as they need other techniques for feature extraction or selection as a prerequisite for subsequent defect classification.

Furthermore, Decor et al. (2019) established that network learning-based approaches in defect classification and detection are now more competitive. They evaluated and compared both RF two CNNs on tunnel linings using images of concrete and masonry walls of tunnels. Their study shows that deep-learning models outperform the RF model on concrete and masonry walls. One of many notable advantages of deep-learning algorithms over other conventional ML algorithms is the automatic feature extraction (Fatma et al., 2016; Lin et al., 2017) in deep-learning without explicitly applying feature extraction techniques (Liu et al., 2019). This has been a source of motivation for people who chose deep learning over other approaches, as feature extraction is usually time-consuming with conventional ML algorithms. This efficiency of deep learning across many of its applications, such as autonomous vehicles (Grigorescu et al., 2020; Ni et al., 2020; Yue et al., 2023), construction (Mostafa & Hegazy, 2021; Yu et al., 2022), healthcare (Jiang et al., 2017), and manufacturing (Bartsch et al., 2021), coupled with the rapid growth of generated data, has scaled up its adoption.

Previous deep-learning approaches focused on defects such as infiltration, root intrusion, and sewage deposits (Pan et al., 2020a), with no specific attention on encrustation. Kumar et al. (2018) proposed defect detection and tracking in sewer pipes using deep learning and metric learning to identify and track defective objects across consecutive frames in the video, but their study only focused on defects such as fractures, root intrusion, and lateral connection. Li et al. (2021) used deep learning with local and global feature fusion to detect the defect in sewer pipes by concatenating the proposed region feature and the global contextual feature from the corresponding image to enhance feature representation fine-grained defect classification network. However, the scope of their model only focused on defects such as Barriers, Deposition, Distortion, Fraction and Foreign bodies. Hence, our proposed model aims to detect encrustation in sewers using deep learning and datasets from CCTV surveys.

More than conventional machine-learning models, deep-learning models require a massive amount of data to train and generalise effectively, but this comes with computational overhead (Fu & Menzies, 2017; Thompson et al., 2020). For example, Xu et al. (2021) combine automatic defect detection and segmentation by applying modified Mask R-CNN to segment defects in tunnels to reveal more detailed information about each defect. Even though their result averaged 90 % generalisation, they noted the computational cost of using their approach as it requires more hours to train the model. Resource usage in deep-learning models has sparked a variety of conversations around the trade-off between model performance and computational cost. Models with large training data often perform better than those with lesser data because the models have more features to learn from, but such models are usually not repeatable in research studies due to hardware requirements and computational costs incurred during training. Invariably, considerations are given to computational cost while choosing the size of training data for deep-learning models. Outside the realm of machine learning, other approaches (Lepot et al., 2017; Redmon et al., 2015; Situ et al., 2023a) have also been applied to defect detection, but this is outside the scope of this study.

3. Methodology

To detect surface encrustation in concrete sewers, we proposed a deep-learning model that automatically detects encrustation from sewer images. To achieve this, we sequenced inspection videos from CCTV footage into streams of images. Sample video sequencing on selected sewers across the UK is summarised in Table 2. The architecture of this approach is pictorially depicted in Fig. 2, which is grouped into data collection, image pre-processing, and defect classification.

3.1. Data collection

We collected several survey videos from 14 different sewers and culverts across the United Kingdom from a UK sewer inspection company that offers a wide range of industry-based services, including CCTV Surveying, Gutter Maintenance, Grease Traps and many more. This survey covers sewers in popular cities like London, Manchester, Glasgow, Nottingham, Leeds, and Derby. In most cases, during a survey, the total sewer length is divided into different sections (from one manhole to another designated manhole), and an additional survey is conducted for each section. Hence, multiple videos are recorded for a sewer based on each surveyed section. The video data was collected as MPEG files with an approximate size of 37GB.

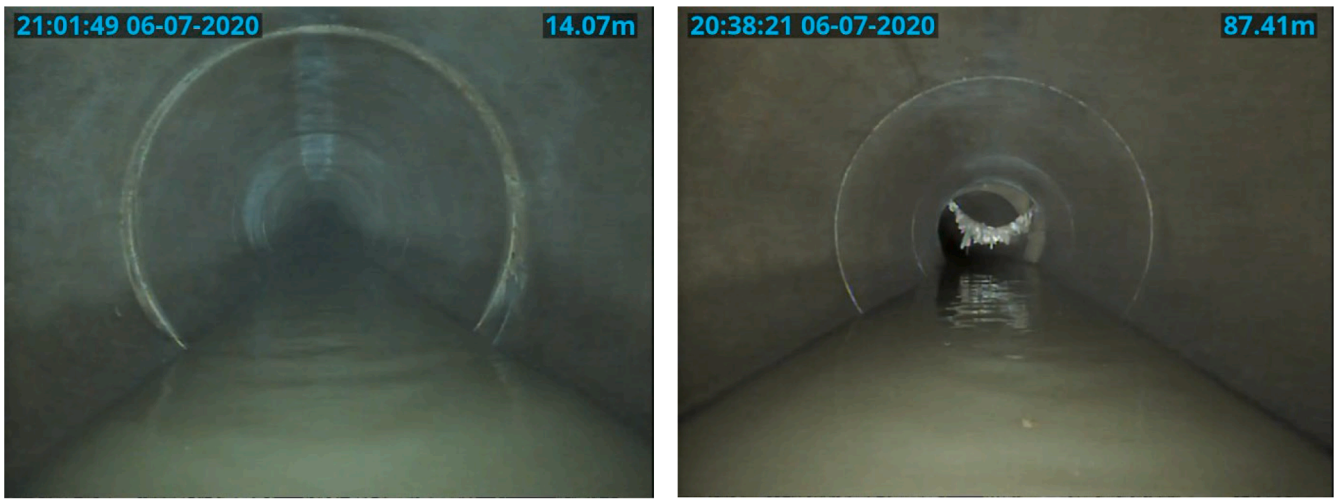


Fig. 3. Sample image extraction.

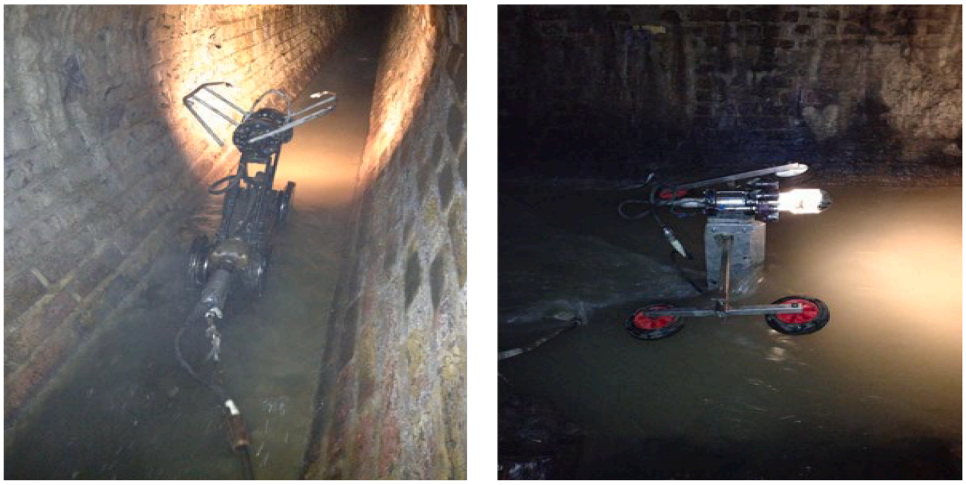


Fig. 4. Robotic system for sewer survey.

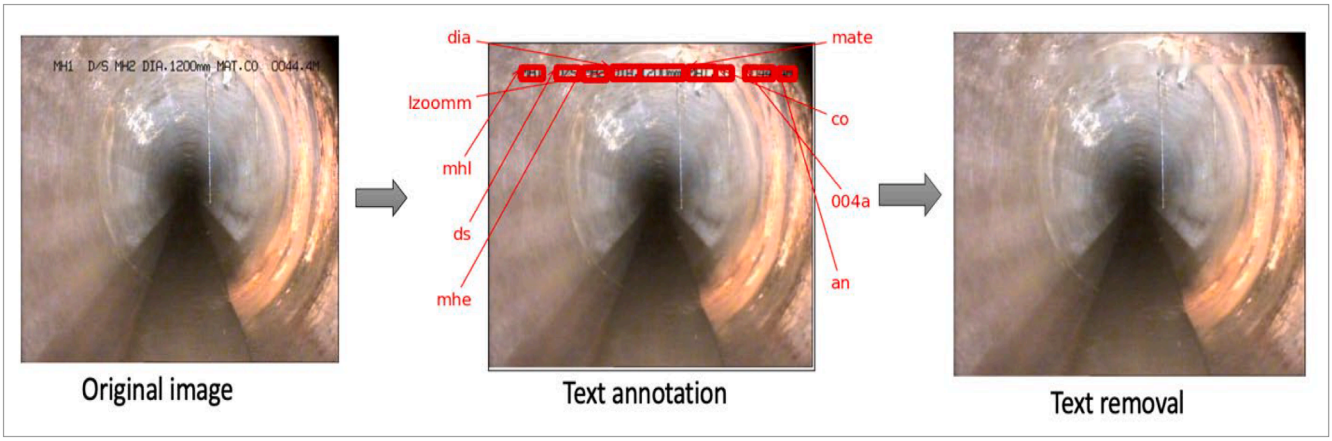


Fig. 5. Text removal from images.

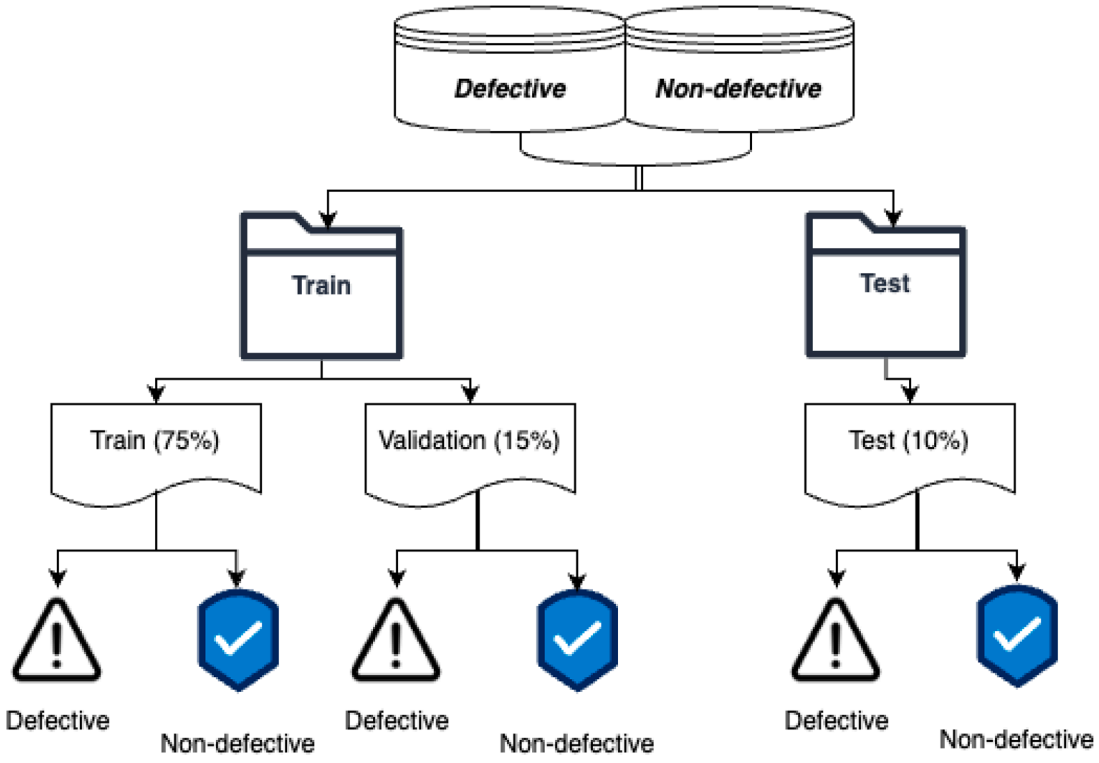


Fig. 6. Annotated dataset for model development.

3.2. *Image extraction*

Extraction of images from videos (i.e., video sequencing) entails converting a video into frames of images. We used *OpenCV*, a Python library for analysing and manipulating videos and images, to extract images from the inspection videos. For example, Fig. 3 shows sample images extracted at a different inspection point in Sewer 1.

We sequenced the videos at a frame rate of approximately 25 images per second using OpenCV's default FPS (frame per second) for recorded videos, which resulted in over 1 million images of sewers surface across different locations within the UK. We further selected images with encrustation since this is the defect in focus.

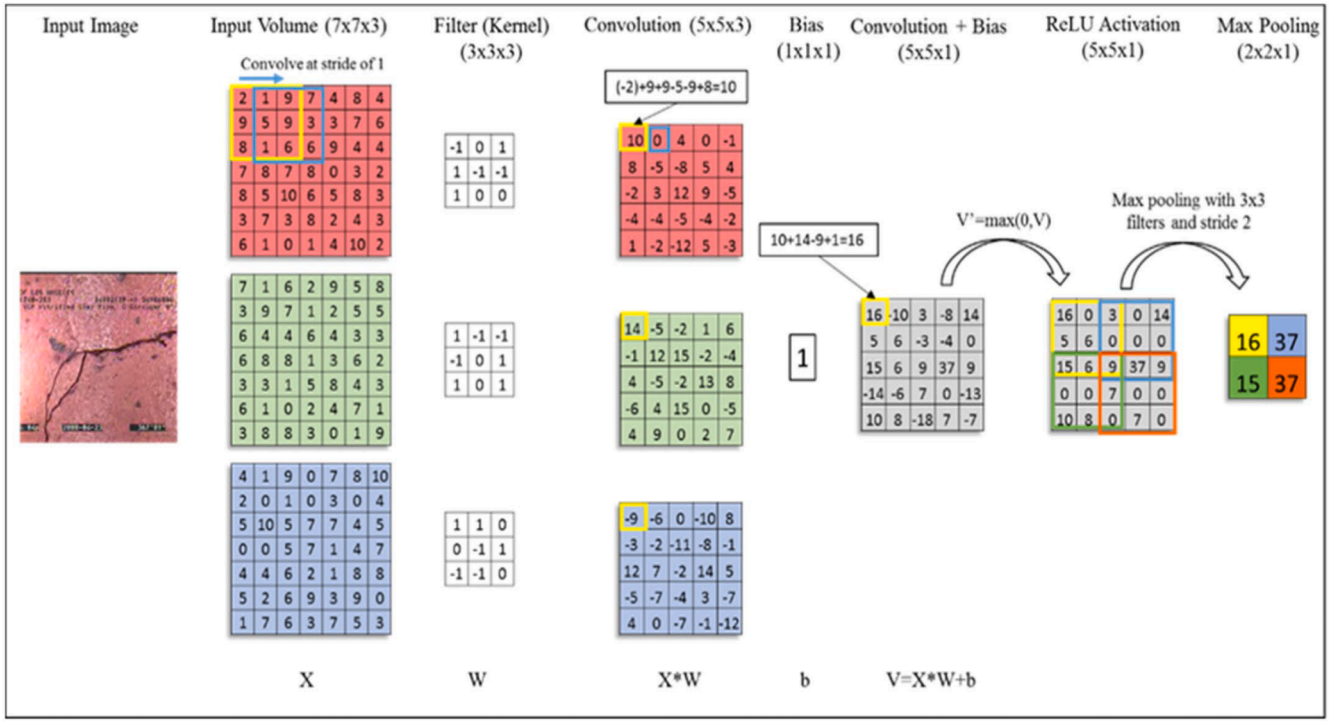


Fig. 7. Convolution layer in CNN (Cheng & Wang, 2018).

3.3. Image pre-processing

Extracted images were in different dimensions (e.g., 750 × 970 pixels) and sizes (e.g., 750 kb), which will negatively affect the model and the pixel-wise spatial information. Even though the robotic system on which the CCTV was mounted has its lightning functionality as seen in Fig. 4, some images were not bright enough to clearly show the surface of the sewer because of light and CCTV alignment distortion whenever the robot hits an obstacle. Owing to this, we applied brightness in images by manipulating the image's colour space in HSV (Hue, Saturation, and Value) using *OpenCV*. HSV saves images' colour information in a rounded representation by depicting the colours as the human eye perceives them. Increasing some dark images' brightness is necessary because darker images reduce model generalisation (Xu et al., 2021).

In addition, all images were resized to a fixed 150 × 150 pixel resolution by specifying the target-size value while building data generators and fetching images from directories using TensorFlow. This abridged image dimension reduces compute resource usage and maintains both the image spatial information and correlation. Similarly, we remove text with survey meta-data information from the images using a fusion of OCR and image inpainting from *OpenCV* and *Keras OCR*. As represented in Fig. 5, the text removal process includes text identification using annotation, removal, and re-painting area where the text was removed.

3.4. Defect classification

To automatically classify encrustation in concrete sewers, a deep-learning model was developed using a 3-fold approach; dataset annotation, model architecture construction, and model training & evaluation.

3.4.1. Dataset annotation

The dataset prepared for this automation was annotated by directory grouping. First, annotation was done by dividing the dataset into training and testing directories. Then, each directory was further divided into defective and non-defective, allowing the model to identify the training and testing set directories. Fig. 6 depicts a hierarchical view of the directory-based annotation. The final dataset was divided between training and testing using stratified hold-out sampling; 75 % for training, 15 % for validation, and the remaining 10 % for testing (Lyu et al., 2021; Xu & Goodacre, 2018). We adopted this splitting ratio due to our small data size and the need to use a sizeable amount to progressively validate the model. For the training portion, the sample size will be automatically augmented (i.e., increased) (Agnieszka & Michał, 2018) using different approaches, however, we can't augment the validation set.

3.3.2. Model architecture

Architecture for the deep-learning model revolves around piecing together algorithms, different parameters, and constraints to be used by the CNN algorithm for both training and testing of the model. This study divided the model architecture into four dependent stages: convolution, activation, pooling, and Flattening.

At the convolution layer, a *filter* (also called the kernel, weight matrix, or receptive field) slides through the pixels of the input image from left to right in a bid to extract relevant features from the image by conducting an element-wise dot operation of the filter weights against each subarray of the image pixel. As shown in Fig. 7, this dot operation

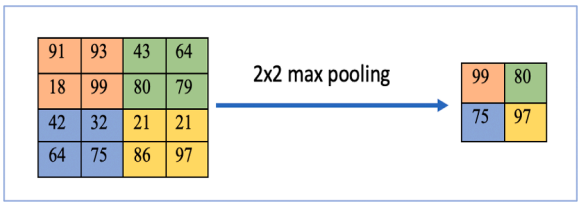


Fig. 8. Max pooling of feature maps.

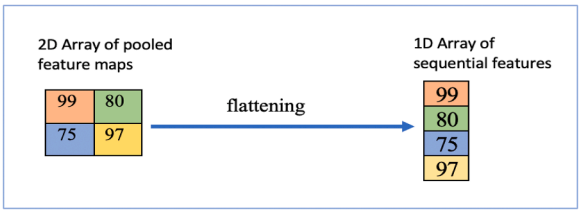


Fig. 9. Flattening pooled feature maps.

results in a stack of *feature maps* $(x * w)$ - a matrix-like representation of the relevant features extracted from the image.

The weight of the filter is randomly generated at the initial stage, and we set a *stride* (i.e., the vertical movement of the filter across the pixels) of 1 to aid in more feature extraction. Filter dimension, weight, bias, and stride are tuned parameters to get more accurate detection. Fig. 6 shows the processes within a typical CNN model during defect detection. Since images are non-linear data, a common way to add non-linearity (i.e., adding more correlations between pixels) is the activation function on feature maps. In this study, we used the Rectified Linear Unit (ReLU) to add non-linearity within the network. We also used ReLU to remove negative values from an activation map by setting the value to zero. Other common activation function includes *sigmoid* and *tanh* (Tan et al., 2021). The choice of ReLU is due to its low time complexity in training the network without serious consequence to generalisation accuracy (Javid et al., 2021).

Furthermore, we introduced pooling between subsequent convolution layers to progressively reduce the spatial size of the image. Pooling is a form of dimensionality reduction that minimises the feature sets and maintains the spatial relationship between pixels. Hence, we used a 2 × 2 filter and stride of two to max pool the spatial size of the input feature map at each convolution layer. Max pooling is common in image recognition and object detection for pooling the maximum value from the pixel subarray in the feature maps (Ajit et al., 2020; Mushtaq & Su, 2020). Other pooling techniques include *average pooling* and *L2 norm*. Fig. 8 illustrates how max pooling reduces the spatial size of our feature maps.

To calculate the output dimension of the feature map after pooling, we used Eq. 1 (Adrian Rosebrock, 2021).

$$(W - F + 2P / S) + 1 \tag{1}$$

Where *W, F, P*, and *S* are input size volume, filter size, number of zero padding, and stride, respectively. Pooled feature maps are usually in a multi-dimensional array of features, as seen in Fig. 8. This must be transformed into a form that is executable by the fully connected layer. We employed flattening in converting this array into a sequential 1D array (i.e., a long vector) to prepare these features for further processing, as seen in Fig. 9.

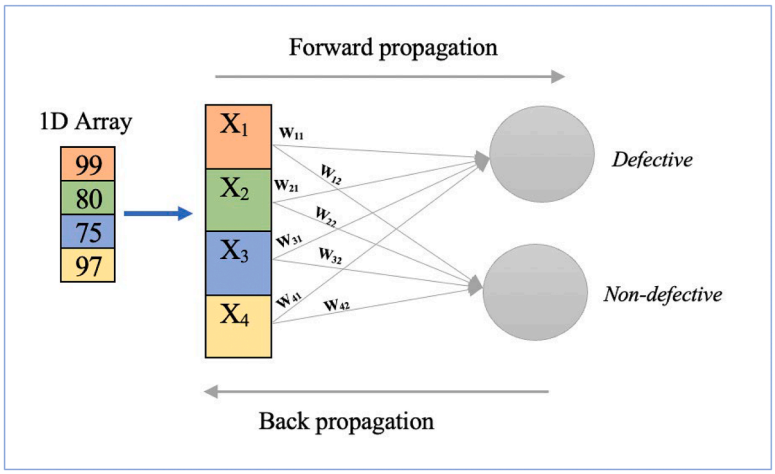


Fig. 10. Fully connected layer.

Table 1
Sample survey videos.

	Surveyed Sewer	No of Videos	Location	Videos	Duration	Size
1	Sewer 1	1	Nottingham	Video 1	28:55 mins	41.3MB
2	Sewer 2	3	Glasgow	Video 1	7:01 mins	175 MB
				Video 2	6:22 mins	158 MB
				Video 3	4:09 mins	103 MB
3	Sewer 3	2	Manchester	Video 1	38:50 mins	918 MB
				Video 2	23:00 mins	402 MB
4	Sewer 4	8	Yorkshire	Video 1	02:09 mins	54.4 MB
				Video 2	13:46 mins	338 MB
				Video 3	16:57 mins	418 MB
				Video 4	33:31 mins	824 MB
				Video 5	8:04 mins	198 MB
				Video 6	02:40 mins	67.1 MB
				Video 7	05:42 mins	141 MB
				Video 8	12:13 mins	301 MB
5	Sewer 5	7	London	Video 1	12:26 mins	294 MB
				Video 2	05:17 mins	125 MB
				Video 3	41:53 mins	990 MB
				Video 4	15:53 mins	376 MB
				Video 5	09:36 mins	227 MB
				Video 6	09:21 mins	221 MB
				Video 7	02:46 mins	65.7 MB

The fully connected layer is a typical Multilayer Perceptron (MLP) sitting at the edge of the architecture. Individual values in the flattened array (1D array) are usually treated as separate features that denote the images by combining the features into more attributes. This attribute is used to predict the class of an image (i.e., defective or non-defective) by using a classifier at the output layer. Also, the error rate is calculated at this layer and sent backward (backpropagation) to the network to adjust the weight and the feature detectors to improve and optimise the model's performance, as depicted in Fig. 10.

This backpropagation is conducted several times to predict the class of an image using the features, and a linear and non-linear transformation is carried out. The linear transformation is of the linear form $(z = \ w^T \cdot X + \ b)$, where w, x, and b are weight, input, and bias respectively.

Since linear transformation alone cannot capture the complex relationship between these parameters, we introduced a sigmoid activation function to add non-linearity to the data and to predict the probability of whether the input image is defective or not. This probability ranges between 0 and 1. A prediction X, such that $X \geq \ 0.5$ indicate a defective image while $X < 0.5$ indicate otherwise. The sigmoid function for calculating this is given as represented in Eq. 2.

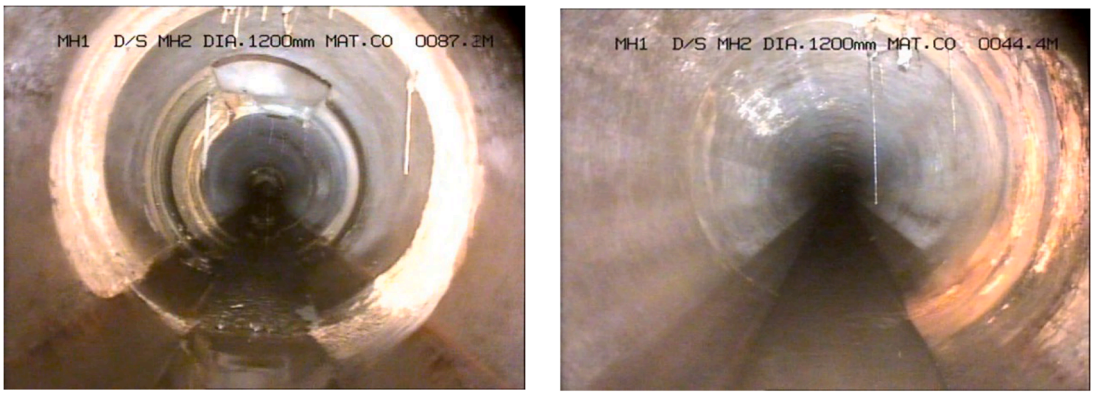
$$f(x) = \frac{1}{(1 + e^{\wedge - X})} \tag{2}$$

4. Experiment and result

The proposed deep-learning classification approach was implemented with TensorFlow, a deep-learning tool for computer vision and deep learning, while the baseline algorithm used for feature extraction was VGG19 (Dang et al., 2021a), due to not just its accuracy in defect detection and classification in sewers (Pan et al., 2020b; Shang et al., 2019; Situ et al., 2023b), but also shows improved performance with reduced parameters and computational time in other tasks (Ahmed

Table 2
Image extraction summary.

	Infrastructure	Duration (mins)	No of Images
1	Sewer 1	28:55	43,371
2	Sewer 2	17:33	26,275
3	Sewer 3	61:50	92,750
4	Sewer 4	95:03	141,050
5	Sewer 5	97:26	145,800
	Total	**301:18**	**303,446**



(a) Encrustation between 7 and 5'o clock (b) Encrustation between 12 and 5'o clock

Fig. 11. Sample defective images.

Table 3
Model summary.

Layer (Type)	Output Shape	Param#
input_1 (InputLayer)	(None, 150, 150, 3)	0
conv2d (Conv2D)	(None, 148, 148, 16)	448
max_pooling2d (MaxPooling2D)	(None, 74, 74, 16)	0
conv2d_1 (Conv2D)	(None, 72, 72, 32)	4640
max_pooling2d_1 (MaxPooling2)	(None, 36, 36, 32)	0
conv2d_2 (Conv2D)	(None, 34, 34, 64)	18,496
max_pooling2d_2 (MaxPooling2)	(None, 17, 17,64)	0
flatten (Flatten)	(None, 18,496)	0
dense (Dense)	(None, 512)	9,470,464
dense_1 (Dense)	(None, 1)	513
Total params:	**9494,561**	
Trainable params:	**9494,561**	
Non-trainable params:	**0**	

Table 4
Model performance.

Epoch	Time (s)	Tr acc	Tr loss	Val acc	Val loss	Avg. batch accuracy
1	95	0.8319	0.4244	0.8992	0.2047	0.66736
2	98	0.9585	0.1308	0.9376	0.1649	
3	100	0.7102	4.4055	0.4990	8.0210	
4	101	0.4962	7.7705	0.5010	7.9901	
5	99	0.4927	7.8237	0.5000	8.0055	
6	101	0.5068	7.6069	0.5146	7.7896	0.50104
7	101	0.5018	7.6845	0.4979	7.4503	
8	101	0.4922	7.8326	0.4938	7.5119	
9	101	0.4980	7.7449	0.4979	7.7433	
10	106	0.5028	7.6678	0.5010	7.6970	
11	101	0.5013	7.6942	0.5031	7.6662	0.50248
12	105	0.5063	7.6171	0.5062	7.3269	
13	211	0.4930	7.8221	0.5000	8.0055	
14	100	0.4962	7.7709	0.5073	7.6045	
15	**101**	**0.4972**	**7.7559**	**0.4958**	**7.7742**	

et al., 2020; Ayanwola et al., 2023; Bouhsissin et al., 2021; Li et al., 2022). It also outperformed others when compared with other models such as VGG16, AlexNet, and EfficientNet (Ahmed et al., 2020; Jain et al., 2023). The next subsections include data preparation, training, and result analysis.

4.1. Dataset preparation

Table 1 summarised 5 out of the 14 surveyed sewer videos, showing the location and duration of the survey.

Subsequently, image extraction was carried out on each of the videos and Table 2 summarised the number of images extracted from survey videos in five sample sewers.

All extracted images were critically analysed and manually grouped into two categories: *defective* and *non-defective*. Defective images indicate only signs of encrustation in some parts of the sewer, while non-defective images are all other images not showing any sign of encrustation but might include other types of defects which are not of focus. We only recorded 1050 defective images compared to more than a million either non-defective or had other types of defects. This poses a widespread problem in machine learning known as *class imbalance* (Li et al., 2021), where data (images in this case) in a particular category (non-defective images) are more than the ones in another category (defective).

Invariably, this means the model will perform well just by classifying all data to belong to the class with the higher amount of data, giving a false performance capability of the model. To mitigate this, we have manually selected 1050 non-defective images to equate the number of defective images, and these cumulatively resulted in 2100 images (defective and non-defective) across 14 different sewers. Sample defective images are shown in Fig. 11.

4.2. Model training

Based on the methodology and architecture described above, we trained a VGG19 by specifying 3 layers of convolution and extracting 16, 32, and 64 filters respectively, with a 3×3 window. We followed each convolution layer with a ReLU and max pooling layer using a 2×2 window to reduce the pixel spatial information. The convolution, ReLU, and max pooling layers were stacked into three separate convolutional modules. This is done to keep the model size minimal and reduce overfitting chances since we have a small training sample. Another attempt was made to further reduce overfitting by augmenting the training data, as explained above. Table 3 depicts the summary of the model.

The output shape illustrates how the spatial size of feature maps changes in each succeeding layer. Convolution layers decrease the volume of the feature maps due to zero padding, and each pooling layer halves the feature map.

We defined two data generators for training and testing, which read images from the - directory, convert them to float32 tensors, and feed them (with their labels) to our network. We normalised the pixel of each image to a range [0,1] which was in the range [0, 255] because of its RGB properties. We fit the model using 15 epochs, 20 *batch_size* and 100 *steps_per_epoch* on 420 validation images, with 210 defective and non-defective images, respectively. We trained the model with binary_crossentropy loss since it's a binary classification problem. Root Mean Square Propagation (RMSprop) optimiser was embedded in the configuration using a learning rate of 0.001. *RMSprop* was preferred to

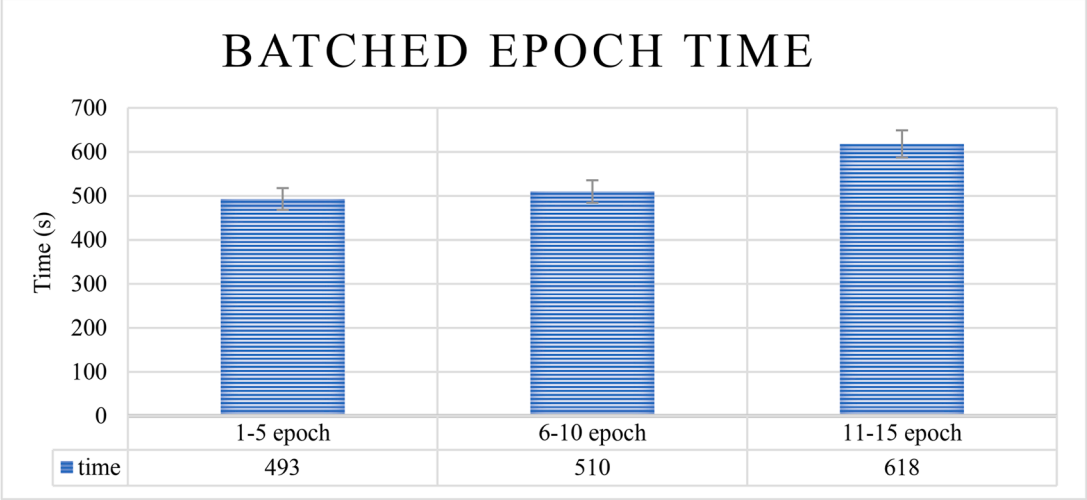


Fig. 12. Completion time for batches of epoch.

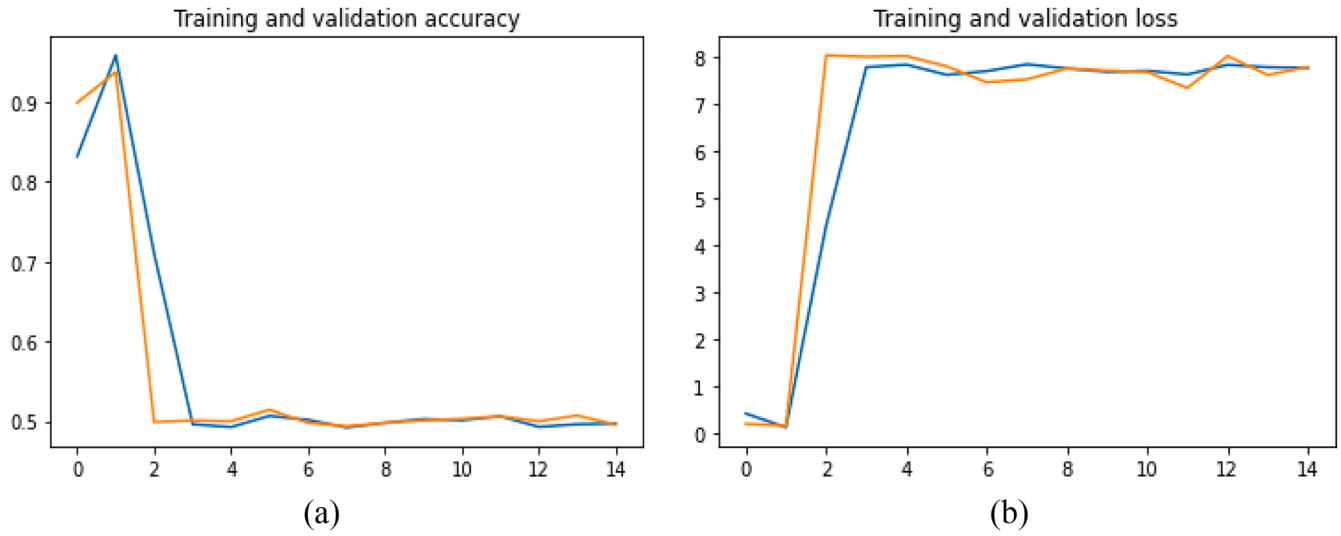


Fig. 13. Training and validation result.

Stochastic Gradient Descent (SDG) because it automates the learning rate during training. Other optimisers that can also serve this purpose include *Adam* and *AdaGrad* (Do et al., 2020; Ottoni et al., 2023).

4.3. Result and discussion

The performance of the model was first monitored and evaluated using accuracy. The first model was developed using the data stated in Section 3.4. It completed 15 epochs in approximately 22 min, as shown in Fig. 11, with the detailed accuracy summarised in Table 4 and Fig. 12. The first epoch completed execution with the least time (95 s) with an accuracy of 89 %. The second epoch shows an improvement with 93 % accuracy in 98 seconds.

However, we recorded a drastic decrease in the model's accuracy from the fourth epoch as the accuracy dropped to 49 % and progressively averaged 50% till the last epoch, as further shown in the table above.

This means the model cannot differentiate between a defective and non-defective image at this point. Pictorially, the accuracy and loss of the model are given below in Fig. 13.

After investigation, we noticed the model was overfitting after the second epoch because there weren't enough unique features to learn from the small dataset to aid its training and generalisation (Aggarwal, 2019; Agnieszka & Michał, 2018). Hence, the model resulted in using irrelevant features to learn and generalise. This is also evident in how the loss (i.e., error rate) value increased after the second epoch, as depicted in Figs. 14 and 15 below shows the wide margin between accuracy and loss for both training and validation. This further indicates the low learning rate of the model.

4.3.1. Data Augmentation

Eliminating overfitting becomes inevitable to achieve better accuracy due to the small size of the selected images. Data augmentation (i.e., generating new images from a sample image) was carried out to increase the number of images, which aids in extensive model training and more generalisation capability (Fabian et al., 2021). At first, the training data was augmented to uniquely increase the size of our training set, which directly increased the number of features the model can learn from to make a better generalisation. At this level, augmentation was implemented in arbitrary rotation, zooming, padding, and flipping of images by setting the new images to a 40° rotation and 20% range in width, height, and zoom. A sample data augmentation using a defective image is shown in Fig. 16(a) – (f).

Augmenting the training data can help reduce the overfitting of the

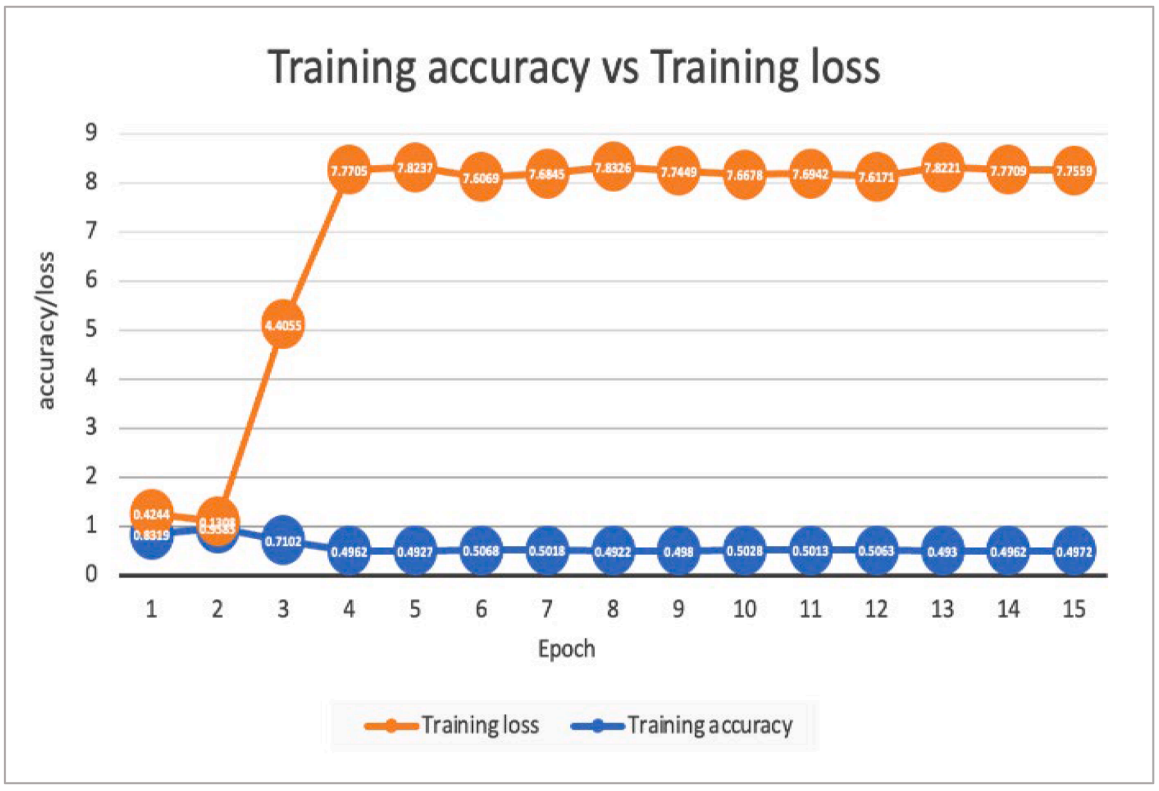


Fig. 14. Training vs Validation accuracy.

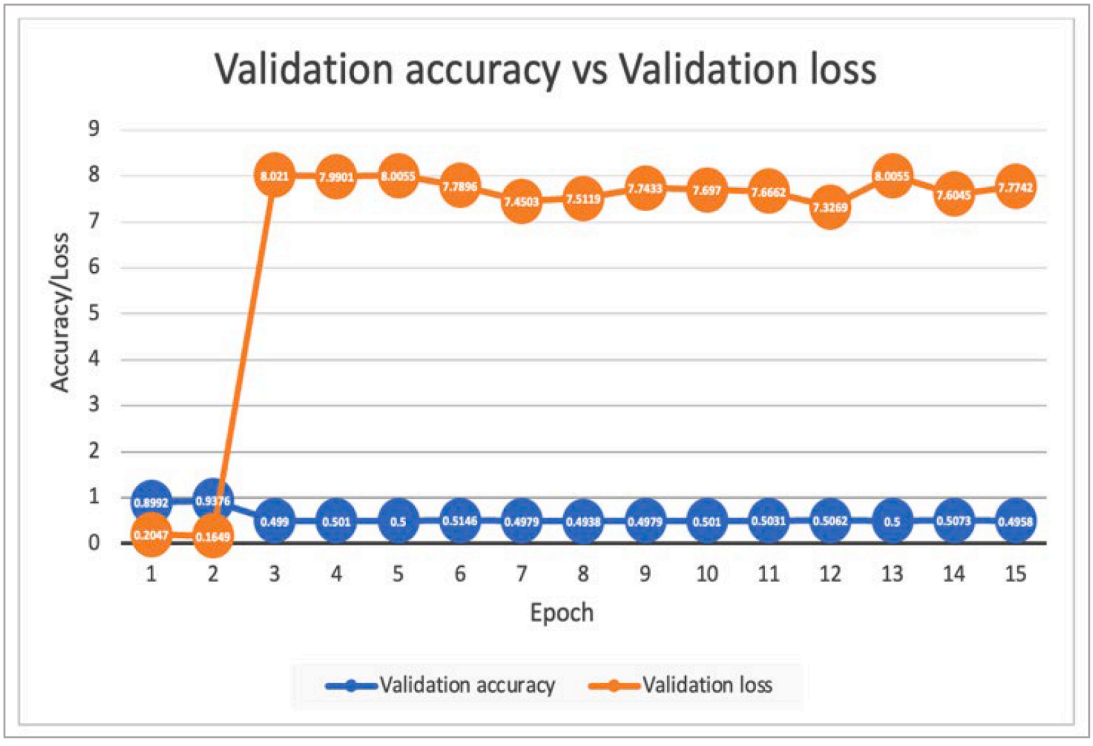


Fig. 15. Training vs Validation loss.

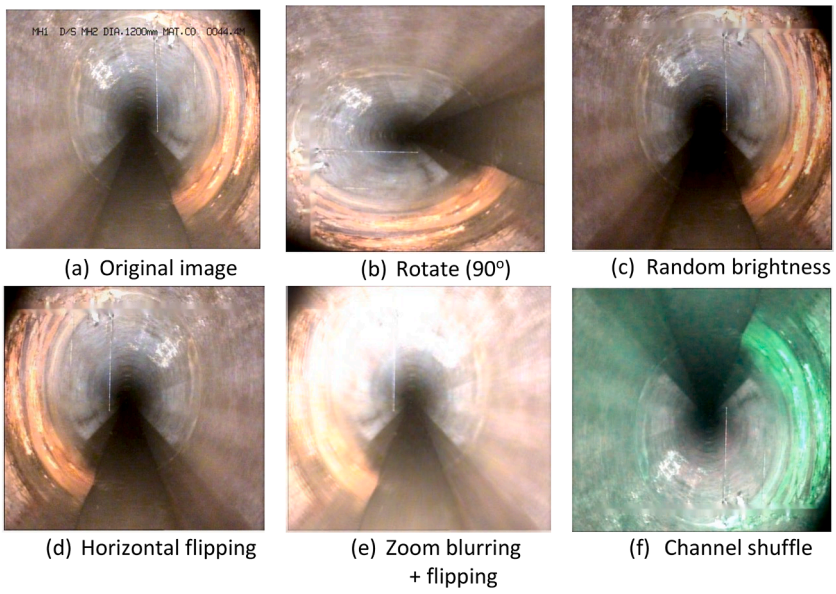


Fig. 16. Data augmentation using image with encrustation.

Table 5
Model performance with augmentation and dropout.

Epoch	Time (s)	Tr accuracy	Tr loss	Val accuracy	Val loss
1	97	0.9632	0.1283	0.9522	0.1348
2	96	0.9631	0.1324	0.9522	0.1386
3	95	0.9632	0.1372	0.9329	0.1721
4	95	0.9587	0.1445	0.9522	0.1575
5	95	0.9748	0.078	0.9587	0.1086
6	98	0.9691	0.1215	0.9502	0.1161
7	97	0.9689	0.142	0.9644	0.1454
8	**97**	**0.9717**	**0.1051**	**0.9665**	**0.1093**
9	102	0.9676	0.1309	0.9685	0.0881
10	95	0.9719	0.0886	0.9502	0.1121
11	93	0.9637	0.149	0.9563	0.1632
12	92	0.9706	0.1034	0.9563	0.1118
13	92	0.9652	0.1273	0.9462	0.1455
14	92	0.9728	0.0954	0.9502	0.1087
15	93	0.9737	0.1449	0.9441	0.164

model by presenting the augmented images that do not correlate with previous ones to the model as new and unique images at every iteration. Nevertheless, these augmented data can still be somehow correlated since each batch stemmed from a single image, as seen in Fig. 16. Hence, we added a network dropout (Srivastava et al., 2014) of 0.5 after every convolution to remove pixel correlation among data input. Then, the model was retrained as previously described for 15 epochs, and the addition of a dropout layer before flattening was conducted on the pooled feature maps and passed to the fully connected layer. This resulted in evident improvement in generalisation and a decrease in error rate. Table 5 shows how the model performed in 15 epochs after applying *augmentation* and *dropout*.

The model improved its generalisation from an average of 50 % to 97 % within 15 epochs. This is because of the increment of training data from 2100 (defective and non-defective) images to 50,400 using data augmentation and embedding network dropout in between convolutions. As a result, the model had more features to learn from the additional images. Hence, the reason for the improvement of the model. Fig. 17 shows this improved model's accuracy and error rate.

The training accuracy of the model increased from 95 % to 97 % (as represented in Figs. 15 and 16) throughout 15 epochs, indicating that the model is learning and capturing the underlying patterns within the training data. This is indicative of the model's ability to adapt and fit the training data, which is essential for the convergence of the network. Validation accuracy, measured on an independent dataset not used for training, also mirrored the trends observed in the training accuracy. The model originally displayed a high validation accuracy of around 93% and, similar to training accuracy, experienced fluctuations. The highest validation accuracy of approximately 96% was achieved around the 9th epoch. This signifies that our model can generalize its learned features to unseen data, highlighting its potential for practical defect classification in real-world applications. Invariably, the model is not just learning to memorize the training data, but it is also learning to identify patterns in the data that can be used to classify new data accurately.

For training loss, which measures the difference between the model's predictions and the true values within the training dataset, started minimal at approximately 0.12 and fluctuated within a narrow range throughout the training epochs, with slight drops at the 5th, 10th, and 14th epoch respectively. The validation loss initially increased slightly, but then it decreased and stabilized at a lower level, which suggests the model is robust and not overfitting the training data. The low validation loss values, combined with high validation accuracy, indicate the model's ability to generalize effectively to new defect samples., which is important for practical defect classification tasks. The combined model's low and stable loss values indicate that it is converging and learning effectively, which is a fundamental characteristic of well-trained neural networks.

In contrast to the first model, the model accuracy improves while the loss reduces for every epoch. This means the model is learning the right features and can generalise effectively. Therefore, training a deep-learning network with more data and additional provisions for

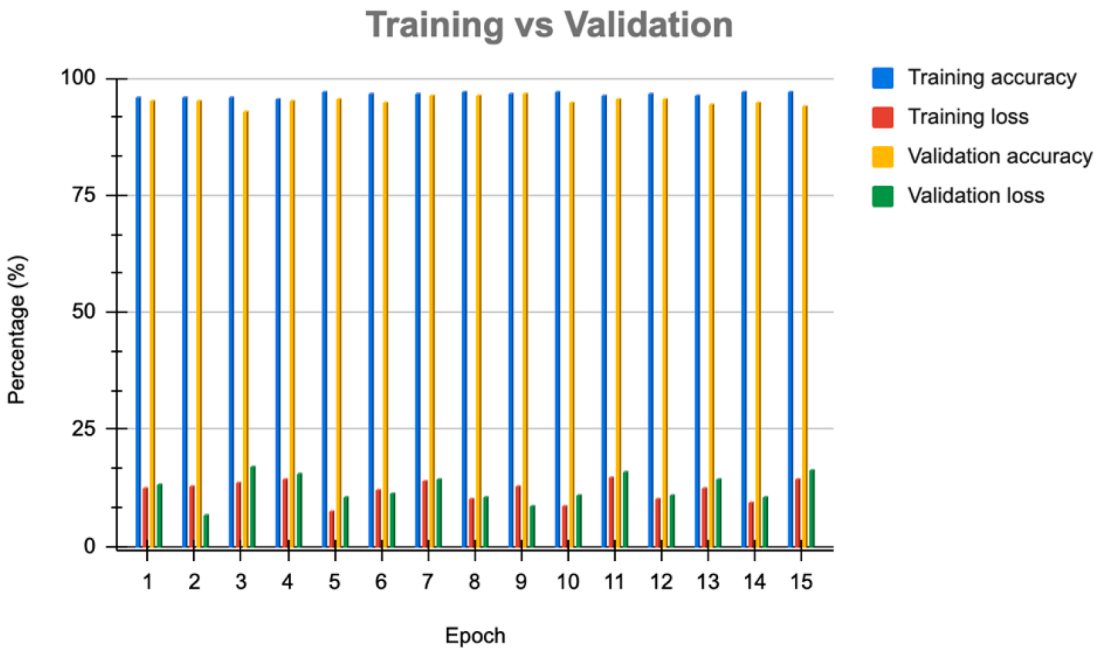


Fig. 17. Training vs Validation (augmentation + dropout).

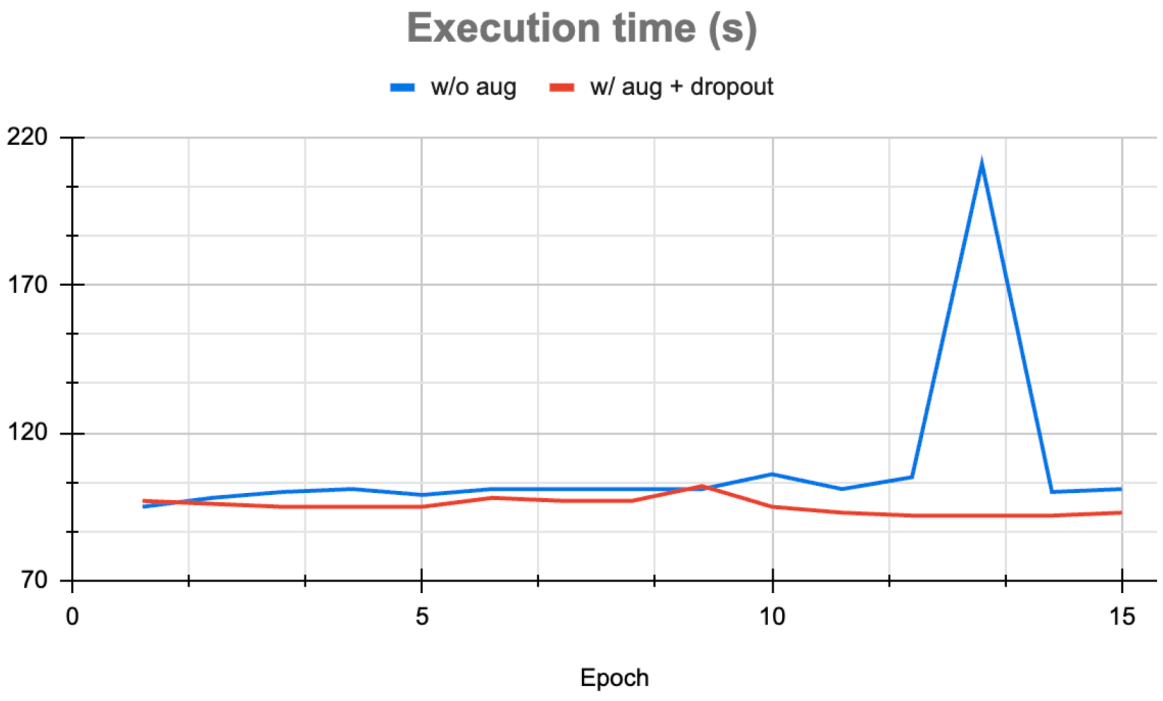


Fig. 18. Model time complexity.

Table 6
Model evaluation result.

Accuracy	Precision	Recall	F1-Score
0.96	0.93	0.94	0.93

overfitting (such as dropout) greatly influences the model's performance. However, despite using a big data cluster to pre-process the images to lower computational time, we recorded an average of 23 minutes of computational time in 15 epochs using more than 50,000 images, compared to the previous 22 min with less than 3000 images in the first model as seen in Fig. 18. Owing to this, we also check the effect of augmentation and dropout on the training time as shown below.

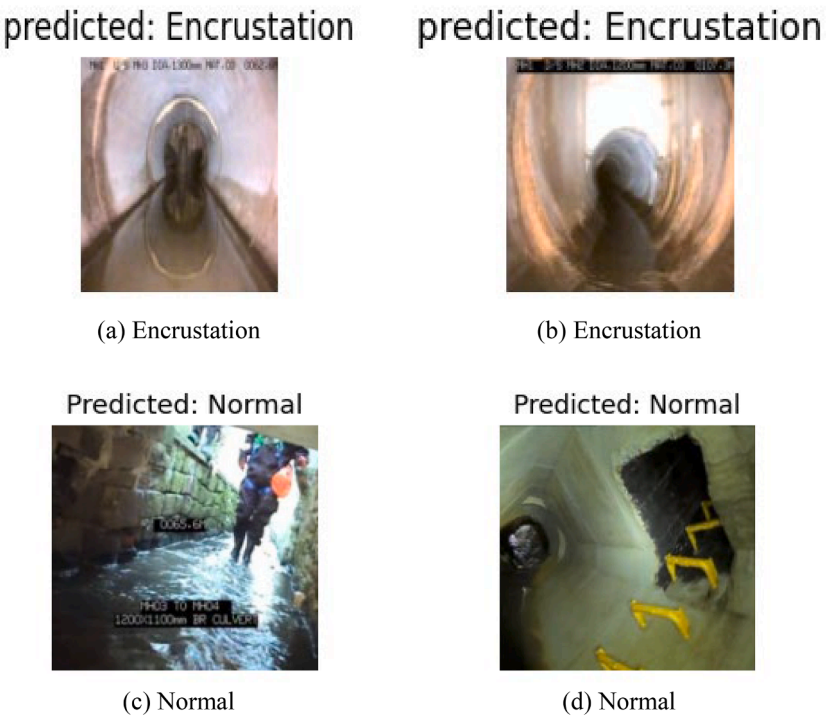


Fig. 19. Classification result from test data.

	Models	Accuracy	Precision	Recall	F1-Score
1.	VGG19	0.967	0.936	0.944	0.940
2.	VGG16	0.828	0.828	0.828	0.828
3.	EfficientNet	0.951	0.922	0.931	0.926
4.	AlexNet	0.617	0.617	0.617	0.617

Fig. 20. Model result comparison.

Table 7
List of studies on defect classification using Accuracy.

Authors	Architecture	Defect type	Data size	Augmentation	Metric	Result
(Dang et al., 2021b)	CNN	crack, debris silty, joint faulty, open joint, surface damage, protruding lateral, pip broken	38,386	Horizontal flip, shear range, and zoom range	Accuracy	97.8
Ours	**VGG19**	**Encrustation**	**2100**	**Rotation, brightness, flipping, zooming, flipping, and channel shuffle.**	**Accuracy**	**96.7**
(Hassan et al., 2019)	CNN	longitudinal defect, debris silty, joint faulty, open joint, surface damage, potruding lateral	24,137	Horizontal flip	Accuracy	96.33
(Khan et al., 2019)	CNN	Crack	627	-	Accuracy	83.3
(Wang et al., 2023)	PIP-CovNET	cracks, breaks, and collapses;	390,078	-	Accuracy	82
(Chen et al., 2018b)	CNN	Deposit, obstacle, blur, intrusion	10,000	Random cropping, PCA, colour enhancement, and rotation transformation	accuracy	81

Table 8
Comparing models on defect classification using F1-Score.

Authors	Architecture	Defect type	Data size	Augmentation	Metric	Result
Ours	**VGG19**	**Encrustation**	**2100**	**Rotation, brightness, flipping, zooming, flipping, and channel shuffle.**	**F1**	**94**
(Ma et al., 2022)	Resnet	crack	5729	Horizontal flipping and scaling	F1	92.8
(Hu et al., 2023)	SPM+CNN	sewer-ml defect	1,700,000	-	F1, F2	91.6
(Xie et al., 2019)	CNN	deposition, stagger, fracture, high water level, disjunction, barrier	40,000	rotation, flipping, and colour jittering	F1	84
(Zhao et al., 2023)	TMSDC	sewer-ml defect	1,700,000	Random flip, jitter of pixel values - brightness, contrast, saturation, and hue	F1	81.5
(Haurum et al., 2021)	DGCNN	Displaced joints, defective rubber rings, and obstacles	17,027	-	F1	-
(Haurum et al., 2022)	Resnet	sewer-ml defect	1,700,000	Horizontal flip	F1, F2	-

Table 9
Comparing models on defect classification using Precision and Recall.

Authors	Architecture	Defect type	Data size	Augmentation	Metric	Result
(Li et al., 2023)	RegNetY + Grad-CAM	root, disjointedness, deformation, deposit, rupture, misalignment, scaling, and interface material failure.	2562	Geometric transformation, colour transformation, mirror, Blur.	F1, Recall	95.4
Ours	**VGG19**	**Encrustation**	**2100**	**Rotation, brightness, flipping, zooming, flipping, and channel shuffle.**	**Precision, Recall**	**94**
(Kumar et al., 2018)	CNN	Tree root, deposits, cracks, infiltration, debris, connections, material change	12,000	Random flips, brightness changes, contrast changes, and motion blur	Accuracy, Precision, Recall	86.2
(Zhou et al., 2022a)	CNN	brick, rubber ring, and displacement	1024	-	Precision, Recall	77

The slight difference in the time complexity is due to the cluster size used. We strongly believe using a powerful cluster with high processing units will take less time to pre-process and analyse these images. Consideration was given to spinning a spot instance (virtual server) on AWS to carry out this experiment, but the associated cost is beyond the scope of this study. This means consideration must be given to computational cost while choosing a training sample size.

4.3.2. Model evaluation

We evaluated the model on the test samples using standardized performance metrics to guarantee validity. The key metrics used included accuracy, precision, recall, and F1-score. Table 6 below provides a summary of the evaluation.

The results indicate a high degree of accuracy on the test set, which suggests that the model is highly effective in correctly classifying encrustation images. The precision of 93 % paired with a recall of 94 % indicates a strong balance between the model's ability to identify relevant instances and its tendency to avoid false positives, which is a critical factor for defect detection systems. The *F1-score*, a harmonic mean of precision and recall, further confirms the model's robust performance in this context. Fig. 19 provided sample classifications from test data.

To further validate and compare our model, we additionally trained new models using separate CNN models like AlexNet (Krizhevsky et al., 2017), EfficientNet (Tan & Le, 2019), and VGG16 (Simonyan & Zisserman, 2014) using the same dataset and evaluation metrics. The results reveal substantial disparities in performance among the models, which is critical in selecting the appropriate model for encrustation detection. As depicted in Fig. 20, VGG19 has superior performance with an accuracy of 96 %, alongside high precision, recall, and F1-score, by using its deeper architecture to capture more convoluted features of encrustation. EfficientNet also provides very good results, with over 95 % accuracy, by systematically scaling network dimensions to enhance both computational efficiency and accuracy.

In contrast, while VGG16 offered moderate performance throughout, AlexNet significantly underperformed on all metrics compared to other models. Their low performance shows the advancement of CNN models towards efficiency, performance, and applicability towards different problem categories. Lastly, we compared our model with other existing classification models that have employed deep learning in classifying defects in sewers. To the authors' best understanding, there have not been any studies in the past that have applied deep learning to encrustation detection. Hence, it became difficult to make a toe-to-toe comparison with existing studies based on the defect type (i.e., encrustation) considered in this study. However, this study has looked deeply into the literature to filter out classification models targeting other types of defects while using different backbone networks, augmentation techniques, and evaluation metrics.

A list of these studies is summarised in Tables 7, 8, 9 and ranked according to the reported result using different metrics. It is evident from the tables that all existing studies have used thousands of raw images more than ours except for (Khan et al., 2019) and (Zhou et al., 2022b) which used 627 and 1024 images encompassing four different defect types, however, this impacted their model performance (83 % and 77 %). In addition, the number of augmentation techniques employed in this study (6) is the joint highest with the work done by Li et al. (2019). However, they applied these augmentations to 6 different defect types, compared to a single defect considered in this study. To drill down into the comparison, we have compared and grouped existing studies into three categories based on the evaluation metrics employed – accuracy,

F1, and a combination of recall and precision.

With accuracy, our model shows robustness in identifying encrustations accurately within our dataset. Nevertheless, there is room for optimization as we recorded a few defect misclassifications. Furthermore, since accuracy alone cannot fully capture the model's performance degrees (Kumar et al., 2018), such as its ability to manage the trade-off between false positives and false negatives, which is critical in defect detection applications where the cost of missing a defect (false negative) could be significant. Consequently, we further compared our model with existing studies using the F1-score.

Notably, our model outperforms all other models as shown in Table 8, which highlights its higher capability to balance precision and recall effectively. This is especially beneficial in defect detection applications where the accurate identification of defects and the minimization of false positives are equally critical (Zhou et al., 2022a). Tables 8-9 also compared our model based on a combination of evaluation metrics. See *Appendix A* (Table 10) for a full list of all studies, including the ones with evaluation metrics not used in this study.

4.3. Impacts on practice

Due to the drawbacks of traditional defect classification, there is a clear industry need for a quicker, cheaper, and safer system to analyse survey videos efficiently and allow multiple surveys yearly to forestall pollution incidents that attract revenue-crushing penalties. To mitigate this, more collaboration is required from sewer inspection and management companies on data sharing. Constructing a unified database for survey videos of hundreds of sewers will provide enough data to extract hundreds of thousands of defective images, thereby eliminating the need for data augmentation. Although, open-source datasets such as SewerML (Harum & Moeslund, 2021) can assist with this, however, the dataset was only from 3 utility companies in a specific region (i.e., Denmark).

This study proposes having a shared repository of survey videos or images by multiple inspection companies which span across different regions. This will aid a more robust detection system that not only improves research in this area but could also automatically trigger automatic report generation after fault detection. This report will be based on unified standards for defect classification and severity grading. The automatic defect classification and reporting process can be automated with little human intervention by building APIs around the model and deploying it on the cloud for real-time usage using the Software as a Service (SaaS) approach. This will only require the survey videos to be uploaded from a web interface like an attachment, and the deep-learning model will break the uploaded videos into streams of images and iterate on the images to detect defects in them. Based on the type and severity of the fault detected, a detailed report of the defect will be generated for end-user review, validation, and feedback. We believe the feedback will guide further model improvements to increase the model's robustness.

5. Conclusion

This study adopted computer vision and deep-learning techniques to develop a model for automatic surface encrustation detection and classification in sewers by leveraging inspection videos from CCTV. We used 2100 encrustation and non-encrustation images from 14 different sewers across the UK to develop CNN models using backbone networks such as AlexNet, VGG16, EfficientNet, and VGG19. This work is the first to specifically address encrustation detection using deep learning, whereas previous research has focused on other types of deposits like settled and attached deposits. Data augmentation was employed on the training data to mitigate the limitations of our small sample size, thereby reducing overfitting and improving model generalization. We authenticated the effectiveness of this approach by conducting experiments on both the small training sample and the augmented sample, achieving an accuracy of 96 %, recall of 94 %, precision of 93 %, and F1-score of 93 %. These results demonstrate that the model can reliably distinguish between images with encrustation and normal images.

Our study also included a comparison of different backbone networks to identify the most effective model for encrustation detection. We found that VGG19 outperformed other networks, showcasing its superior capability in this application. Additionally, this research highlights the impact of data augmentation and network dropout not only on reducing overfitting and improving accuracy but also on the time complexities involved in training models with and without data augmentation. In conclusion, we have demonstrated that it is feasible to develop an accurate model for surface encrustation detection in sewers using various CNN models, with VGG19 performing the best among the networks tested. While a larger dataset would further enhance model robustness and generalization, data augmentation remains a viable method to compensate for limited data. However, the choice of augmentation techniques must be carefully considered, as they are not universally applicable. Future studies should aim to utilize larger datasets to enhance feature learning and generalization capabilities. Additionally, employing multi-class defect detection for images containing encrustation and other defects using more advanced deep-learning algorithms could further improve detection accuracy and efficiency.

CRediT authorship contribution statement

Wasiu Yusuf: Methodology, Investigation, Writing – original draft, Writing – review & editing, Visualization. **Hafiz Alaka:** Conceptualization, Supervision, Funding acquisition. **Mubashir Ahmad:** Data curation. **Wusu Godoyon:** Validation. **Saheed Ajayi:** Project administration. **Luqman Olalekan Toriola-Coker:** Formal analysis. **Abdullahi Ahmed:** Resources.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Funding

This research was supported by Innovate UK [grant number: 10004446].

Appendix A

Table 10
List of deep learning studies on defect classification.

Authors	Architecture	Defect type	Data size	Augmentation techniques	Metrics	Result
(Dang et al., 2021b)	CNN	crack, debris silty, joint faulty, open joint, surface damage, potruding lateral, pip broken	38,386	Horizontal flip, shear range, and zoom range"	Accuracy	97.8
(Hassan et al., 2019)	CNN	longitudinal defect, debris silty, joint faulty, open joint, surface damage, potruding lateral	24,137	Horizontal flip	Accuracy	96.33
Ours	VGG19	Encrustation	2100	Rotation, brightness, flipping, zooming, flipping, and channel shuffle.	Accuracy	95
(Khan et al., 2019)	CNN	Crack	627	-	Accuracy	83.3
(Chen et al., 2018b)	CNN	Deposit, obstacle, blur, intrusion	10,000	Random cropping, PCA, colour enhancement, and rotation transformation	accuracy	81
(Wang et al., 2023)	PIP-CovNET	cracks, breaks, and collapses;	390,078	-	Accuracy, F1	82
(Kumar et al., 2018)	CNN	Tree root, deposits, cracks, infiltration, debris, connections, material change	12,000	Random flips, brightness changes, contrast changes, and motion blur	Accuracy, Precision, and Recall	86.2
(Li et al., 2019)	Resnet	Deposit, settlement, joint offset, broken, obstacles, water level stag, and deformation	18,352	Random horizontal flip, random cropping, Gaussian blur, contrast normalization, additive Gaussian noise and channel scaling	AUC	64.8
(Meijer et al., 2019)	CNN	Fissures, Surface damage, intruding connection, Defective connection, Intruding sealing material, Displaced joint, Porous pipe, Roots Attached deposits, Settled deposits Ingress of soil, Infiltration	17,662	-	AUROC, AUPR, specificity, precision	90
(Ma et al., 2022)	Resnet	crack	5729	Horizontal flipping and scaling	F1	92.8
(Xie et al., 2019)	CNN	deposition, stagger, fracture, high water level, disjunction, barrier	40,000	rotation, flipping, and colour jittering	F1	84
(Zhao et al., 2023)	TMSDC	sewer-ml defect	1,700,000	Random flip, jitter of pixel values - brightness, contrast, saturation, and hue	F1	81.5
(Haurum et al., 2021)	DGCNN	Displaced joints, defective rubber rings, and obstacles	17,027	-	F1	-
(Hu et al., 2023)	SPM+CNN	sewer-ml defect	1,700,0000	-	F1, F2	91.6
(Haurum et al., 2022)	Resnet	sewer-ml defect	1,700,000	Horizontal flip	F1, F2	-
(Li et al., 2023)	RegNetY + Grad-CAM	root, disjointedness, deformation, deposit, rupture, misalignment, scaling, and interface material failure	2562	Geometric transformation, colour transformation, mirror, Blur.	F1, Recall	95.4
(Zhong, et al., 2019)	CNN	-	36,000	Flipping, rotating and scaling, and colour adjustment.	FNR, ATC	95.9
(Zhou et al., 2022a)	CNN	brick, rubber ring, and displacement	1024	-	precision, recall, F1-score	77

References

A. Rosebrock. (2021, May). *Convolutional Neural Networks (CNNs) and Layer Types*. Https://Pyimagesearch.Com/2021/05/14/Convolutional-Neural-Networks-Cnns-and-Layer-Types/.

Aggarwal, P. (2019). Data augmentation in dermatology image recognition using machine learning. *Skin Research and Technology, 25*(6), 815–820. https://doi.org/10.1111/srt.12726

Agnieszka, M., & Michał, G. (2018). Data augmentation for improving deep learning in image classification problem. *International Interdisciplinary PhD Workshop (IIPhDW), 9–12. May 2018*.

Ahmed, T., Das, P., Ali, M. F., & Mahmud, M.-F. (2020). A Comparative Study on Convolutional Neural Network Based Face Recognition. In *2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)* (pp. 1–5). https://doi.org/10.1109/ICCCNT49239.2020.9225688

Ajit, A., Acharya, K., & Samanta, A. (2020). A review of convolutional neural networks. In *, 2020. International Conference on Emerging Trends in Information Technology and Engineering, Ic-ETITE*. https://doi.org/10.1109/ic-ETITE47903.2020.049

Aşchilean, I., Iliescu, M., Ciont, N., & Giurca, I. (2018). The unfavourable impact of street traffic on water distribution pipelines. *Water (Switzerland), 10*(8). https://doi.org/10.3390/w10081086

Ayanwola, T., Oludele, A., & Agbaje, M. (2023). Enhancing face spoofing attack detection: Performance evaluation of a VGG-19 CNN Model. *Acadlore Transactions on AI and Machine Learning, 2*(2), 84–98. https://doi.org/10.56578/ataiml020204

Bartsch, K., Pettke, A., Höbert, A., Lakämper, J., & Lange, F. (2021). On the digital twin application and the role of artificial intelligence in additive manufacturing: A systematic review. In *JPhys materials, 4*. IOP Publishing Ltd. https://doi.org/10.1088/2515-7639/abf3cf

Bouhsissin, S., Sael, N., & Benabbou, F. (2021). Enhanced VGG19 model for accident detection and classification from video. In *2021 International Conference on Digital Age & Technological Advances for Sustainable Development (ICDATA)* (pp. 39–46). https://doi.org/10.1109/ICDATA52997.2021.00017

Chen, K., Hu, H., Chen, C., Chen, L., & He, C. (2018a). An intelligent sewer defect detection method based on convolutional neural network. In *, 2018. 2018 IEEE International Conference on Information and Automation, ICIA* (pp. 1301–1306). https://doi.org/10.1109/ICInfA.2018.8812445

Chen, K., Hu, H., Chen, C., Chen, L., & He, C. (2018b). An intelligent sewer defect detection method based on convolutional neural network. In *, 2018. 2018 IEEE International Conference on Information and Automation, ICIA* (pp. 1301–1306). https://doi.org/10.1109/ICInfA.2018.8812445

Cheng, J. C. P., & Wang, M. (2018). Automated detection of sewer pipe defects in closed-circuit television images using deep learning techniques. *Automation in Construction, 95*, 155–171. https://doi.org/10.1016/j.autcon.2018.08.006

ClearView Survey. (2023). *Sewer cleaning*. Https://Www.Clearviewsurveys.Co.Uk/Services/Sewer-Cleaning/.

Dang, L. M., Kyeong, S. J., Li, Y., Wang, H., Nguyen, T. N., & Moon, H. (2021a). Deep learning-based sewer defect classification for highly imbalanced dataset. *Computers and Industrial Engineering, 161*. https://doi.org/10.1016/j.cie.2021.107630

Dang, L. M., Kyeong, S., Li, Y., Wang, H., Nguyen, T. N., & Moon, H. (2021b). Deep learning-based sewer defect classification for highly imbalanced dataset. *Computers and Industrial Engineering, 161*. https://doi.org/10.1016/j.cie.2021.107630

Decor, G., Bah, M. D., Foucher, P., Charbonnier, P., & Heitz, F. (2019). Defect detection in tunnel images using random forests and deep learning. In *International Conference on Pattern Recognition Systems (ICPRS)*.

DEFRA. (2014). Sewage treatment in the UK UK implementation of the EC urban waste water treatment directive. *Department for Environment, Food & Rural Affairs*.

Do, T.-D., Tuyet-Doan, V.-N., Cho, Y.-S., Sun, J.-H., & Kim, Y.-H. (2020). Convolutional-neural-network-based partial discharge diagnosis for power transformer using UHF sensor. *IEEE Access, 8*, 207377–207388. https://doi.org/10.1109/ACCESS.2020.3038386

Jain, Dr. Madhur, Bora, Mayank Singh, Chandnani, Sameer, Grover, Sanidhay, & Sadwal, Shivank (2023). Comparison of VGG-16, VGG-19, and ResNet-101 CNN models for the purpose of suspicious activity detection. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, 121–130. https://doi.org/10.32628/CSEIT2390124

Fabian, Z., Heckel, R., & Soltanolkotabi, M. (2021). Data augmentation for deep learning based accelerated MRI reconstruction with limited data. In *, 2021. Proceedings of the 38th International Conference on Machine Learning, PMLR 139*. https://github.com/MathFLDS/.

Fang, X., Guo, W., Li, Q., Zhu, J., Chen, Z., Yu, J., Zhou, B., & Yang, H. (2020). Sewer pipeline fault identification using anomaly detection algorithms on video sequences. *IEEE Access, 8*, 39574–39586. https://doi.org/10.1109/ACCESS.2020.2975887

Fatma, S., Brijesh, V., & Md, A. (2016). *2016 International conference on digital image computing : techniques and applications (DICTA)*. IEEE.

Fu, W., & Menzies, T. (2017). Easy over hard: A case study on deep learning. In *Proceedings of the ACM SIGSOFT Symposium on the Foundations of Software Engineering, Part F130154* (pp. 49–60). https://doi.org/10.1145/3106237.3106256

Gao, X., Jian, M., Hu, M., Tanniru, M., & Li, S. (2019). Faster multi-defect detection system in shield tunnel using combination of FCN and faster RCNN. *Advances in Structural Engineering, 22*(13), 2907–2921. https://doi.org/10.1177/1369433219849829

Grigorescu, S., Trasnea, B., Cocias, T., & Macesanu, G. (2020). A survey of deep learning techniques for autonomous driving. *Journal of Field Robotics, 37*(3), 362–386. https://doi.org/10.1002/rob.21918

Guo, W., Soibelman, L., & Garrett, J. H. (2009). Automated defect detection for sewer pipeline inspection and condition assessment. *Automation in Construction, 18*(5), 587–596. https://doi.org/10.1016/j.autcon.2008.12.003

Gutierrez-Mondragon, M.A., Garcia-Gasulla, D., Alvarez-Napagao, S., Brossa-Ordoñez, J., & Gimenez-Esteban, R. (2020). *Obstruction level detection of sewer videos using convolutional neural networks*. http://arxiv.org/abs/2002.01284.

Harum, B. J., & Moeslund, T. (2021). Sewer-ML: A multi-label sewer defect classification dataset and benchmark. In *Computer Vision Proceedings*. http://vap.aau.dk/sewer-ml.

Hassan, S. I., Dang, L. M., Mehmood, I., Im, S., Choi, C., Kang, J., Park, Y.-S., & Moon, H. (2019). Underground sewer pipe condition assessment based on convolutional neural networks. *Automation in Construction, 106*. https://doi.org/10.1016/j.autcon.2019.102849

Haurum, J. B., Allahham, M. M. J., Lynge, M. S., Henriksen, K. S., Nikolov, I. A., & Moeslund, T. B. (2021). Sewer defect classification using synthetic point clouds. In *, 5. VISIGRAPP 2021 - Proceedings of the 16th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications* (pp. 891–900). https://www.scopus.com/inward/record.uri?eid=2-s2.0-85102968084&partnerID=40&md5=98e3dd3754e9b467bafd0a5aa1f702f0.

Haurum, J. B., Madadi, M., Escalera, S., & Moeslund, T. B. (2022). Multi-scale hybrid vision transformer and Sinkhorn tokenizer for sewer defect classification. *Automation in Construction, 144*. https://doi.org/10.1016/j.autcon.2022.104614

Hawari, A., Alamin, M., Alkadour, F., Elmasry, M., & Zayed, T. (2018). Automated defect detection tool for closed circuit television (cctv) inspected sewer pipelines. *Automation in Construction, 89*, 99–109. https://doi.org/10.1016/j.autcon.2018.01.004

Hu, C., Dong, B., Shao, H., Zhang, J., & Wang, Y. (2023). Toward purifying defect feature for multilabel sewer defect classification. *IEEE Transactions on Instrumentation and Measurement, 72*. https://doi.org/10.1109/TIM.2023.3250306

Hu, M., Liu, Y., Sugumaran, V., Liu, B., & Du, J. (2019). Automated structural defects diagnosis in underground transportation tunnels using semantic technologies. *Automation in Construction, 107*. https://doi.org/10.1016/j.autcon.2019.102929

Hughes, J., Cowper-Heays, K., Olesson, E., Bell, R., & Stroombergen, A. (2021). Impacts and implications of climate change on wastewater systems: A New Zealand perspective. In *Climate risk management, 31*. Elsevier B.V. https://doi.org/10.1016/j.crm.2020.100262

Javid, A. M., Das, S., Skoglund, M., & Chatterjee, S. (2021). A relu dense layer to improve the performance of neural networks. In *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings, 2021-June* (pp. 2810–2814). https://doi.org/10.1109/ICASSP39728.2021.9414269

Jiang, F., Jiang, Y., Zhi, H., Dong, Y., Li, H., Ma, S., Wang, Y., Dong, Q., Shen, H., & Wang, Y. (2017). Artificial intelligence in healthcare: Past, present and future. In *Stroke and vascular neurology, 2* pp. 230–243). BMJ Publishing Group. https://doi.org/10.1136/svn-2017-000101

Khan, M. S., Zeng, K., Wu, N., & Unwala, I. (2019). Robotics and deep learning framework for structural health monitoring of utility pipes. In *2019 22nd IEEE International Symposium on Measurement and Control in Robotics: Robotics for the Benefit of Humanity, ISMCR 2019*. https://doi.org/10.1109/ISMCR47492.2019.8955723

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM, 60*(6), 84–90. https://doi.org/10.1145/3065386

Kumar, S. S., Abraham, D. M., Jahanshahi, M. R., Iseley, T., & Starr, J. (2018). Automated defect classification in sewer closed circuit television inspections using deep convolutional neural networks. *Automation in Construction, 91*, 273–283. https://doi.org/10.1016/j.autcon.2018.03.028

Lepot, M., Stanić, N., & Clemens, F. H. L. R. (2017). A technology for sewer pipe inspection (Part 2): Experimental assessment of a new laser profiler for sewer defect detection and quantification. *Automation in Construction, 73*, 1–11. https://doi.org/10.1016/j.autcon.2016.10.010

Li, D., Cong, A., & Guo, S. (2019). Sewer damage detection from imbalanced CCTV inspection data using deep convolutional neural networks with hierarchical classification. *Automation in Construction, 101*, 199–208. https://doi.org/10.1016/j.autcon.2019.01.017

Li, D., Xie, Q., Yu, Z., Wu, Q., Zhou, J., & Wang, J. (2021). Sewer pipe defect detection via deep learning with local and global feature fusion. *Automation in Construction, 129*. https://doi.org/10.1016/j.autcon.2021.103823

Li, M., Li, M., Ren, Q., Liu, H., & Liu, C. (2023). Intelligent identification and classification of sewer pipeline network defects based on improved RegNetY network. *Journal of Civil Structural Health Monitoring, 13*(2–3), 547–560. https://doi.org/10.1007/s13349-022-00660-7

Li, Y., Wang, H., Dang, L. M., Song, H.-K., & Moon, H. (2022b). Vision-Based Defect Inspection and Condition Assessment for Sewer Pipes_ A Comprehensive Survey _ Enhanced Reader. *Sensors*.

Lin, Y. Z., Nie, Z. H., & Ma, H. W. (2017). Structural damage detection with automatic feature-extraction through deep learning. *Computer-Aided Civil and Infrastructure Engineering, 32*(12), 1025–1046. https://doi.org/10.1111/mice.12313

Liu, Y., Yao, J., Lu, X., Xie, R., & Li, L. (2019). DeepCrack: A deep hierarchical feature learning architecture for crack segmentation. *Neurocomputing, 338*, 139–153. https://doi.org/10.1016/j.neucom.2019.01.036

Lyu, Y., Li, H., Sayagh, M., Jiang, Z. M.(Jack), & Hassan, A. E. (2021). An empirical study of the impact of data splitting decisions on the performance of AIOps solutions. *ACM Transactions on Software Engineering and Methodology, 30*(4), 1–38. https://doi.org/10.1145/3447876

Ma, B., Zuo, X., Shen, J., Xin, S., Shucheng, H., & Yuanhao, L. (2022). Training a lightweight CNN model for fine-grained sewer pipe cracks classification based on

knowledge distillation. In *International Conference on Pattern Recognition and Artificial Intelligence (PRAI)*. https://ieeexplore.ieee.org/abstract/document/9904106/.
Meijer, D., Scholten, L., Clemens, F., & Knobbe, A. (2019). A defect classification methodology for sewer image sets with convolutional neural networks. *Automation in Construction, 104*, 281–298. https://doi.org/10.1016/j.autcon.2019.04.013
Michael, F. (2017). *Five years after the Deep Learning revolution of computer vision: State of the art methods for online image and video*.
Moradi, S., & Zayed, T. (2017). Real-time defect detection in sewer closed circuit television inspection videos. In *Pipelines 2017: Condition Assessment, Surveying, and Geomatics - Proceedings of Sessions of the Pipelines 2017 Conference* (pp. 295–307). https://doi.org/10.1061/9780784480885.027
Moskalenko, V., Zaretskyi, M., Moskalenko, A., & Lysyuk, V. (2020). Sewer pipe defects classification based on deep convolutional network with information-extreme error-correction decision rules. *Communications in Computer and Information Science, 1158*, 253–263. https://doi.org/10.1007/978-3-030-61656-4_16
Mostafa, K., & Hegazy, T. (2021). Review of image-based analysis and applications in construction. *Automation in construction*. Elsevier B.V. https://doi.org/10.1016/j.autcon.2020.103516
Mushtaq, Z., & Su, S. F. (2020). Environmental sound classification using a regularized deep convolutional neural network with data augmentation. *Applied Acoustics, 167*. https://doi.org/10.1016/j.apacoust.2020.107389
Myrans, J., Everson, R., & Kapelan, Z. (2018a). Automated detection of faults in sewers using CCTV image sequences. *Automation in Construction, 95*, 64–71. https://doi.org/10.1016/j.autcon.2018.08.005
Myrans, J., Kapelan, Z., & Everson, R. (2018b). Using automatic anomaly detection to identify faults in sewers. In *International WDSA /CCWI Joint Conference*.
Ni, J., Chen, Y., Chen, Y., Zhu, J., Ali, D., & Cao, W. (2020). A survey on theories and applications for self-driving cars based on deep learning methods. In *Applied sciences (Switzerland), 10*. MDPI AG. https://doi.org/10.3390/APP10082749
Office for National Statistics. (2021, January). *Overview of the UK population: January 2021*. Https://Www.Ons.Gov.Uk/Peoplepopulationandcommunity/Populationandmigration/Populationestimates/Articles/Overviewoftheukpopulation/January2021.
Ofwat. (2023, February 20). *Thames Tideway*. Https://Www.Ofwat.Gov.Uk/Households/Supply-and-Standards/Thames-Tideway/#:~:Text=Each%20year%20around%2039%20million,And%20its%20aging%20sewer%20system.
Ottoni, A. L. C., Novo, M. S., & Oliveira, M. S. (2023). A statistical approach to hyperparameter tuning of deep learning for construction machine classification. *Arabian Journal for Science and Engineering*. https://doi.org/10.1007/s13369-023-08330-6
Pan, G., Zheng, Y., Guo, S., & Lv, Y. (2020a). Automatic sewer pipe defect semantic segmentation based on improved U-Net. *Automation in Construction, 119*. https://doi.org/10.1016/j.autcon.2020.103383
Pan, Y., Zhang, G., & Zhang, L. (2020b). A spatial-channel hierarchical deep learning network for pixel-level automated crack detection. *Automation in Construction, 119*. https://doi.org/10.1016/j.autcon.2020.103357
RapidResponse Drain Care. (2023). *Blocked Pipes and Drains*. Https://Www.Blockage.Co.Uk/Blocked-Pipes-Drains/.
Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2015). *You only look once: unified, real-time object detection*. http://arxiv.org/abs/1506.02640.
Ren, Z., Fang, F., Yan, N., & Wu, Y. (2022). State of the art in defect detection based on machine vision. *International Journal of Precision Engineering and Manufacturing-Green Technology, 9*(2), 661–691. https://doi.org/10.1007/s40684-021-00343-6
Savino, P., & Tondolo, F. (2021). Automated classification of civil structure defects based on convolutional neural network. *Frontiers of Structural and Civil Engineering, 15*(2), 305–317. https://doi.org/10.1007/s11709-021-0725-9
Shang, X., Li, C., Liu, M., Jiang, C., & Yang, F. (2019). Automatic drainage pipeline defect detection method using handcrafted and network features. In *International Conference on Unmanned Systems and Artificial Intelligence*.
Shiv Kumar, S., Abraham, D. M., Jahanshahi, M., & Iseley, T (2018). A deep learning framework for automated defect detection using sewer CCTV videos. *NASTT's 2018 No-Dig Show*.
Simonyan, K., & Zisserman, A. (2014). *Very deep convolutional networks for large-scale image recognition*. http://arxiv.org/abs/1409.1556.
Situ, Z., Teng, S., Feng, W., Zhong, Q., Chen, G., Su, J., & Zhou, Q. (2023a). A transfer learning-based YOLO network for sewer defect detection in comparison to classic object detection methods. *Developments in the Built Environment, 15*, Article 100191. https://doi.org/10.1016/j.dibe.2023.100191
Situ, Z., Teng, S., Liao, X., Chen, G., & Zhou, Q. (2023b). Real-time sewer defect detection based on YOLO network, transfer learning, and channel pruning algorithm. *Journal of Civil Structural Health Monitoring*. https://doi.org/10.1007/S13349-023-00681-W
Smith, M. L., Smith, L. N., & Hansen, M. F. (2021). The quiet revolution in machine vision - a state-of-the-art survey paper, including historical review, perspectives, and future directions. *Computers in Industry, 130*. https://doi.org/10.1016/j.compind.2021.103472
Srivastava, N., Hinton, G., Krizhevsky, A., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research, 15*.
Statista. (2021, August 21). *United Kingdom: Degree of urbanization from 2010 to 2020*. Https://Www.Statista.Com/Statistics/270369/Urbanization-in-the-United-Kingdom/.
Sutton, M. A. (2013). Computer vision based, non-contacting deformation and shape measurements: A revolution in progress. *Journal of the South Carolina Academy of Science, 11*(1).
Tan, M., & Le, Q (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. Eds.. In K. Chaudhuri, & R. Salakhutdinov (Eds.), *97. Proceedings of the 36th International Conference on Machine Learning* (pp. 6105–6114). PMLR https://proceedings.mlr.press/v97/tan19a.html
Tan, Y., Cai, R., Li, J., Chen, P., & Wang, M. (2021). Automatic detection of sewer defects based on improved you only look once algorithm. *Automation in Construction, 131*. https://doi.org/10.1016/j.autcon.2021.103912
Thiyagarajan, K., Kodagoda, S., & Ulapane, N. (2016). Data-driven machine learning approach for predicting volumetric moisture content of concrete using resistance sensor measurements. In *, 2016. Proceedings of the 2016 IEEE 11th Conference on Industrial Electronics and Applications, ICIEA* (pp. 1288–1293). https://doi.org/10.1109/ICIEA.2016.7603783
Thompson, N.C., Greenewald, K., Lee, K., & Manso, G.F. (2020). *The computational limits of deep learning*. http://arxiv.org/abs/2007.05558.
Varghese, S. (2023, February 20). *London's super sewer won't solve the city's epic poop problem*. Https://Www.Wired.Co.Uk/Article/Sewage-Environment-Climate-Change-London.
Wang, M., Kumar, S. S., & Cheng, J. C. P. (2021a). Automated sewer pipe defect tracking in CCTV videos based on defect detection and metric learning. *Automation in Construction, 121*. https://doi.org/10.1016/j.autcon.2020.103438
Wang, M., Luo, H., & Cheng, J. C. P. (2021b). Towards an automated condition assessment framework of underground sewer pipes based on closed-circuit television (CCTV) images. *Tunnelling and Underground Space Technology, 110*. https://doi.org/10.1016/j.tust.2021.103840
Wang, M. Z., & Cheng, J. C. P. (2019). Semantic segmentation of sewer pipe defects using deep dilated convolutional neural network. In *Proceedings of the 36th International Symposium on Automation and Robotics in Construction, ISARC 2019* (pp. 586–594). https://doi.org/10.22260/isarc2019/0078
Wang, X., Thiyagarajan, K., Kodagoda, S., & Zhang, M. (2023). PIPE-CovNet: Automatic in-pipe wastewater infrastructure surface abnormality detection using convolutional neural network. *IEEE Sensors Letters, 7*(4). https://doi.org/10.1109/LSENS.2023.3258543
Westphal, E., & Seitz, H. (2021). A machine learning method for defect detection and visualization in selective laser sintering based on convolutional neural networks. *Additive Manufacturing, 41*, Article 101965. https://doi.org/10.1016/j.addma.2021.101965
White, J., Hurlebaus, S., & Wimsatt, A. (2013). Structural impairment detection of tunnel linings using ultrasonic sensors. In *, 2. Structural Health Monitoring 2013: A Roadmap to Intelligent Structures - Proceedings of the 9th International Workshop on Structural Health Monitoring, IWSHM 2013* (pp. 1721–1728).
Xie, Q., Li, D., Xu, J., Yu, Z., & Wang, J. (2019). Automatic detection and classification of sewer defects via hierarchical deep learning. *IEEE Transactions on Automation Science and Engineering, 16*(4), 1836–1847. https://doi.org/10.1109/TASE.2019.2900170
Xu, Y., & Goodacre, R. (2018). On splitting training and validation set: A comparative study of cross-validation, bootstrap and systematic sampling for estimating the generalization performance of supervised learning. *Journal of Analysis and Testing, 2*(3), 249–262. https://doi.org/10.1007/s41664-018-0068-2
Xu, Y., Li, D., Xie, Q., Wu, Q., & Wang, J. (2021). Automatic defect detection and segmentation of tunnel surface using modified Mask R-CNN. *Measurement: Journal of the International Measurement Confederation, 178*. https://doi.org/10.1016/j.measurement.2021.109316
Xu, Y., Yang, J., Zhao, S., Wu, H., & Sawan, M. (2020). An end-to-end deep learning approach for epileptic seizure prediction. In *, 2020. Proceedings, 2020 IEEE International Conference on Artificial Intelligence Circuits and Systems : AICAS*.
Yu, J., de Antonio, A., & Villalba-Mora, E. (2022). Deep learning (CNN, RNN) applications for smart homes: A systematic review. In *Computers, 11*. MDPI. https://doi.org/10.3390/computers11020026
Yue, W., Li, C., Wang, S., Xue, N., & Wu, J. (2023). Cooperative incident management in mixed traffic of CAVs and human-driven vehicles. *IEEE Transactions on Intelligent Transportation Systems, 24*(11), 12462–12476. https://doi.org/10.1109/TITS.2023.3289983
Zhao, C., Hu, C., Shao, H., & Z, W. (2023). Towards trustworthy multi-label sewer defect classification via evidential deep learning. *Ieeexplore.Ieee.Org*. https://ieeexplore.ieee.org/abstract/document/10096569/.
Zhou, Q., Situ, Z., Teng, S., Chen, W., Chen, G., & Su, J. (2022a). Comparison of classic object-detection techniques for automated sewer defect detection. *Journal of Hydroinformatics, 24*(2), 406–419. https://doi.org/10.2166/hydro.2022.132
Zhou, Y., Ji, A., & Zhang, L. (2022b). Sewer defect detection from 3D point clouds using a transformer-based deep learning model. *Automation in Construction, 136*. https://doi.org/10.1016/j.autcon.2022.104163

Further reading

Chen, Y., Yin, X., Zhang, Q., Bouferguene, A., Zaman, H., Al-Hussein, M., Russell, R., & Kurach, L. (2019a). Productivity analysis of manual condition assessment for sewer pipes based on CCTV monitoring. In *CSCE Annual Conference*.
Chen, Y., Zhong, S., Chen, K., Chen, S., & Zheng, S. (2019b). Automated detection of sewer pipe defects based on cost-sensitive convolutional neural network. In *ACM International Conference Proceeding Series* (pp. 8–17). https://doi.org/10.1145/3372806.3372816
Li, Q., He, W., Zhang, X., Yang, Y., & Toe, T. T. (2022a). Comparative study of Transfer Learning and VGGNet based classification of Alzheimer's disease. In *2022 12th International Conference on Information Technology in Medicine and Education (ITME)* (pp. 88–93). https://doi.org/10.1109/ITME56794.2022.00029