International Journal of Scientific Research and Modern Technology (IJSRMT) ijsrmt.com
Volume 4, Issue 11, 2025
DOI: https://doi.org/10.38124/ijsrmt.v4i11.965

Bayesian-Optimised Latent Encoding and Agent-Based Simulation for Enhanced Medical Image Character Recognition

Efosa Osagie[1]; Wei Ji[2]; Na Helian[3]

[1]Department of Computer Science & Data Science, York St. John University, London UK
[2,3]Department of Computer Science, University of Hertfordshire, Hertfordshire, UK

Publishing Date: 2025/11/19

Abstract

This paper presents a Bayesian-optimised Conditional Variational Autoencoder (CVAE) for synthetic data augmentation, embedded within an agent-based simulation framework. The CVAE systematically refines latent-space representations, generating high-quality synthetic character images that enhance dataset diversity and reduce the risk of overfitting. Bayesian optimisation ensures optimal latent variable selection, improving reconstruction accuracy while enabling scalable MICR training. The proposed agent-based system introduces autonomous agents: patient agents, doctor agents, imaging device agents, and recognition agents that collaborate to simulate real-world MICR workflows. This structured pipeline enables dynamic dataset augmentation while supporting medical diagnostics and automated text extraction. Experimental evaluations demonstrate significant performance improvements, with CNN models achieving accuracy gains of +3.2%, +3.5%, and +1.79% on the public dataset and +2.41%, +6.85%, and +1.60% on the private dataset when augmented with 50, 100, and 150 synthetic images per class, respectively. This research validates the effectiveness of Bayesian-tuned latent-space encoding and a supporting agent-based data augmentation, offering a scalable, computationally efficient solution for MICR enhancement.

Keywords: *Conditional Variational Autoencoder CVAE, Medical Image Character Recognition (MICR), Agent-Based Simulation Framework, Optical Character Recognition OCR, Data Augmentation, Latent Variable Modelling.*

I. INTRODUCTION

Medical Image Character Recognition (MICR) is the automated identification of alphanumeric characters embedded in medical imaging modalities (MIM), such as radiographs, ultrasound scans, and pathology slides. Unlike general Optical Character Recognition (OCR), which typically operates on structured, high-resolution text documents, MICR must contend with low-resolution, noisy, and spatially irregular character data. These characters often encode essential clinical metadata, making their accurate recognition critical for diagnostic workflows and data integrity. MICR faces persistent challenges due to limited dataset availability, which directly impacts deep learning (DL) model performance. Small datasets limit pattern generalisation, increase the risk of overfitting [1] and reduce the reliability of OCR models in medical contexts. Privacy constraints and high acquisition costs further limit access to annotated medical image datasets, necessitating the use of effective augmentation strategies.

This study proposes a Conditional Variational Autoencoder (CVAE) as a targeted solution for synthetic data augmentation in MICR. The CVAE approximates the probability distribution ($P(X)$) over high-dimensional image data (Doersch, 2016), learning pixel dependencies [2] to generate realistic samples that match the original data distributions [3]. The objective is to develop a generative model (P) that closely approximates ($P(X)$), producing synthetic images that expand training datasets and improve classification accuracy. Despite the promise of generative models, many rely on strong assumptions [4] and require computationally intensive inference methods [5]. Neural networks, used as numerical approximators [6, 7], offer more stable training. Among these, the Variational Autoencoder (VAE) is notable for its fast convergence and

Osagie, E., Ji, W., & Helian, N. (2025). Bayesian-Optimised Latent Encoding and Agent- Based Simulation for Enhanced Medical Image Character Recognition. *International Journal of Scientific Research and Modern Technology, 4*(11), 84–94. https://doi.org/10.38124/ijsrmt.v4i11.965

minimal assumptions [8], making it suitable for latent space encoding in MICR augmentation. Unlike traditional autoencoders, VAEs encode inputs as probability distributions rather than fixed points, enabling structured learning. The encoder computes mean (μ) and covariance (Σ), forming a Gaussian latent space that supports stable training [9]. Figure 1 illustrates the generative process, showing how latent space transformations contribute to MICR-specific augmentation [5, 10].

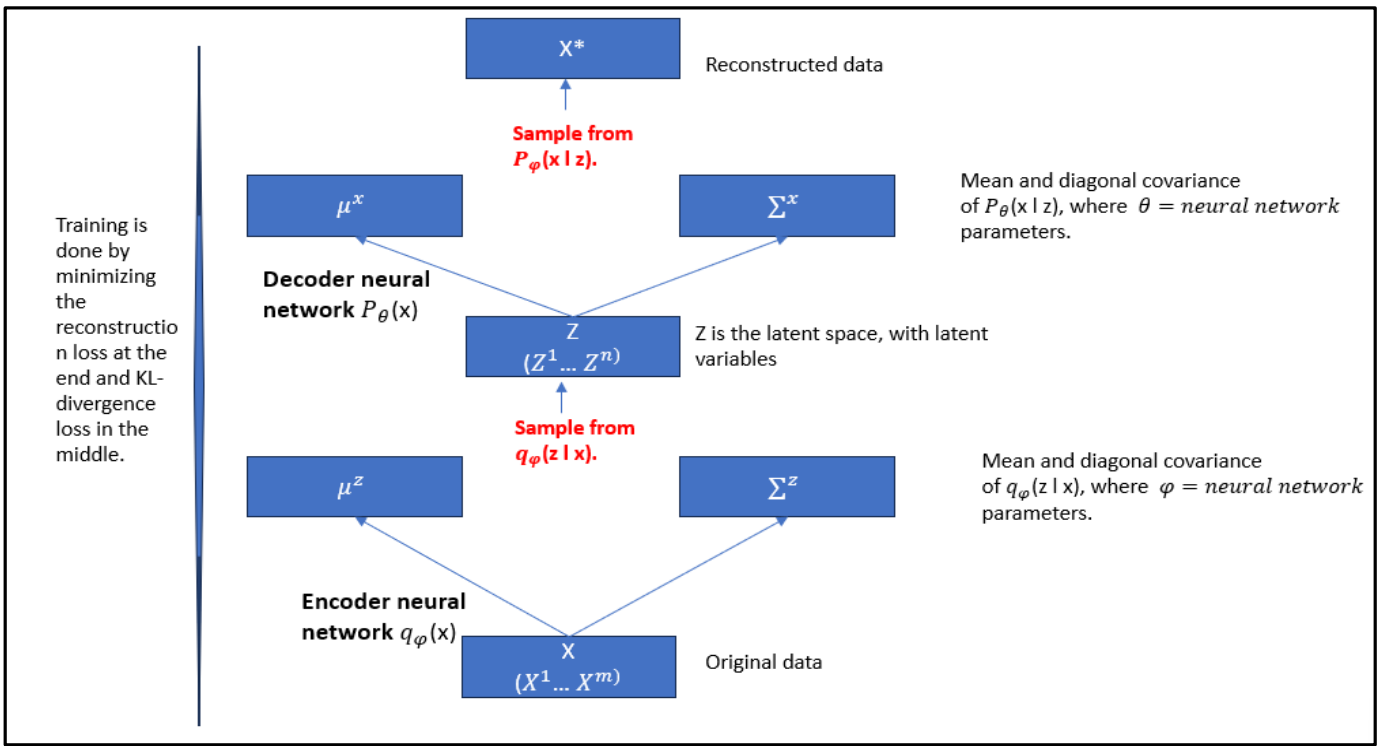

Fig 1 VAE Generative Modelling Process

Latent variables represent compressed representations of high-dimensional data in a continuous, lower-dimensional space. In image modelling, this means that a set of latent variables. $Z^1 \ldots Z^n$ encodes the essential structure of an input image $X^1 \ldots X^m$. where n < m. For example, a 28×28 image contains 784 observed pixel values, but its latent representation may consist of far fewer variables that capture the hidden features responsible for pixel dependencies. These latent variables are not directly observable but are inferred during training. Neighbouring pixels in an image exhibit strong spatial correlation, which can determine visual properties such as colour, shape, and layout. Latent space modelling aims to capture these spatial correlations and dependencies in a compact form suitable for generative synthesis. VAEs learn such representations by minimising a composite loss function that combines a reconstruction loss and a Kullback-Leibler (KL) divergence. This formulation encourages the model to generate outputs that are similar to the input while regularising the latent space to follow a known distribution. Although Generative Adversarial Networks (GANs) are widely used for data augmentation, they are difficult to train on small datasets due to discriminator overfitting and architectural complexity [11, 12]. For instance, CycleGAN employs 26 layers and 18 residual blocks, which increases the risk of mode collapse and training instability [13]. In contrast, CVAEs offer a more stable alternative for low-resolution image synthesis, as they optimise reconstruction loss directly and avoid the adversarial feedback loop that characterises GAN training. This adversarial loop, where the generator and discriminator compete, demands extensive and diverse datasets to maintain balanced learning dynamics; without sufficient data, the discriminator tends to overfit, destabilising the generator and leading to the production of irrelevant or incoherent images, as noted in Ref [14]. By removing this dependency, CVAEs maintain consistent training behaviour and are better suited for small, class-imbalanced datasets where architectural simplicity and convergence stability are critical. Here, a CVAE is proposed to generate class-conditioned outputs [15, 16] that align with the dataset characteristics of this study, namely the low-resolution constraint. Their standard neural architecture and compatibility with stochastic gradient descent make them computationally efficient and easier to deploy in constrained settings [17]. Prior research has demonstrated CVAE's effectiveness in digit and character recognition tasks [18, 19], but its application to low-resolution MICR was not explored in these works. This underexplored area shows the need for targeted investigation, positioning the current study as a relevant and timely contribution to MICR augmentation.

The primary focus of this work is to propose and critically evaluate a CVAE-based solution as a targeted data augmentation method for MICR under low-resolution constraints as low as 96 dpi. Unlike conventional augmentation techniques, CVAE enables structured latent space encoding, producing realistic synthetic character images that improve model generalisation and reduce

overfitting. An agent-based simulation framework is introduced to support dataset augmentation through modular role-specific agents. However, its integration with the CVAE is limited and not technically central to the generative process. The framework simulates realistic augmentation workflows but lacks programmatic control over latent-space modelling and image synthesis. Bayesian optimisation (BO) is applied to refine latent space configurations, aiming to minimise reconstruction loss and improve encoding efficiency. While this approach enhances CVAE performance, the optimisation process requires a more precise specification. Key elements such as the acquisition function, hyperparameter search space, and convergence criteria are discussed in Section 3 to enable reproducibility and validation. This paper is organised as follows: Section 2 reviews related literature, Section 3 details the proposed method, Section 4 presents experimental results, and Section 5 concludes with future directions.

II. RELATED WORK

MICR suffers from limited dataset availability, particularly in low-resolution imaging modalities. This constraint negatively impacts the generalisation of DL models and increases the risk of overfitting. Traditional augmentation techniques such as rotation, scaling, flipping, blurring, and image blending have been widely used to mitigate class imbalance and improve generalisation. However, these methods operate at the pixel level and do not introduce new semantic variations, rendering them inadequate for tasks such as MICR classification, where structural consistency and class-specific features are critical. Moreover, they often fail to capture the diversity needed in small datasets, leading to limited gains in model robustness and increased risk of overfitting to superficial transformations. Ref [20] demonstrated that multiscale convolutional neural networks (CNNs) combined with geometric augmentation can improve MICR performance. Similarly, Ref [21] and Ref [22] applied online and offline augmentation strategies to CNN and CRNN architectures, respectively. While these methods enhance robustness, they rely on deterministic transformations and offer limited diversity, which restricts generalizability across unseen medical imaging datasets. To address these limitations, generative models have gained traction for synthetic data augmentation. VAEs and GANs have shown promise in medical imaging tasks. While VAE-based approaches are frequently cited in the literature, their relevance to MICR-specific constraints is rarely examined in depth. For example, Ref [23] employed VAEs for feature learning in content-based medical image retrieval, focusing on global image descriptors rather than character-level synthesis. Similarly, Ref [24] applied a Vector Quantised VAE (VQ-VAE) to improve Gram-stain image classification, targeting texture-rich bacterial images rather than sparse alphanumeric characters. These studies demonstrate the versatility of VAEs in modelling complex image structures, but their domains and resolutions differ markedly from the structural and semantic demands of MICR. In parallel, GANs have gained popularity for image synthesis tasks due to their ability to produce visually compelling outputs. However, their reliance on large datasets and deep architectures makes them less suitable for MICR applications, which often involve low-resolution inputs and limited class diversity as noted in Ref [14]. This instability, coupled with the computational overhead of tuning deep GAN architectures, limits their practicality for character-level augmentation. Unlike the broader VAE applications cited earlier, CVAE-based methods directly address the challenges of character-level synthesis, thereby aligning more closely with MICR requirements. However, existing CVAE studies have not focused on MICR or low-resolution modalities, leaving a gap in domain-specific validation. This study addresses that gap by demonstrating how Bayesian-optimised CVAE architectures can be tailored to MICR constraints, offering an efficient semantically coherent augmentation strategy.

While generative models such as VAEs and CVAEs address the challenge of data scarcity through synthetic augmentation, they do not inherently model the contextual workflows in which medical image data is generated, processed, and interpreted. To complement these limitations, agent-based modelling has emerged as a promising approach for simulating healthcare environments and task-specific interactions. Ref [31] explored multi-agent systems for autonomous decision-making in clinical settings, demonstrating their potential for workflow automation and decision support. However, such frameworks are rarely integrated with generative augmentation pipelines. Most implementations lack structured mechanisms for character recognition or dataset expansion. This study addresses that gap by embedding CVAE-generated synthetic images within an agent-based simulation framework. The agents representing patients, clinicians, imaging devices, and recognition modules facilitate dynamic data retrieval and augmentation. However, the agent-based system serves a supporting role and does not directly influence the CVAE architecture.

Conclusively, though VAEs and CVAEs are frequently cited in the literature, their relevance to MICR varies significantly. Many VAE studies focus on global image features, whereas CVAE-based character synthesis offers a more targeted solution, especially for low-resolution medical images. Traditional augmentation methods lack the diversity needed for robust generalisation, and agent-based modelling remains underutilised in character recognition pipelines. This study addresses these gaps by combining CVAE-based augmentation with a structured simulation framework, offering a scalable, context-aware solution for MICR enhancement.

➢ *Contribution of this Study*

The core contributions of this study can be summarised as follows, highlighting the most novel and impactful elements of the proposed approach:

- The study applies Bayesian Optimisation to refine latent space configurations within a CVAE framework, offering a principled and empirically validated strategy

for tuning latent dimensionality in MICR-specific augmentation. While alternative automated methods exist, this approach demonstrates that a compact latent dimension yields optimal reconstruction fidelity and encoding efficiency under constrained, low-resolution imaging conditions.

- It adapts the CVAE architecture to address modality-specific constraints in MICR, including low resolution at 96 dpi, which has not been reported in existing literature to our knowledge, spatial noise, and limited sample availability, thereby extending prior work on general-purpose digit synthesis to a structurally constrained medical imaging domain.
- An agent-based simulation framework is integrated to contextualise the augmentation process, modelling real-world MICR workflows through autonomous agents. While it does not directly influence the CVAE's generative mechanics, it supports dynamic data retrieval and scalable dataset expansion.

III. PROPOSED WORK

This section outlines the integrated framework combining CVAE augmentation with an agent-based simulation system for MICR. The method is designed to address dataset scarcity, structural variability, and workflow realism in low-resolution medical imaging.

➢ *Agent-Based Simulation Framework*

The agent-based simulation framework models data flow and task delegation in MICR environments. It consists of four autonomous agents:

- Patient Agent: Represents individuals undergoing imaging and initiates data generation requests.
- Doctor Agent: Structures clinical imaging requirements and defines augmentation parameters.
- Imaging Device Agent: Simulates acquisition of raw medical images and metadata.
- Recognition Agent: Applies CVAE-generated synthetic data to train, evaluate MICR models and carry out character recognition.

Each agent within the simulation framework operates in a defined state space:

$$S = \{s_1, s_2, \ldots, s_n\}$$

Where augmentation actions A_i govern transitions between states. The agent's behaviour is modelled by the transition function:

$$P(s_{t+1} \mid s_t) = f(A_i, s_t)$$

To formalise the interaction between agents and the CVAE module, we define the probability of generating a synthetic image $\boldsymbol{X\_sy}$ from a real image $\boldsymbol{X\_re}$ using a Bayesian likelihood model:

$$P(X\_sy \mid X\_re) = \int P(X\_re \mid z, y) \cdot P(z \mid y)\, dz$$

Here, $\boldsymbol{z}$ represents the latent variable sampled from the CVAE's posterior distribution, and $\boldsymbol{y}$ is the class label provided by the agent. The agent's role is to supply contextual parameters, such as class labels, imaging conditions, or augmentation volume, that condition the CVAE's generative process. This interaction can be expressed as:

$$X\_sy = CVAE(z, y)$$
$$where\ z \sim q\_\varphi(z \mid X\_re, y), and\ y = Agent(s_t)$$

This formulation captures how agents influence the generation pipeline by dynamically assigning class labels and augmentation triggers based on their current state $\boldsymbol{s_t}$. Figure 2 illustrates this workflow, showing how autonomous agents initiate augmentation requests, define imaging parameters, and coordinate with the CVAE module to produce synthetic character samples for MICR training and post-MICR processes.

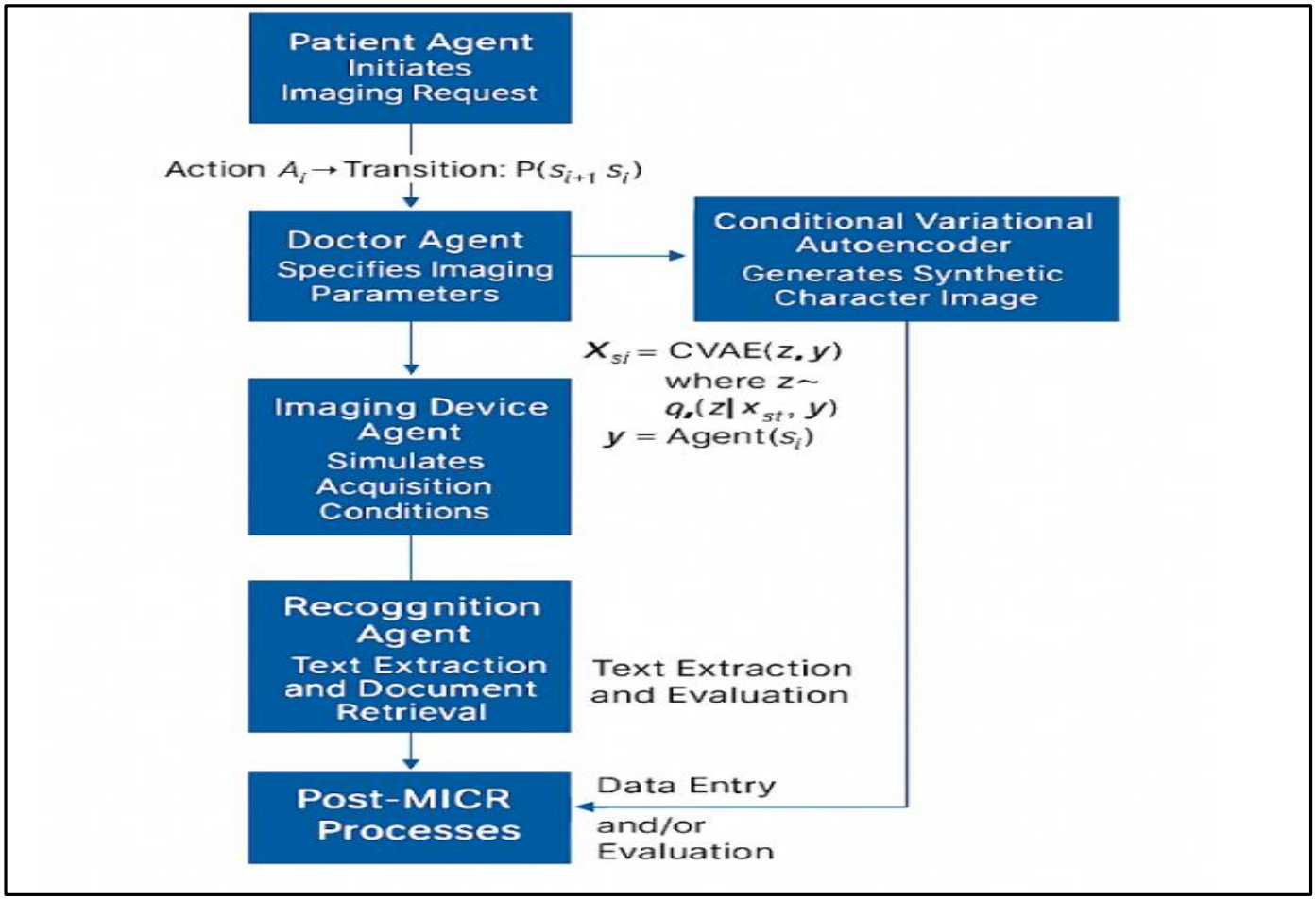

Fig 2 Agent-Based Simulation Workflow

Figure 2 presents a structured overview of the agent-based simulation framework designed to support synthetic data augmentation for MICR. The diagram illustrates a sequential flow of tasks initiated by the Patient Agent, who triggers an imaging request, and coordinated by the Doctor Agent, who defines imaging parameters and contextual labels. These parameters guide the CVAE, which generates synthetic character samples conditioned on latent variables and agent-supplied labels. The Imaging Device Agent simulates acquisition conditions, while the Recognition Agent receives the synthetic samples directly from the CVAE for MICR training and evaluation. This stage marks the transition from generative modelling to recognition, where extracted features are used to optimise MICR performance. The final stage, labelled Post-MICR Processes, includes downstream tasks such as data entry, validation, or integration into document retrieval systems in Electronic Health Record (EHR) systems. The CVAE module is centrally positioned to reflect its role as a bridge between raw image simulation and recognition. Figure 2 emphasises modular task delegation, explicitly distinguishing between generative, acquisition, recognition, and post-recognition phases. Each agent contributes uniquely to the augmentation pipeline, with clearly defined transitions and no architectural redundancy.

➢ *CVAE Architecture and Training Process*

Given the limitations of small medical datasets, we propose a CVAE as a generative modelling approach to synthesise realistic character images for medical text recognition in low-resolution medical imaging modalities of 96 dpi. The CVAE extends traditional VAEs by incorporating labels into the generative process, enabling class-conditional image generation. Let (X) represent an observed image and (y) its corresponding class label. The encoder network transforms the input (X) into a latent space representation (Z), governed by a normal distribution:

$$q\_\varphi(z \mid x, y) = N(\mu\_\varphi(x, y), \Sigma\_\varphi(x, y))$$

where μφ(x, y) is the mean and Σφ(x, y) represents the covariance matrix. The latent space sampling follows:

$$z \sim q\varphi(z \mid x, y)$$

The decoder reconstructs an output image conditioned on the latent variable z and label y:

$$p_\theta(x \mid z, y) = N(\mu\theta(z, y), \Sigma\theta(z, y))$$

Therefore, a CVAE provides controlled image generation, ensuring that synthetic images closely reflect the characteristics of the original dataset.

➢ *Bayesian Optimisation for Latent Space Tuning*

To train the CVAE, a composite loss function is minimised that balances reconstruction fidelity with latent-space regularisation. This objective integrates two components: the reconstruction loss, which ensures that generated images resemble the original input, and the KL

divergence, which encourages the latent space distribution to approximate a standard Gaussian prior.

The Reconstruction Loss ensures generated images match the original input, as given by:

$$E[q\varphi(z \mid x, y)] \, [-log \, p\theta(x \mid z, y)]$$

Where, $E[\cdot]$ denotes the expectation over the approximate posterior distribution

The KL divergence loss regularises the latent space by encouraging the posterior distribution to approximate a standard Gaussian prior, as given by: :

$$D\_KL \, [\, q\varphi(z \mid x, y) \, || \, p(z) \,]$$

Where $p(z) = N(0, I)$ represents the prior distribution.

This formulation ensures that the learned distribution remains close to the prior $p(z) = N(0, I)$, promoting smoothness and generalisation in the latent representation. The total loss function is computed as the sum of the reconstruction loss and the KL divergence loss, weighted by a beta coefficient. The beta coefficient modulates the trade-off between reconstruction accuracy and latent space regularisation. A higher value of the coefficient encourages disentanglement and smoother latent representations, while a lower value prioritises reconstruction fidelity. This formulation enables the CVAE to generate diverse yet semantically coherent synthetic character samples, which are subsequently used for MICR training and evaluation.

BO is mentioned in recent works on generative modelling [25], yet its application to latent space design for MICR remains underexplored and underdeveloped. Prior CVAE implementations often rely on fixed or heuristically selected latent dimensions, which limit their adaptability to domain-specific challenges. For instance, Ref [26] applied Convolutional VAEs to detect and eliminate eye blinks from EEG signals, a task focused on temporal noise suppression rather than spatial character reconstruction. While effective in that context, their approach did not incorporate latent space tuning or address the structural sparsity and resolution constraints inherent to MICR. In contrast, our study introduces BO as a principled mechanism for latent space refinement, using a structured search space (including latent size, dropout rate, and activation functions) and the Expected Improvement (EI) acquisition function to guide convergence. This represents a novel contribution, offering a reproducible, domain-aware alternative to the static configurations used in earlier VAE-based models. EI was chosen for its ability to balance exploration and exploitation in low-sample, noise-prone environments [27], making it particularly effective for MICR, where each model evaluation can be computationally expensive. Its probabilistic nature directs the search toward configurations with statistically meaningful gains, avoiding speculative tuning and premature convergence, while being computationally cheaper [28]. Optimisation trials were conducted over 20 iterations, with convergence assessed by stabilisation of the reconstruction loss and consistency across validation folds. The selection of z = 2 as the optimal latent dimension reflects empirical tuning rather than arbitrary choice, though statistical significance testing is still required to confirm robustness. Section 4 provides comparative metrics and clustering visualisations to support this claim.

Following BO, the CVAE architecture was finalised to balance reconstruction fidelity and generalisation for MICR augmentation. The optimal architectural configuration includes:

- *Encoder:*

A sequential stack of fully connected layers beginning with a 256-unit dense layer, followed by a 128-unit dense layer. Regularisation is applied via a dropout layer with a rate of 0.3, followed by batch normalisation and a ReLU activation function to introduce non-linearity.

- *Latent Space:*

The latent variable z is two-dimensional, with its optimal dimensionality determined through Expected Improvement-based Bayesian optimisation, ensuring efficient encoding and reconstruction fidelity.

- *Decoder:*

The decoder mirrors the encoder structure, starting with a 128-unit dense layer followed by a 256-unit dense layer. It includes a dropout layer (0.3), batch normalisation, and a final dense layer with the exact resolution as the image. A sigmoid activation function is applied at the output to constrain pixel values between 0 and 1.

This architecture was empirically validated across 20 optimisation trials, with convergence assessed via reconstruction loss stabilisation and validation consistency. The input comprises original character images and one-hot encoded labels, enabling class-conditional generation. Figure 3 illustrates the final architecture, highlighting the encoding-decoding flow and regularisation components. It visually outlines the encoder-decoder structure, highlights the selected latent dimension z = 2, and illustrates how dropout and batch normalisation are integrated to improve generalisation. This architecture reflects a data-driven configuration tailored to MICR augmentation, rather than a heuristic or static design.

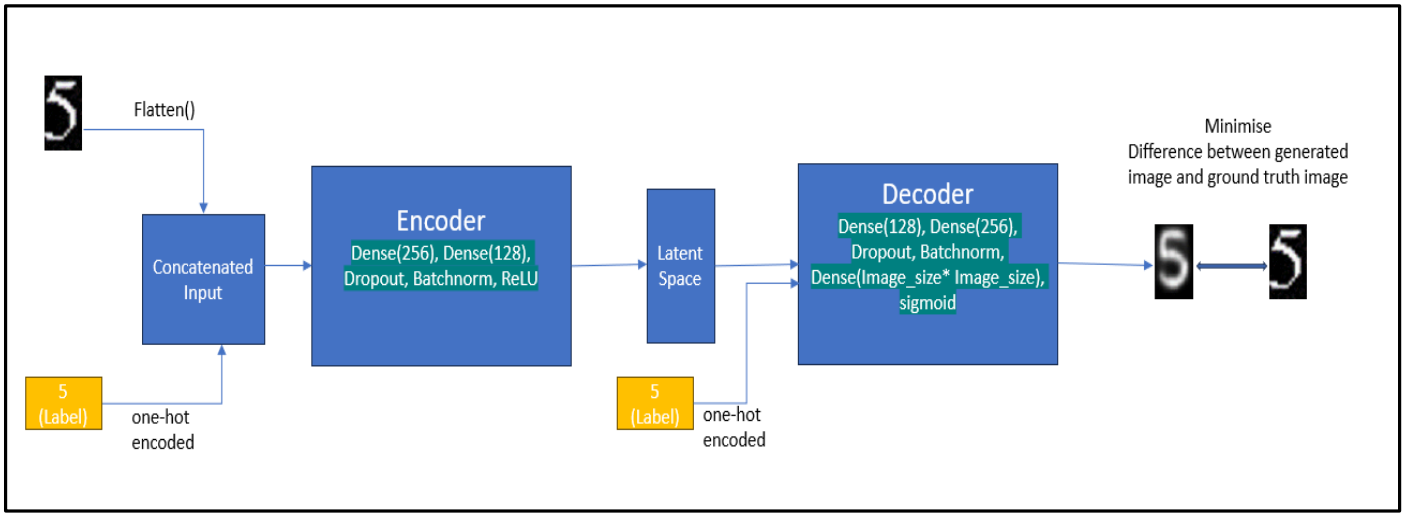

Fig 3 Final CVAE Architecture Determined Through Bayesian Optimisation

Figure 3 illustrates the final CVAE architecture selected via BO iterations, reflecting a configuration tailored to MICR's structural constraints. The encoder compresses the input, comprising a flattened character image and its one-hot encoded label, through two dense layers, followed by dropout (0.3), batch normalisation, and ReLU activation. These components were chosen to stabilise training and reduce overfitting [29], particularly in small medical datasets. The latent space dimension (z = 2), determined via EI-guided tuning, offers a compact representation that preserves essential character features while minimising reconstruction loss. The decoder mirrors the encoder's structure, culminating in a sigmoid output layer that reconstructs the image at the original resolution. This architecture balances expressiveness and generalisation, enabling the generation of structurally valid synthetic samples that enhance MICR performance.

Conclusively, the proposed method integrates agent-based simulation with a Bayesian-optimised CVAE architecture to address the challenges of MICR augmentation. By aligning architectural design with domain-specific and modality-specific constraints while validating through iterative optimisation, this study offers a promising solution for enhancing character recognition in low-resolution medical imaging. The following section presents experimental results that demonstrate the effectiveness of this approach across reconstruction quality, clustering consistency, and performance.

IV. RESULTS AND DISCUSSION

All experiments were conducted on a Google Compute Engine instance with 12.7 GB of system RAM, using Python 3, TensorFlow, and Keras. The study utilised two datasets: MEDPIX (a publicly available medical imaging dataset) and PRIVATEDT (a curated private dataset). Together, they provided a total of 5,126-character patches, comprising 3,050 samples from MEDPIX and 2,076 samples from PRIVATEDT, spanning 62 alphanumeric classes (A–Z, a–z, 0–9). Each image was standardised to dimensions (28 × 28 × 3) at 96 dpi resolution, reflecting typical constraints in low-resolution medical imaging. To ensure reproducibility and statistical validity, all reported metrics averaged over 20 independent experimental runs. Each run used a consistent 70:30 train–test split, allowing for controlled variation and reducing the influence of random initialisation and sampling bias.

➢ *Latent Variable Investigation*

To identify the optimal latent variable size for MICR augmentation, we conducted a series of controlled experiments to minimise reconstruction loss. The goal was to ensure that CVAE-generated synthetic images closely resemble their real counterparts in structure and intensity. Pixel normalisation was applied to standardise image brightness and contrast across samples, while conventional augmentation techniques were deliberately excluded to avoid bias during latent-space optimisation. The CVAE model was evaluated across a range of latent dimensions, and the minimum reconstruction loss (MRL) was recorded for both the MEDPIX and PRIVATEDT datasets. The results of these experiments are presented in Table 1 below.

Table 1 Latent Variables and Minimum Reconstruction Loss (MRL) for the CVAE Model

Latent Variables	MLR (Medpix)	95% CI (Medpix)	MRL (Privatedt)	95%CI (Privatedt)
2	27.03 ±0.02	[27.02, 27.04]	13.23 ±0.11	[13.18, 13.28]
3	28.22 ±0.03	[28.21, 28.23]	14.58 ±0.04	[14.56, 14.60]
4	28.78 ±0.04	[28.76, 28.80]	14.69 ±0.02	[14.68, 14.70]
5	28.65 ±0.02	[28.64, 28.66]	14.65 ±0.04	[14.63, 14.67]
6	28.72 ±0.03	[28.71, 28.73]	14.87 ±0.04	[14.85, 14.89]
7	28.86 ±0.14	[28.79, 28.93]	14.77 ±0.10	[14.72, 14.82]
8	29.62 ±0.39	[29.44, 29.80]	15.29 ±0.07	[15.26, 15.32]

The optimal latent variable size was determined to be z = 2, as larger configurations consistently led to increased reconstruction errors and diminished convergence efficiency due to excessive dimensionality [30]. Latent sizes beyond 8 were particularly unstable, often requiring longer training cycles with no measurable gain in reconstruction fidelity. Principal Component Analysis (PCA) confirmed that two latent variables captured 58.34% and 41.66% of the total variance, respectively, indicating that the tuned latent space was compact and sufficient to encode the structural complexity of MICR characters even at the low resolution of 96 dpi. To support this latent-variable size selection, 95% confidence intervals were calculated for each latent variable using the mean and standard deviation across 20 runs. For z = 2, the intervals were [27.02, 27.04] for MEDPIX and [13.18, 13.28] for PRIVATEDT, both of which were non-overlapping with those of higher latent sizes, providing strong descriptive evidence of its significance in reducing reconstruction error. Additionally, class-conditioned latent clustering shows that synthetic samples retained distinctive class features, supporting models' ability to encode small-sample-size classes while preserving the discriminative structure. The latent space distribution is visualised in Figure 4, showing clusters of character embeddings and maintaining semantic separation across classes. These findings align with prior work on dimensionality-constrained generative modelling in medical imaging [30].

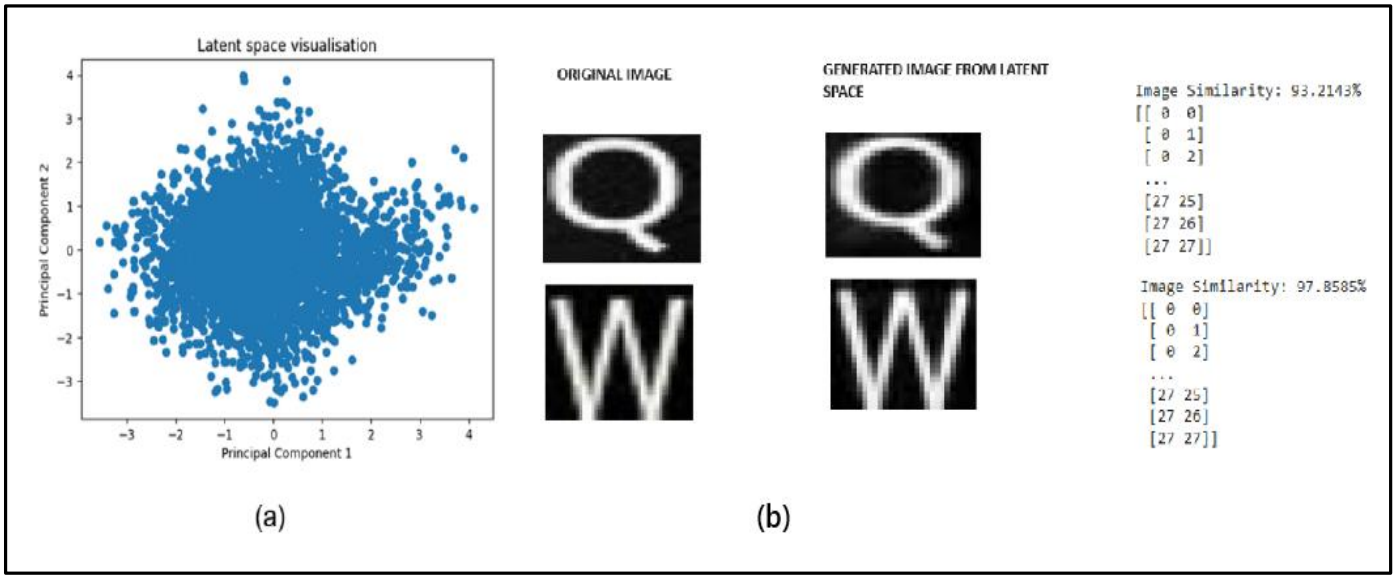

Fig 4 (a) Scatter Plot of Latent Space. (b)Similarity Assessment of Images

As seen in Figure 4, which provides a visual assessment of the latent space structure and image similarity, it offers insight into how different latent dimensions encode character features. Building on this, Figure 5 below presents the optimal latent-variable analysis, reinforcing the trends reported in Table 1. It demonstrates that a latent space of size z = 2 yields the lowest reconstruction loss across both MEDPIX and PRIVATEDT datasets.

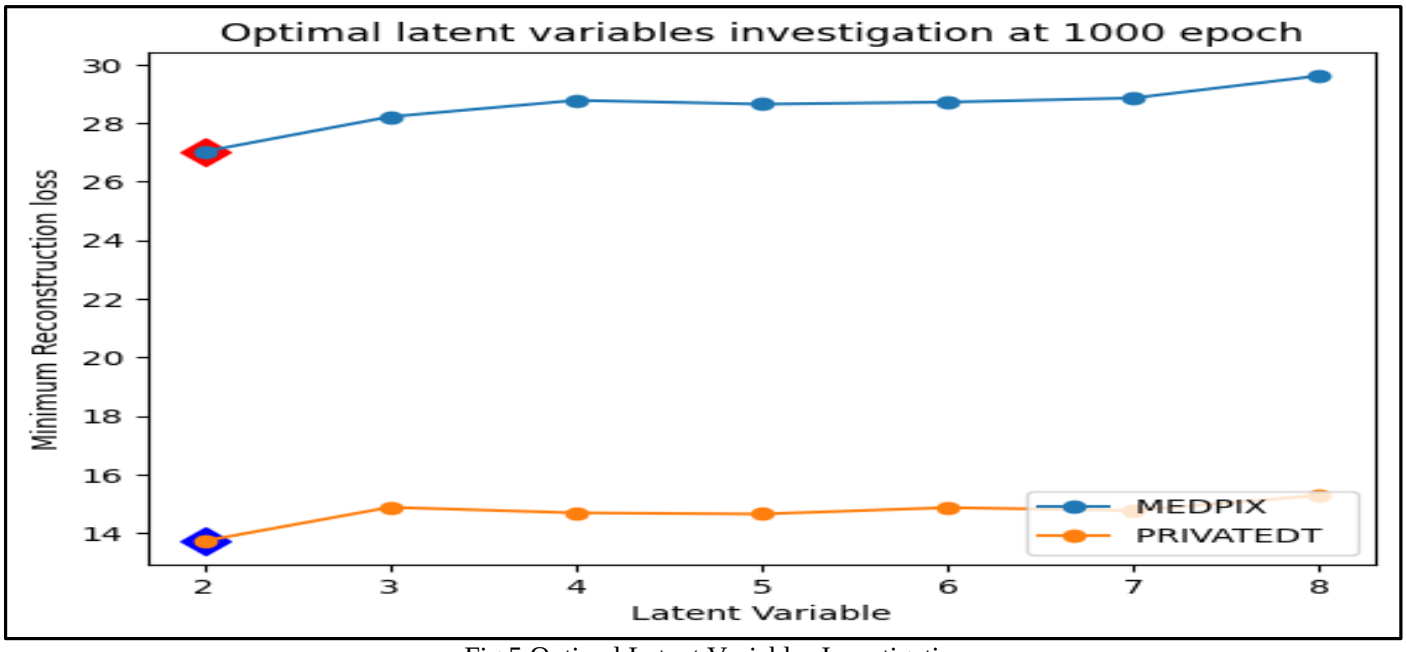

Fig 5 Optimal Latent Variables Investigation

By incorporating class labels during training, the CVAE achieves localised clustering in latent space while maintaining global packing, ensuring structural similarity between nearby encodings. This behaviour is evident in Figure 4a, where PCA-based visualisation reveals distinct groupings of character embeddings, suggesting semantic coherence across MICR classes. Although clustering metrics such as the silhouette score or Davies–Bouldin index were not computed in this study, visual inspection supports the CVAE model's ability to preserve class structure under constrained latent dimensionality. Future work may incorporate t-SNE and UMAP projections to validate non-linear separability further. Figure 4b complements this analysis by comparing original and generated images of "Q" and "W," with structural similarity scores of 93.21% and 97.86%, respectively. These results confirm the CVAE's ability to preserve essential image features, even for small-sample-size classes, aligning with the perceptual similarity framework proposed by Ref. [32] for assessing image quality. Together, the scatter plot and pixel-level evidence reinforce the model's capacity to encode and regenerate structurally valid MICR characters.

- *Quantitative Analysis - Augmenting Datasets with Synthetic Images*

To assess the impact of CVAE-generated synthetic images, we augmented the training datasets by adding N synthetic samples per class. Model performance was then evaluated using a standard CNN-based MICR classifier with basic hyperparameter tuning. As the classifier architecture is not the focus of this study, detailed configuration is omitted. The primary objective was to measure the relative performance gains attributable to the synthetic data augmentation. The results are presented in Table 2.

Table 2 Accuracy (%) of a CNN Classifier on Augmented Datasets Averaged on 20 runs.

	NUMBER OF SYNTHETIC IMAGES PER CLASS (N)			
	0	50	100	150
MEDPIX	87.13 ±0.18	90.33 ±0.12	90.63 ±0.10	88.92 ±0.02
PRIVATEDT	91.42+0.14%	93.83 ±0.02	98.27 ±0.06	93.02+0.06%

Table 2 shows that CVAE-based augmentation consistently improved CNN classification accuracy across both datasets. For MEDPIX, accuracy increased from 87.13% to 90.33%, 90.63%, and 88.92% when 50, 100, and 150 synthetic images per class were added, corresponding to gains of +3.2%, +3.5%, and +1.79%, respectively. Similarly, PRIVATEDT showed improvements from 91.42% to 93.83%, 98.27%, and 93.02%, yielding gains of +2.41%, +6.85%, and +1.60%. However, a decline in performance was observed beyond N = 100, indicating diminishing returns as synthetic samples began to outweigh original data. This trend underscores the importance of maintaining a balanced ratio of real to synthetic images to preserve data diversity and prevent overfitting to generated patterns. These results validate the effectiveness of CVAE augmentation while highlighting the need to carefully calibrate augmentation volume for constrained medical imaging tasks.

- *Comparison with Geometric Data Augmentation Methods*

To evaluate the effectiveness of different data augmentation strategies for MICR, comparative experiments were conducted using geometric transformations (GT), specifically random translation, scaling, and rotation, alongside the proposed method. The results are presented in Table 3 below.

Table 3 Accuracy (%) of a Quantitative Comparison Averaged on 20 runs.

(N = 100)	**GT**	**CVAE**
MEDPIX	87.82 ±0.13	90.58 ±0.18
PRIVATEDT	92.02 ±0.04	98.41 ±0.08

Results in Table 3 show that CVAE-based augmentation significantly outperformed GT methods across both datasets. When augmented with 100 synthetic images per class, character classification accuracy improved from $87.82\% \pm 0.13$ to $90.58\% \pm 0.18$ for MEDPIX and from $92.02\% \pm 0.04$ to $98.41\% \pm 0.06$ for PRIVATEDT. These gains confirm CVAE's better performing ability to preserve structural features and enhance dataset diversity, particularly under the 96-dpi low-resolution constraint. Unlike GT, which applies pixel-level distortions such as translation, scaling, and rotation, CVAE generates class-conditioned samples that retain semantic integrity and reflect the underlying data distribution. This is made possible by its compact latent space. Furthermore, GT do not introduce new semantic content and may degrade legibility at low resolutions, whereas CVAE augmentation maintains edge sharpness and spatial coherence. GAN-based augmentation was excluded from this comparison due to training instability, susceptibility to model collapse, and high computational overhead. This makes it unsuitable for small, class-imbalanced MICR datasets.

While the proposed model showed effective augmentation for MICR, its generative capacity is constrained by the simplicity of the underlying network and the low resolution of input data. The model employs dense layers without convolutional layers, which may restrict its ability to capture fine-grained spatial features. Additionally, although BO was used to tune latent space

parameters, the search space was limited to a small set of regularisation variables, leaving room for broader exploration and exploitation. The agent-based simulation framework, while helpful for contextualising augmentation workflows, remains decoupled from the generative process and does not directly influence latent-space modelling. Finally, the CNN classifier used for evaluation was a standard configuration with minimal tuning, selected to directly isolate the impact of synthetic augmentation rather than optimise classification performance. However, we prioritised interpretability and reproducibility, which may limit generalisability across more complex imaging modalities. Our future work will consider expanding the CVAE architecture to include convolutional layers, broadening the optimisation search space, and integrating agent feedback mechanisms to enable adaptive augmentation workflows.

Conclusively, the agent-based simulation framework served a supporting role in this study by structuring the augmentation workflow and modelling realistic MICR interactions. Its modular design enabled task delegation across autonomous agents, thereby contextualising the deployment of synthetic data. The central focus of this research remains the principled design and empirical validation of a Bayesian-optimised CVAE architecture tailored to low-resolution MICR augmentation. This emphasis on latent space refinement, structural fidelity, and augmentation efficiency defines the methodological core and primary contribution of the study.

V. CONCLUSION

This study demonstrates the effectiveness of Bayesian-optimised CVAE augmentation for MICR classification, integrated within an agent-based simulation framework to support text recognition and extraction in medical imaging workflows. By refining latent-space configurations via BO, the model achieved efficient synthetic data generation, improved reconstruction accuracy, and maintained dataset diversity. The agent-based simulation enabled dynamic interactions among autonomous agents, streamlining the augmentation pipeline and enabling adaptive control over data generation. The findings have broader implications for low-resource, low-resolution (up to 96 dpi) imaging environments, where conventional augmentation strategies often fail to capture class-specific variability. The structured latent space representation not only enhances model generalisation and reduces overfitting risk but also improves downstream OCR reliability, especially in scenarios with limited annotated data and low-resolution constraints. While the agent-based framework was not the primary methodological focus, its inclusion proved instrumental in automating and scaling the augmentation process. It offers a modular, extensible foundation for future research into intelligent data curation, adaptive sampling, and reinforcement-driven augmentation.

- *Author contribution*

E.O. conducted the experiment and wrote the manuscript. W.J. and N.H reviewed the manuscript and refined the concept.

- *Ethical Statement*

This study used both private and public MICR datasets. Private data were ethically approved by the University of Hertfordshire and anonymised prior to analysis. Public data were open-source and freely available. All procedures complied with institutional and international standards for ethical research and responsible data handling.

REFERENCES

[1]. Safonova, A., Ghazaryan, G., Stiller, S., Main-Knorn, M., Nendel, C., & Ryo, M. (2023). Ten deep learning techniques to address small data problems with remote sensing. International Journal of Applied Earth Observation and Geoinformation, 125, 103569. https://doi.org/10.1016/j.jag.2023.103569

[2]. Cromey, D.W. (2012) Digital Images Are Data: And Should Be Treated as Such. Methods in Molecular Biology, 1–27.

[3]. Ruthotto, L., Haber, E. (2021) An Introduction to Deep Generative Modeling.

[4]. Sabuncu, M.R., Yeo, B.T.T., Van Leemput, K., Fischl, B., Golland, P. (2010) A Generative Model for Image Segmentation Based on Label Fusion. IEEE Transactions on Medical Im-aging. 29(10), 1714–1729.

[5]. Bond-Taylor, S., Leach, A., Long, Y., Willcocks, C.G. (2022) Deep Generative Modelling: A Comparative Review of VAEs, GANs, Normalizing Flows, Energy-Based and Auto-regressive Models. IEEE Transactions on Pattern Analysis and Machine Intelligence. 44(11), 7327–7347.

[6]. Tang, S., Yang, Y. (2021) Why neural networks apply to scientific computing? Theoretical and Applied Mechanics Letters. 11(3), 100242.

[7]. DeVore, R., Hanin, B., Petrova, G. (2020) Neural Network Approximation.

[8]. Kingma, D.P., Welling, M. (2019) An Introduction to Variational Autoencoders. Founda-tions and Trends® in Machine Learning. 12(4), 307–392.

[9]. Liu, Y., Yang, Z., Yu, Z., Liu, Z., Liu, D., Lin, H., Li, M., Ma, S., Avdeev, M., Shi, S. (2023) Generative artificial intelligence and its applications in materials science: Current situation and future perspectives. Journal of Materiomics. 9(4), 798–816.

[10]. Doersch, Carl. (2016). Tutorial on Variational Autoencoders.

[11]. Sharma, P., Kumar, M., Sharma, H. K., & Biju, S. M. (2024). Generative adversarial networks (GANs): Introduction, Taxonomy, Variants, Limitations, and Applications. Multimedia Tools and Applications, 83(41), 88811–88858. https://doi.org/10.1007/s11042-024-18767-y

[12]. Li, D.-C., Chen, S.-C., Lin, Y.-S., & Huang, K.-C. (2021). A Generative Adversarial Network Structure for Learning with Small Numerical Data Sets. Applied Sciences, 11(22), 10823. https://doi.org/10.3390/app112210823
[13]. Goodfellow, I., Bengio, Y. and Courville, A. (2016) Deep Learning. MIT Press.
[14]. Kim, J., & Park, H. (2023). Limited Discriminator GAN using explainable AI model for overfitting problem. ICT Express, 9(2), 241–246. https://doi.org/10.1016/j.icte.2021.12.014
[15]. He, L. (2023) Comparison of improved variational autoencoder models for human face gen-eration. Journal of Physics: Conference Series. 2634(1), 012042.
[16]. Chen, S., Guo, W. (2023). Auto-Encoders in Deep Learning—A Review with New Perspectives. Mathematics. 11(8), 1777.
[17]. Pu, Y., Gan, Z., Henao, R., Li, C., Han, S., Carin, L. (2017) VAE Learning via Stein Variational Gradient Descent.
[18]. Rezende, D.J., Mohamed, S. and Wierstra, D. (2014). Stochastic backpropagation and approximate inference in deep generative models, in Proceedings of the 31st International Conference on Machine Learning - Volume 32. Beijing, China: JMLR.org (ICML'14), p. II-1278-II–1286.
[19]. Kingma, D.P., Rezende, D.J., Mohamed, S., Welling, M. (2014) Semi-Supervised Learning with Deep Generative Models.
[20]. Xu, X., Wang, W., Liu, Q. (2021) Medical Image Character Recognition Based on Multi-scale Neural Convolutional Network. 2021 International Conference on Security, Pattern Analysis, and Cybernetics（SPAC), 408–412.
[21]. Osagie, E., Ji, W., Helian, N. (2023) Ensemble Learning for Medical Image Character Recognition based on Enhanced Lenet-5. 2023 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB), 1–8.
[22]. Gifu, D. (2022). AI-backed OCR in Healthcare, Procedia Computer Science, 207, pp. 1134–1143. Available at: https://doi.org/10.1016/j.procs.2022.09.169.
[23]. Alves, C., Traina, A.J.M. (2022) Variational Autoencoders for Medical Image Retrieval. 2022 International Conference on INnovations in Intelligent SysTems and Applications (INISTA), 1–6.
[24]. Shwetha et al. (2024) 'Data augmentation for Gram-stain images based on Vector Quantized Variational AutoEncoder', Neurocomputing, 600, p. 128123. Available at: https://doi.org/10.1016/j.neucom.2024.128123.
[25]. Khater, T., Alkhatib, S. A., AlShehhi, A., Pitsalidis, C., Pappa, A. M., Ngo, S. T., Chan, V., & Truong, V. K. (2025). Generative artificial intelligence based models optimization towards molecule design enhancement. Journal of Cheminformatics, 17(1). https://doi.org/10.1186/s13321-025-01059-4
[26]. Criscuolo, S., Apicella, A., Prevete, R., & Longo, L. (2025). Interpreting the latent space of a Convolutional Variational Autoencoder for semi-automated eye blink artefact detection in EEG signals. Computer Standards & Interfaces, 92, 103897. https://doi.org/10.1016/j.csi.2024.103897
[27]. Di Fiore, F., Nardelli, M., & Mainini, L. (2024). Active Learning and Bayesian Optimisation: A Unified Perspective to Learn with a Goal. Archives of Computational Methods in Engineering, 31(5), 2985–3013. https://doi.org/10.1007/s11831-024-10064-z
[28]. Khondaker, R. M., Gow, S., Kanza, S., Frey, J. G., & Niranjan, M. (2022). Robustness under parameter and problem domain alterations of Bayesian optimisation methods for chemical reactions. Journal of Cheminformatics, 14(1). https://doi.org/10.1186/s13321-022-00641-4
[29]. Ruhland, J. B., Masoudian, I., & Heider, D. (2025). Enhancing deep neural network training through learnable adaptive normalisation. Knowledge-Based Systems, 326, 113968. https://doi.org/10.1016/j.knosys.2025.113968
[30]. Ji, Y. and Lu, Z. (2021). The Theoretical Breakthrough of Self-Supervised Learning: Variational Autoencoders and Their Application in Big Data Analysis, Journal of Physics: Conference Series, 1955 (1), pp. 012062.
[31]. Mehdizadeh, M., Nordfjaern, T. and Klöckner, C.A. (2022) A systematic review of the agent-based modelling/simulation paradigm in mobility transition, Technological Forecasting and Social Change, 184, p. 122011.
[32]. Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image Quality Assessment: From Error Visibility to Structural Similarity. In IEEE Transactions on Image Processing (Vol. 13, Issue 4, pp. 600–612). Institute of Electrical and Electronics Engineers (IEEE).