Cognitive Computation (2024) 16:2580–2595
https://doi.org/10.1007/s12559-024-10283-3

RESEARCH

# Generative Model-Driven Synthetic Training Image Generation: An Approach to Cognition in Railway Defect Detection

**Rahatara Ferdousi[1] · Chunsheng Yang[2,6] · M. Anwar Hossain[3] · Fedwa Laamarti[1,4] · M. Shamim Hossain[5] · Abdulmotaleb El Saddik[1,4]**

Received: 22 December 2023 / Accepted: 7 April 2024 / Published online: 17 May 2024


## Abstract

Recent advancements in cognitive computing, through the integration of artificial intelligence (AI) techniques, have facilitated the development of intelligent cognitive systems (ICS). This benefits railway defect detection by enabling ICS to emulate human-like analysis of defect patterns in image data. Although visual defect classification based on convolutional neural networks (CNN) has achieved decent performance, the scarcity of large datasets for railway defect detection remains a challenge. This scarcity stems from the infrequent nature of accidents that result in defective railway parts. Existing research efforts have addressed the challenge of data scarcity by exploring rule-based and generative data augmentation approaches. Among these approaches, variational autoencoder (VAE) models can generate realistic data without the need for extensive baseline datasets for noise modeling. This study proposes a VAE-based synthetic image generation technique for training railway defect classifiers. Our approach introduces a modified regularization strategy that combines weight decay with reconstruction loss. Using this method, we created a synthetic dataset for the Canadian Pacific Railway (CPR), consisting of 50 real samples across five classes. Remarkably, our method generated 500 synthetic samples, achieving a minimal reconstruction loss of 0.021. A visual transformer (ViT) model, fine-tuned using this synthetic CPR dataset, achieved high accuracy rates (98–99%) in classifying the five railway defect classes. This research presents an approach that addresses the data scarcity issue in railway defect detection, indicating a path toward enhancing the development of ICS in this field.

**Keywords** Generative AI · Synthetic data · Railway · Defect classification

## Introduction

In recent years, cognitive computing has advanced through the integration of artificial intelligence (AI), aiming to mimic human thought processes in a computational context [1, 2]. This has prompted the development of intelligent cognitive systems (ICS).

One potential application of ICS is in the detection and monitoring of railway defects, where the system can understand visual patterns of defective parts similarly to how a real-life quality inspector would. This process is illustrated in Fig. 1. In this figure, sensors (e.g., drone cameras) collect data akin to the tools used by quality inspectors. The generative AI component identifies and reproduces defect patterns, much like an inspector's initial assessment and hypothesis formation. The discriminative AI component then analyzes these patterns, similar to an inspector taking a closer look to verify their initial hypotheses. This focused analysis allows the model to determine the likelihood of actual

✉ Rahatara Ferdousi
rferd068@uottawa.ca

✉ M. Shamim Hossain
mshossain@ksu.edu.sa

1 The University of Ottawa, Ottawa, ON, Canada

2 National Research Council Canada, Ottawa, ON, Canada

3 School of Computing, Queen's University, Kingston, ON, Canada

4 Computer Vision Department, Mohamed Bin Zayed University of Artificial Intelligence, Abu Dhabi, United Arab Emirates

5 Department of Software Engineering, College of Computer and Information Sciences, King Saud University, 12372 Riyadh, Saudi Arabia

6 Institute of Artificial Intelligence, Guangzhou University, Guangzhou 510006, China

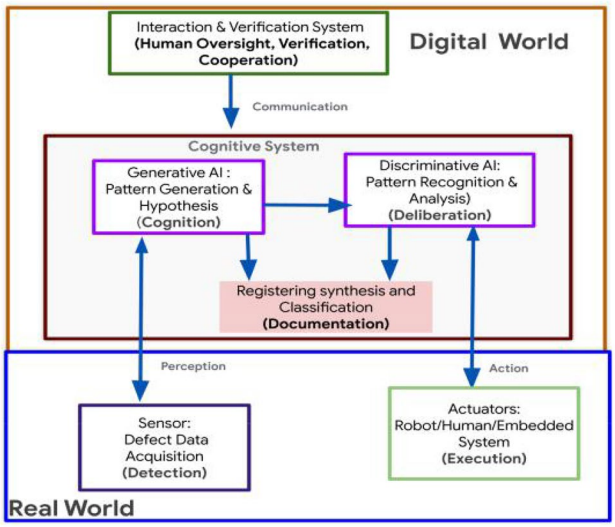


**Fig. 1** Basic scheme of an AI-integrated cognitive system for railway defect detection

defects, categorizing them with precision. Actuators, such as robotic or embedded systems, perform repairs by emulating an inspector's corrective actions. During the documentation phase, the types and locations of defects are cataloged, much like how a human inspector would document findings in a report. The entire process is overseen and verified by human experts, ensuring a collaborative and accurate system. This setup mimics the collaborative nature of inspection teams, ensuring that decisions generated by AI are validated by human expertise, thus maintaining the system's reliability and accuracy.

Despite the success of convolutional neural networks (CNN) in visual classification [3], CNN models and their variations often encounter generalizability issues in railway defect classification [4]. This is particularly true when the classification models are often trained and evaluated on an insufficient number of sample images [5]. One reason for incorporating limited samples in existing approaches is the scarcity of original defect samples or the challenge of collecting such samples [6]. This is attributed to the rare occurrence of accidents thanks to high-quality control and inspection measures in railway system [7, 8]. Moreover, the quality of images characterizing railway defects can deteriorate due to dynamically changing backgrounds and lighting conditions during data collecting [9]. This data scarcity issue ultimately limits the ability of ICS to "learn" from real-world examples to achieve human-like accuracy and decision-making capabilities [10].

Although pre-trained models in similar domains are commonly advocated in the literature to address data scarcity issues, such an option may not always be effective for specialized cases like railway defect detection [11]. Specifically, a model fine-tuned in a relevant domain (such as image recognition in industrial settings) may perform well in identifying common defects, like cracks, but could struggle with rare or specific defect types, such as those affecting brake shoes, which occur in less common objects (e.g., Canadian freight trains). Recent studies have proposed synthetic data generation as an approach to tackle the data scarcity issue for training classifiers. Generally, there are two main approaches: (i) fixed-rule-based [12] and (ii) data-driven models [13, 14]. Traditional fixed rule-based data augmentation methods (e.g., zooming, rotation) may not adequately capture the intricate patterns and variations present in actual defect images [15]. Generative adversarial network (GAN) and variational auto encoder (VAE) models have been rapidly proposed for synthetic data generation across various domains. However, synthetic data generation is not a magical solution. GAN are sensitive to data distribution and require a significant amount of data for effective noise modeling [16]. In contrast, VAE can still effectively model the statistical properties and patterns of the available limited samples within the latent space [17]. Nonetheless, if a VAE model learns from a limited sample size [18], it may face constraints that lead to the reconstruction of unrealistic or erroneous samples [19].

Therefore, in this research, we aim to produce high-fidelity synthetic images of railway defects for precise defect detection from a limited number of samples. This approach seeks to reduce overfitting and achieve generalizability in multi-class defect recognition, even within imbalanced datasets.

To address the stated problem, we propose a VAE-based synthetic data generation approach. The model processes defective images as input, encodes high-level features (e.g., edge, shape, intensity), learns the latent distribution of the defects, and ultimately reconstructs synthetic samples from the learned patterns. To mitigate the issue of erroneous sample generation, we have updated the regularization technique of the traditional VAE model by incorporating weight decay into the reconstruction training loss. Moreover, we designed an algorithm (Algorithm 1) for synthetic defect generation, utilizing a VAE to produce synthetic images from a dataset of original defect images. This algorithm methodically transforms original images into a specified number of synthetic variants for each defect class, thereby enhancing the diversity and volume of data available for training.

Our experimental setup utilized the main CPR dataset [6], comprising 50 samples across three normal classes and two defect classes. Initially, we evaluated the VAE model's ability to reconstruct images and tested its overall loss. Subsequently, we assessed the performance of a fine-tuned vision transformer-based defect classifier, trained using the CPR synthetic dataset. This dataset includes 500 images, of which 450 are reconstructed images generated from the 50 original images of the main CPR dataset. We

discovered that our modified regularization technique effectively reduces loss when compared to the traditional VAE model's regularization technique. Additionally, the proposed VAE approach demonstrated superior qualitative results by generating more realistic images, improving the distribution of the latent space, and addressing issues related to incorrect image reconstruction due to overfitting. In the second phase of our evaluation, we achieved high scores ranging from 98 to 99% in test accuracy, training accuracy, F-measure, precision, and recall. Our findings indicate that the VAE model can significantly contribute to overcoming the data collection challenge for rarely occurring samples, such as defects.

The rest of the paper is organized as follows. In "Literature Review" section we outline the existing approach for synthetic data generation to identify the gaps and requirements for the existing approach. "Method" sect. details the proposed VAE-based approach and synthetic data generation algorithm incorporating the VAE. In "Evaluation" sect. we present the data, experiments, and results. We discuss the impact of the outcome of this study in "Discussion" sect. and the limitations and viable future work in "Limitations and Future Work" sect. Finally, we conclude this study in "Conclusion" sect.

## Literature Review

The challenge of data scarcity in training classifiers is a significant hurdle in automating defect inspection for large-scale transportation systems. Synthetic image generation has emerged as a popular solution in the industry for addressing this issue [16]. We have reviewed the existing literature and found that researchers have extensively explored various approaches to handle this challenge, which we outline in the following sections.

### Fixed-Rule Based Approach

The fixed-ruled-based data augmentation typically involves geometric transformations (e.g., zooming, rotation, cropping, scaling) or basic image processing (e.g., color contrasting) [20]. However, the generated defects from a fixed-rule-based approach often lack randomness to mimic real defects [14, 20]. Slight variations in existing defects may not be fully captured or adequately described by the preset rules, leading to degraded quality of the generated defects and limitations in providing appropriate synthetic data for the defect classifier [14].

### Data-Driven Approach

Data-driven synthetic defect generation methods that learn directly from existing data are another approach [15]. From the literature on defect classification, we identified two candidate data-driven approaches: i) Generative Adversarial Networks (GAN) [21] and ii) Variational Autoencoders (VAE) [22]. We conducted a comparison between GAN and VAE to gain an understanding of the effectiveness and suitability of the GAN and VAE approaches specifically for synthetic data generation for railway defects.

Table 1 contrasts the two data-driven generative models, making it clear that VAEs offer a higher degree of control and flexibility, especially for cases where data variation might be limited, and where preservation of semantic information in generated data is crucial.

## Related Work

The use of GenAI for synthetic data generation is a novel approach to address the issue of limited datasets. In this section, we provide a summary of the few existing studies that are related to this topic. The application and methods in these works are tabulated in Table 2

The authors in [12] propose a framework for data augmentation by creating synthetic images using GAN. The synthetic images are used for training classification algorithms, improving the performance of CNN for the classification of surface defects. This approach is similar to the one used in [13] and [21], but it specifically focuses on surface defects. In the context of railway defect detection, this approach could be particularly useful as surface defects are the most common type of railway defect.

The study in [21] presents Defect Transfer GAN (DTGAN), a framework that can learn to depict different types of defects using different backdrop products. It can also

**Table 1** Comparative table: GAN vs. VAE for defect data generation [21, 22]

| Subject | GAN | VAE |
|---|---|---|
| Latent Space Control | No Controlling of the latent space | Has control over the latent space. |
| Noise Modeling | Complex and Data Sensitive | Not Required. |
| Variation | Not suitable for limited varia tion | Can learn latent distribution from limited variation |
| Semantic Rep-resentations of generated data | Not preserved | Preserved and modifiable |

**Table 2** Comparison of papers on synthetic data generation

| Paper | Method Used | Application |
|---|---|---|
| [12] | GAN | Surface defect detection |
| [?] | DT-GAN | Defect synthesis |
| [13] | Defect-GAN | Defect inspection |
| [14] | VAEs | Compound property prediction |
| [23] | Generative AI models | Healthcare |
| [?] | reviewed Various methods (including decision trees, deep learning techniques, and iterative proportional fitting) | Various applications |

apply styles particular to defects to create realistic representations of defects. When compared to cutting-edge picture synthesis approaches, it generates more diverse defects and has higher sample fidelity. This study, along with the study in [13], provides a more specialized approach to defect synthesis compared to the general synthetic data augmentation approach presented in [12].

The authors in Defect-GAN[13] present an automated defect synthesis network that generates realistic and diverse defect samples for training accurate and robust defect inspection networks. It shows that Defect-GAN can synthesize various defects with diversity and fidelity.

The authors [14] suggest a technique to enhance the accuracy of machine learning models in predicting chemical properties. This is achieved by integrating supplementary data on associated molecular descriptors into the learned representations of variational autoencoders. This work stands out as it focuses on improving the performance of VAEs, which is a different approach compared to the GAN-based methods used in the other papers. Although the application domain is not relevant to defect data generation, the findings of the paper helped to understand the role of latent-space-based VAE data generation for limited samples.

Similarly, the study in [23], examines generative AI models for creating realistic, anonymized patient data for research and training. It explores synthetic data generation strategies for a use case where data collection is challenging due to privacy concerns.

Another comprehensive systematic review in [16] examines existing studies that employ machine learning models for generating synthetic data. It explores different machine learning methods, with particular emphasis on neural network approach and deep generative models, and addresses privacy and fairness concerns related to synthetic data generation.

Although the existing literature provides valuable insights into synthetic data generation using GAN and VAEs, there remains a distinct gap in applying these methods specifically to the challenges of railway defect classification, especially in the context of limited defect data, which we have outlined in the following section.

## Gaps and Requirements

Most existing work has well-documented the generation of synthetic images. However, there is an apparent gap in terms of evaluating and implementing these techniques within a complete pipeline, specifically for railway defect classification. A methodology that encompasses data generation, model training, and defect classification, tested in a real-world railway setting, is yet to be explored.

After reviewing recent research that proposed GAN and VAEs for synthetic data generation, we found that GAN are limited to scenarios when the noise in the data is not properly modeled with small or less diverse training datasets [?]. In such cases, GAN face challenges in accurately capturing the underlying data distribution due to an incomplete representation of noise patterns, leading to model collapse [21]. VAEs address the challenge of data scarcity in machine learning through a combination of efficient latent space representation and regularization, enhanced by their generative capabilities [14]. Unlike traditional autoencoders, which aim to minimize reconstruction error, VAEs introduce a probabilistic graphical model approach. They enforce a distribution (usually Gaussian) over the latent space, ensuring that similar inputs result in close points in this space, thus facilitating smoother interpolation and the generation of new data points [18].

For example, consider a dataset of images depicting freight body parts with subtle defects. Such datasets are often small due to the rarity of defects and the high cost of data collection. A VAE trained on this dataset would not only learn to compress and reconstruct these images but also understand the underlying distribution of features that characterize both normal and defective equipment. By sampling from the latent space, the VAE can generate new images that realistically vary along dimensions that matter for identifying and analyzing defects, such as the size, shape, and location of anomalies. This is required for augmenting the dataset, providing a richer set of examples for training more robust defect detection models without the need for extensive real-world data collection [24].

In contrast, while GAN are capable of generating high-fidelity images, they may struggle with training stability and

mode collapse, especially in data-scarce scenarios, potentially making them less reliable for augmenting datasets with realistic variations of rare defects [21]. Traditional autoencoders might efficiently compress and reconstruct data; however, they lack the structured latent space and generative properties that enable the creation of a semantic distribution of the images. Because of this, they struggle to create new and varied examples of defects.

Therefore, we opted for VAEs due to their ability to learn a latent space from limited variation and fewer samples [22]. However, when a model learns from fewer samples, there is a risk of overfitting issues [25]. Generally, reconstruction loss is measured throughout the training process of VAEs to keep the difference between the original and reconstructed image low [16].

Nevertheless, if any model parameters are abnormally prioritized by the model, it will cause an overfitting issue, resulting in a wrong or low-quality reconstructed image, which is a concern when creating training data for the defect classifier [25]. The overfitting of the VAE model needs to be regularized while training the model.

Therefore, in this paper, we propose a VAE-based synthetic data generation method with a modified regularization mechanism to address these issues. In the following sections, we detail the methodology for synthetic data generation, model training, and defect classification. In addition, we rigorously test our approach with real railway defect datasets to address the identified research gap.

## Method

In this section, we present the proposed VAE-based method for synthetic railway defect data generation.

Overall, the input layer takes the original images as input, then the encoder CNN captures its hidden representation, such as shape, texture, and other high-level features. After that, this encoding is used to learn a latent space. From this latent space, a pattern or distribution is learned by the decoding neural network. Finally, the output is pre-processed to reconstruct an image of the original image. Throughout this process, a regularization technique is applied to minimize the reconstruction loss and penalize large weights, mitigating the risk of overfitting. In Fig. 2, we illustrate the proposed VAE approach for railway defect image reconstruction. The detail of the proposed approach is presented in the following sections.

### Encoder

In the context of a Variational Autoencoder (VAE), the hidden representation (h0) refers to the output of the encoder network after processing the input defect images. It captures the compressed representation of the input data (e.g., texture, shape, pattern, and edge). The hidden layers of the encoder CNN transform the hidden representation (h0) to obtain the parameters that define the probability distribution of the latent space. Let us consider that there are three different types of defects—bottom defect, side defect, and top defect.

The VAE encoder-

- first, divides these input images into three regions, each represented by a different color in Fig. 3.
- then processes each region separately and extracts features using convolutional layers.
- after that, it maps the extracted features of each region into a lower-dimensional vector (latent space).
- then the encoder generates multiple latent space vectors for each input image, allowing for probabilistic encoding. ?? The variation in latent space allows the VAE to learn a meaningful and continuous latent space for better defect representation and image generation.

### Latent Space

The latent space captures the essence of the input image. In the latent space, images that are similar or share common characteristics are expected to be closer to each other. Conversely, images that are dissimilar or have distinct features

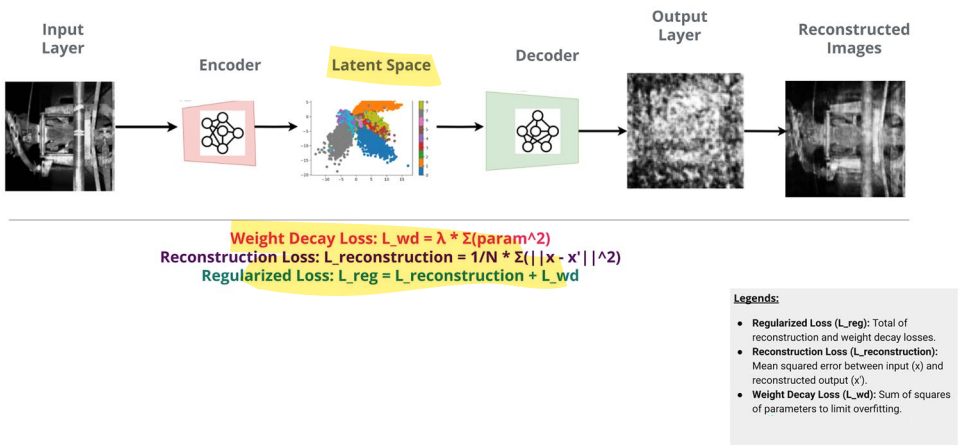


**Fig. 2** Proposed VAE approach for railway defect image reconstruction

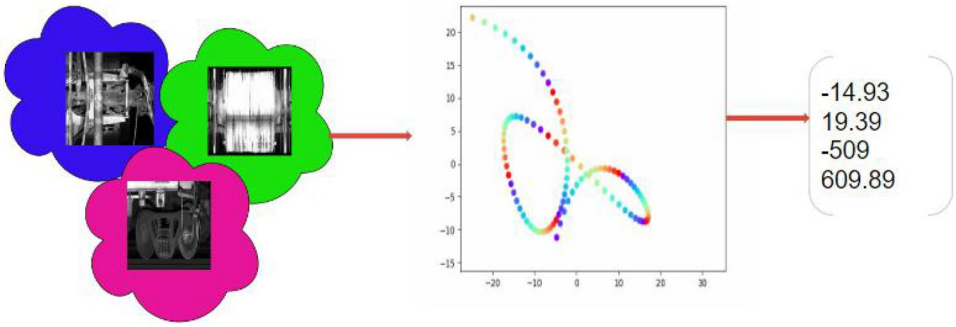


**Fig. 3** Illustraion of encoding process for railway defect generation

are expected to be farther apart. This property of latent space ensures that images with the same defect type are encoded close to each other in the latent space.

As demonstrated in Fig. 4, neighboring points in the latent space correspond to similar representations or characteristics in the generated data. By exploring different points in the latent space, such as sampling along a straight line or within a region, synthetic samples can be generated through gradual transitions or variations in their features. This property enables the VAE to perform smooth interpolation, allowing for the generation of new data points with controlled changes or combinations of attributes.

## Decoder

The Decoder of VAE maps the latent space back to the data space. Unlike the encoder, the decoder takes the latent vector (z), maps it to the hidden representation (h1) and then generates the reconstructed defect image. To break it down, the decoder network (g dec).

- takes the sampled latent vector (z) as input.
- then it maps the sampled latent vector (z) to a hidden representation (h1). Here h1, serves as a compressed and abstract representation of the input data, capturing important features such as texture, shape, pattern, and edges.
- after that, the decoder network uses several hidden layers to transform the hidden representation (h1) to generate the learned pattern as demonstrated in Fig. 5

Finally The reconstructed output ($x'$) is further processed by the output layer function (f out) to obtain the final output (y), which is the reconstructed Images in desired representation.

## Regularization Technique

In general, a VAE-based method tries to keep the reconstruction loss low while learning. In our case, we propose adding weight decay and reconstruction loss to regularize the training. The loss to activate the regularization are computed as given in the following equations. ??

- Reconstruction Loss:

$$L_{reconstruction} = \frac{1}{N} \| X_{train} - X_{train}' \|^2$$

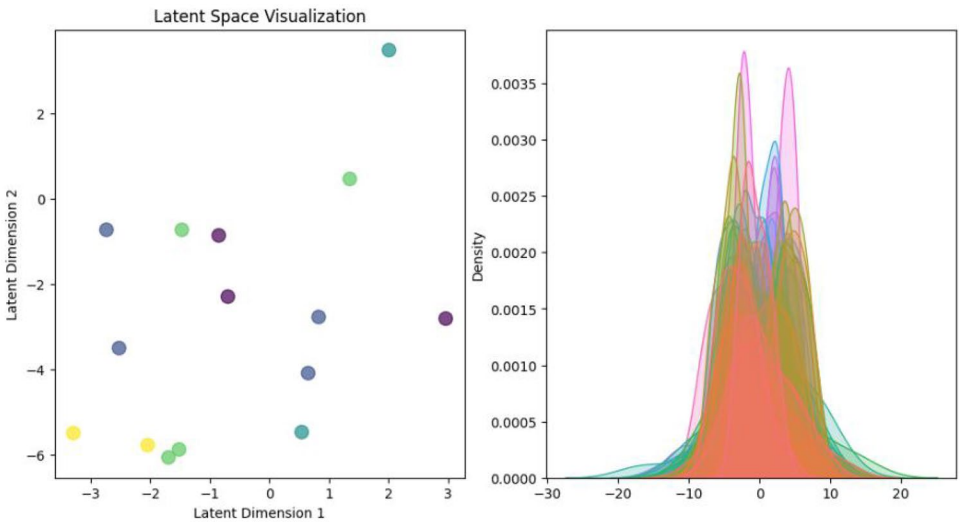


**Fig. 4** Example of a learned distribution of the latent space: On the left side, the same color represents samples of the same class, indicating that similar data points are clustered together in the latent space. The probability distribution on the right side shows the density and spread of the learned representations for different defect classes

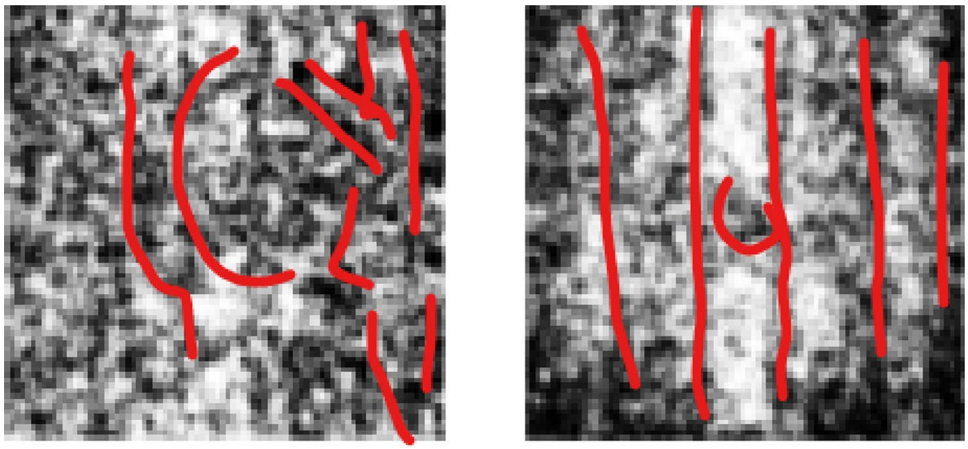

**Fig. 5** Example out of a decoder learned patterns (e.g., shapes) of a defective part. The red lines are to highlight the pattern in these images

- Measures the discrepancy between the input data (x) and the reconstructed output ($x^{'}$).
- A VAE-based method aims to minimize this loss during training to improve the quality of the reconstructions.

- Weight Decay Loss:

$$L_{wd} = \lambda \, param^2$$

- Penalizes large parameter values to control model complexity and avoid overfitting.
- Example: If $\lambda = 0.001$ and a parameter value is 2, the weight decay loss contribution for that parameter is $0.001 \times 2^2 = 0.004$.

- Combined Regularization:

$$\text{Total Loss} = L_{reconstruction} + L_{wd}$$

- We propose to add weight decay ($L_{wd}$) along with the reconstruction loss to regulate the training process.
- By adjusting the regularization weight $\lambda$, we can control the relative importance of the reconstruction term and overfitting prevention.

The combination of the reconstruction loss and weight decay in the regularization process can mitigate the trade-off between model complexity and generalization capability. In scenarios with limited training data, the proposed technique allows the VAE to effectively learn the underlying probability distribution from the available samples, avoiding the risk of overfitting, and enhancing its capacity to generalize to new data.

## VAE-Based Synthetic Data Generation Algorithm

The proposed Algorithm 1, employs a trained VAE model to reconstruct images of defects. The algorithm takes original images, the desired number of reconstructed images for each original image, and the total number of defect classes as inputs. Then it reconstructs the specified quantity of synthetic images for each class. Further details of this algorithm are given in "Input specification", "Image preprocessing", "Reconstruction" and "Computational complexity" sections.

**Algorithm 1** VAE-based synthetic image generator

Require: dataset path: Path to the original dataset
Require: defect classes: List of defect classes
Require: reconstructed path: Path to save the reconstructed and original images
Require: vae model: Variational Autoencoder model
Require: num images per sample: Number of synthetic images per original sample
1: Create a directory at reconstructed path to save the reconstructed and original images
2: Set vae model to evaluation mode
3: Define image transformation steps: grayscale conversion, resizing, and conversion to tensor
4: for each defect class in defect classes do
5: Set class path as dataset path/defect class
6: Create a directory at reconstructed path/defect class + " reconstructed" to store the reconstructed images
7: Get the list of image files inclass path
8: Shuffle the image files randomly
9: for each image in the shuffled image files do
10: Load and pre-process the image
11: Reshape the image to match the expected input shape of the VAE model
12: Reconstruct the image using the VAE model
13: Save the reconstructed image and the original image in the reconstructed directory
14: end for
15: end for
16: return Reconstructed and original images saved in the reconstructed path directory

#### Input Specification

The algorithm accepts original defect images, the desired number of reconstructions for each image, and the total number of defect classes. It systematically processes each class to enhance the dataset with synthetically generated, yet realistic, images. Then directories are created to save the reconstructed alongside the original images, ensuring organized storage and easy retrieval for subsequent processes like training and validation.

#### Image Preprocessing

Prior to reconstruction, images undergo a series of transformation steps. These include grayscale conversion, resizing to match VAE input requirements, and tensor conversion, which is essential for maintaining image quality and ensuring compatibility with the VAE's architecture. Such a pre-processing stage also supports the ready-to-use dataset feature in the generated dataset.

#### Reconstruction

The core of the algorithm is the reconstruction loop. For each class, the original images are processed through the VAE to produce the synthetic counterparts. Both original and reconstructed images are then saved, building a comprehensive dataset that encompasses a wide spectrum of defect variations. To avoid bias and pattern repetition, a randomization step is integrated. This ensures that the synthetic images produced do not simply replicate the input but instead introduce variability, reflecting a realistic distribution of railway defects. The VAE model is set to evaluation mode, which optimizes its performance for reconstruction tasks. In this mode, the encoder part of the VAE approximates the posterior distribution and the decoder reconstructs the defect images from sampled latent codes.

#### Computational Complexity

The variation in defect classes, represented as a set C is leveraged by the algorithm so that each defect class $\in$C contributes to a diverse set of images. By iterating over each defect class the proposed algorithm ensures variability in the defect characteristics of generated images.

In addition, it includes a randomization step to avoid repeating the same patterns during the reconstruction process. The Algorithm 1 activates the VAE in evaluation mode to allow its encoder to approximate the posterior distribution $q_\phi(z|x')$. Latent codes are sampled from this distribution, creating a compact representation: $z \sim q_\phi(z|x')$.

The complexity of the algorithm for synthetic image generation is influenced by several factors including the number of defect classes (|C|), the total number of original images across all classes (N), the VAE's latent dimension (d), and the number of synthetic images generated per original image (num images per sample). Moreover, a reconstruction process involves both encoding and decoding with time complexity of O(d). For an entire dataset, the total time complexity is calculated as:

$$O(|C| \cdot N \cdot \overline{\text{num}}\ \overline{\text{images}}\ \overline{\text{per}}\ \text{sample} \cdot d) \quad (1)$$

For a large number of defect classes, the latent spaces can be high-dimensional. Consequently, the memory, model approach, and hardware resources may impact the scalability of the algorithm. Therefore, optimizations such as parallel processing and GPU acceleration could be considered to enhance the algorithm's scalability and performance for using the proposed algorithm for large-scale datasets.

## Evaluation

This section details the dataset, system configuration, evaluation metric, and qualitative and quantitative results of the synthetic data generation approach. In addition, we present a further evaluation of a ViT defect classifier [26] trained on the generated synthetic dataset.

### Data

The original dataset used in this study is the CPR main dataset. In Fig. 6 we demonstrate sample images from each of 5 defect classes representing the visual characteristics of the defects. It is spectacular from this illustration these images are hard to characterize manually by textures or color. However, the shape of the patterns is distinguishable.

Moreover, the bar diagram demonstrates that the sample distribution per class was not reasonable for training a CNN model. The distribution of samples per class in Fig. 7 demonstrates that it is infeasible to train a multi-class defect classifier, even with a transfer learning model. Because after splitting the dataset for training and testing there were significantly low samples (6–8 images in each class).

abit sus

### Evaluation Metrics

We employed a combination of various loss and accuracy measures, as presented in Table 3, to evaluate the quality of the generated outputs by the proposed VAE-based approach, as well as the classification accuracy of the ViT model.

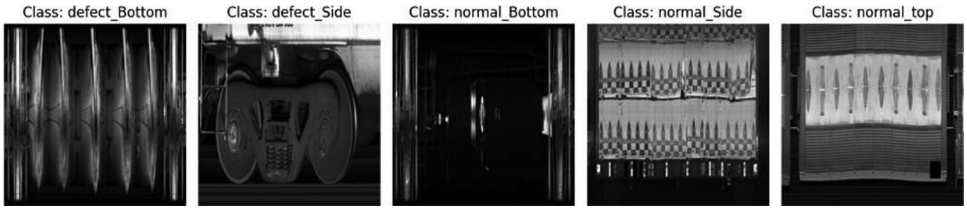


**Fig. 6** Example CPR original image samples for different classes

The selected loss functions and performance measures facilitate specific purposes in evaluating the VAE generator and ViT model. The Reconstruction Loss is used to measure the similarity between the original and reconstructed images by the generative models, while the KLD demonstrates the ability of the VAE model to capture underlying data distributions. We considered widely used performance measures, including Precision, Recall, and F1-Score, for assessing the performance of the ViT defect classifier in terms of false positives and false negative scores for each class.

## Evaluation of VAE Model

In for the proposed approach. First, we compare the performance between the traditional and proposed approach. Second, we compare the qualitative output of the proposed VAE approach with two other contemporary synthetic data generation approaches.

### Comparison Between Traditional and Proposed VAE Model

We utilized several aspects to compare the traditional and proposed VAE models, including latent space analysis, training loss, quality of generated output, and test loss. These are elaborated upon in the following sections.

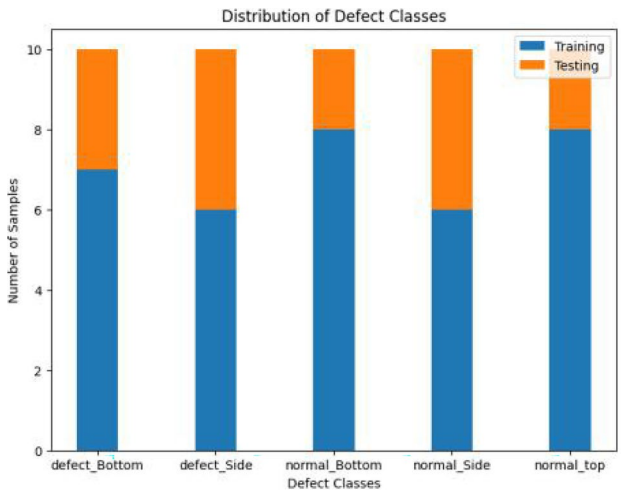


**Fig. 7** 80:20 split of training and testing data for the CPR main dataset

- Latent space analysis: The plots in Fig. 8 represent the learning process of the latent space for different classes.

Compared to the traditional VAE's latent space, the proposed VAE's latent space demonstrates a clear separation between different classes because the points in Fig. 8a are closer for the same classes. This outcome indicates that the model has learned to encode the input data in a way that captures the underlying class-specific features. The t-SNE (t-Distributed Stochastic Neighbor Embedding) is performed to obtain these plottings [25]. t-SNE performs dimensionality reduction of the latent vectors from their original high-dimensional space to a lower-dimensional space to compress the representation of the original image. Here, the resulting vectors have two dimensions representing the pattern of the encoded images in the latent space. The better one group of (same color) points is linearly separable from the other groups of data by a straight line, the better the model learns about the variation in patterns of different classes [25].

However, some overlaps can be observed in both of the latent spaces, which suggests that the model needs improvement to create distinct representations for each class. Overlapping points between classes indicate that the model struggles to differentiate between certain instances the classes have similarities, which is one of the limitations of the CPR dataset. Because some of the samples in this dataset from different classes are visually identical.

- Training loss: Fig. 9 plots the loss scores for each approach on the y-axis against the epoch numbers on the x-axis. Although both of the curves exhibit a consistent downward trend, there is a subtle difference in terms of consistency in performance. From the loss graph, we can observe the traditional VAE approach demonstrates a relatively high initial loss value of around 12.13 and experiences a rapid decrease in loss over the first few epochs. In contrast, the line for the proposed approach starts with a lower initial loss value of approximately 0.1087 and steadily reduces the loss throughout the training process. At the end of the 100th epoch, both of the approaches achieve a significantly lower final loss value of about 0.0081 after 100 epochs. However, the high loss and fluctuation of the performance by the traditional approach are clearly an indication of an overfitting issue,

**Table 3** Loss Functions and Classifier Performance Metrics

| Metrics | Description/Formula |
|---|---|
| 1. Training Loss | |
| Reconstruction Loss + Weight Decay | $L_{reconstruction} + L_{wd}$ |
| Weight Decay Loss ($L_{wd}$) | $L_{wd} = \lambda \ (param^2)$ |
| Reconstruction Loss ($L_{reconstruction}$) | $L_{reconstruction} = \frac{1}{N} \left( \lVert X_{train} - X_{train}' \rVert^2 \right)$ |
| 2. Test Loss | |
| Reconstruction Loss + Kullback–Leibler Divergence | $L_{reconstruction} + KLD$ |
| Reconstruction Loss ($L_{reconstruction}$) | $L_{reconstruction} = \frac{1}{N} \ (\lVert X_{test} - X_{test}' \rVert^2)$ |
| Kullback–Leibler Divergence (KLD) | $KLD = P(x)\log\frac{P(x)}{Q(x)}$ |
| P (True distribution) | |
| Q (Learned distribution) | |
| 3. Classifier Performance Metrics | |
| Precision | $\frac{TP}{TP+FP}$ |
| Recall | $\frac{TP}{TP+FN}$ |
| F1-Score | $\frac{2 \cdot Precision \cdot Recall}{Precision+Recall}$ |
| Test Accuracy | Accuracy on test data |

*TP* True positive, *FP* False Positive, *TN* True Negative, *FN* False Negative

which has also been reflected in the qualitative output, which we discuss next.

- Quality of generated output: We illustrate the output of the VAE model with the traditional regularization technique in Fig. 10a and the out of the updated regularization technique in Fig. 10b in order to observe the quality of the synthetic images. It is evident from the output images that the VAE model with the traditional regularization technique provides better resolution and correct output in comparison with the proposed VAE model with the updated regularization technique. To elaborate, having a closer look at the images in Fig. 10a, we can observe that the first constructed image in the list is blurry and the features of the images are hardly visible. In addition, the second generated image in Fig. 10b is a different image. As a consequence, training a classifier with such poor-quality images leads to lower accuracy in the model ranging from 58%–61%.

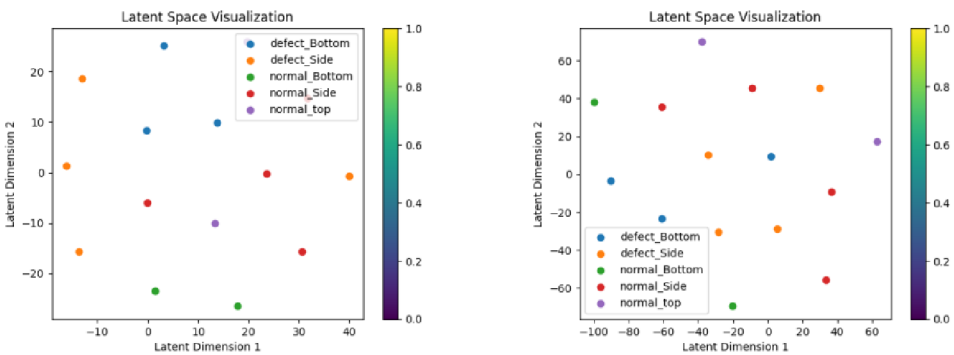


**Fig. 8** The latent vectors represent the encoded latent space representations of the images, and the classes are represented by different colors. The x-axis represents the first dimension of the reduced latent vectors, and the y-axis represents the second dimension. Each point in the scatter plot

(a) Latent space of the VAE model with traditional regularization technique

(b) Latent space of the VAE model with the proposed regularization technique

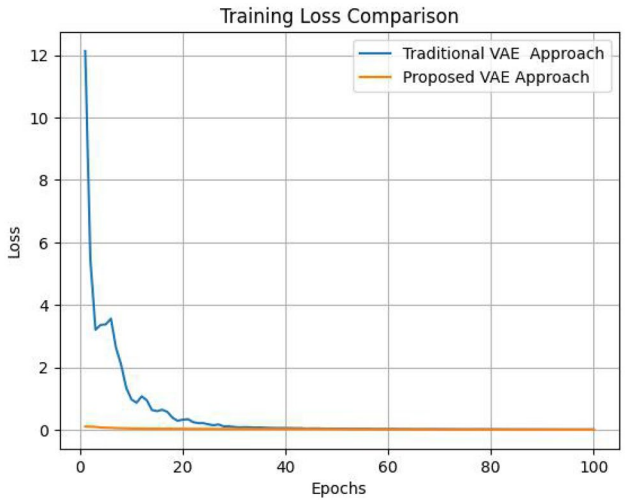


**Fig. 9** Training loss comparison graph of VAE model with traditional regularization technique VS VAE model with proposed regularization technique

- Test loss: We obtained the test loss by combining the reconstruction loss and KLD to observe the difference between the original and reconstructed images. While generating samples using the traditional VAE we obtained low loss values, nearly converging to 0. By contrast, with the incorporation of weight decay into our training the test loss was reduced from an initial value of 0.030 to a significantly improved 0.021. Interestingly, the training loss for the proposed VAE approach was more stable and lower compared to the traditional VAE approach.

### Comparison Between the Proposed and Existing Data Generation Approach

Due to the scarcity of a benchmark dataset for railway defect generation, we explored alternative synthetic data generation methods as proposed in the literature.

- Synthetic data generation by fixed-rule based geometric transformation: We have demonstrated the results of employing geometric transformations on images in Fig. 11a. To achieve this, we employed random rotation techniques on the original images sourced from another CPR dataset, consisting of 1000 defective and 2000 normal images. In a previous study [6], we employed this approach in the data pre-processing phase for differentiating between defective and non-defective images. Although this geometric augmentation led to improved accuracy in binary defect classification, we encountered challenges related to significant overfitting when employing CNN models.

  Basically, the augmented images from fixed-rule-based geometric transformations do not obtain substantial alterations in their patterns or textures. From Fig. 11a, it is visible that the augmented images lack authenticity in terms of their patterns and textures (e.g., there is no difference between the synthetic images other than the angle). This absence of realistic variation contributed to a defect classifier in a preceding paper achieving high accuracy during training. However, during testing, the

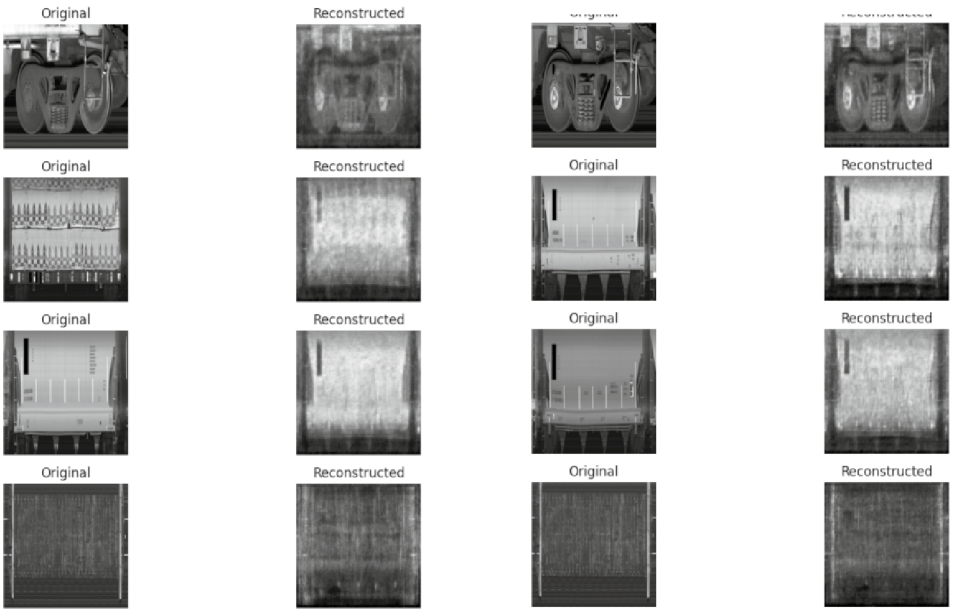


**Fig. 10** Comparison of qualitative results after applying traditional vs proposed regularization technique

(a) Synthetic images generated by the VAE model with traditional regularization technique

(b) Synthetic images generated by the VAE model with proposed regularization technique

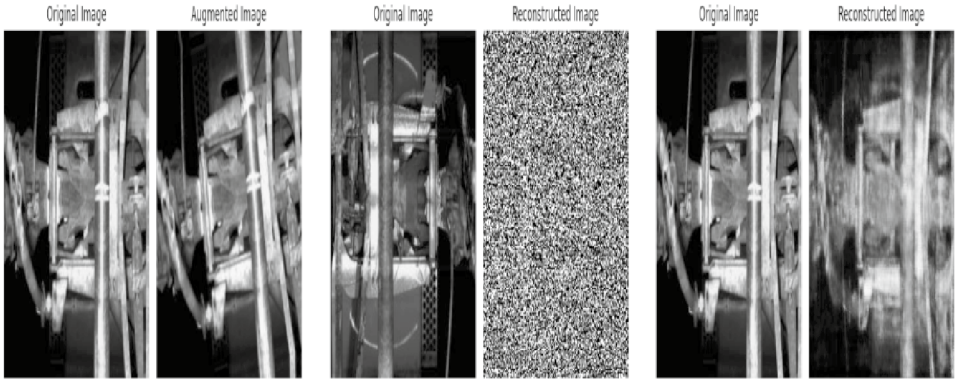


**Fig. 11** Comparison of synthetic images generated by different approaches

(a) Synthetic images gener-ated by fixed-rule based ge-ometric rotation

(b) Synthetic images generated by GAN based approach

(c) Synthetic images gener-ated by the proposed ap-proach

model experienced elevated loss and fluctuating performance. As a result, the application of geometric transformations did not sever the purpose of generalizing the railway defect classifier.

- Synthetic data generation by GAN model: We extended our exploration to the CPR dataset by applying a GAN-based transformation. The result of this process is illustrated in Fig. 11b. We intended to utilize GAN to generate synthetic datasets from the given CPR dataset. However, we found that the GAN is limited to dealing with a relatively small sample size. In our case, the GAN model was unable to generate synthetic samples. With only 50 samples, the GAN model could not capture the full range of diversity inherent in the target image dataset. In other words, the model collapsed and could not generate meaningful samples due to improper noise modeling.

  These findings underscored the dependency of the GAN model on availability of a sufficiently large and diverse dataset for effective training. Therefore, in scenarios where defect data is infrequent and hard to collect GAN may not be suitable for generating synthetic defect samples due to their sensitivity to sample size.
- Synthetic data generation by the proposed VAE model: The synthetic images generated by our proposed VAE approach are depicted in Fig. 11c. In this technique, we leveraged the latent space of the VAE to generate meaningful and diverse synthetic images from limited samples.

  Geometric transformations, as evidenced in Fig. 11a exhibit exact patterns and textures akin to original images; which usually leads to high accuracy but overfitted model. By contrast, the proposed approach by capturing the underlying distribution in the latent space of the image. The result in Fig. 11c illustrates that the proposed VAE-generated synthetic images exhibit realistic patterns and textures similar to those of the original images. Such output subdues the limitations posed by the fixed-rule-based geometric transformation method where hardly any variation in texture and pattern is observed. Moreover, the weight decay-based regularization technique strengthened the proposed VAE approach to handle the overfitting issue.

  In summary, The comparative analysis of the three synthetic data generation methods highlights the shortcomings of fixed-rule-based geometric transformations and the sensitivity of GAN-based transformation to sample size limitations. In our subsequent evaluations, we demonstrate the superior performance of the defect classifier trained on the proposed VAE-generated synthetic data compared to the models trained on the other two approaches.

## Evaluation of ViT Defect Classifier

We generate a synthetic dataset from the CPR main dataset using Algorithm 1. This dataset comprises 500 images, with 50 being original and the remaining being reconstructed images.

### Training

Based on the findings from our precedent research detailed in [7] and [5], we underscored the need for a reusable model for defect classification. Therefore, we trained a MobileViT model. MobileViT is a visual transformer, that integrates MobileNet CNN with the visual transformer approach to create an efficient and lightweight model [16]. We utilize the feature extraction of MobileNetV3, succeeded by the attention-driven capabilities of ViT to focus on the salient features within defect images. The integration of MobileNetV3 is strategically chosen for its lightweight nature and efficient processing capabilities, making it suitable for real-time deployment scenarios. After the feature extraction through MobileNetV3, a 768-dimensional dense layer

processes the high-level features which are then refined for classification through a series of dense layers.

The dense layers are fine-tuned to process the rich representations from the output of ViT. The first dense layer with 512 units and Rectified Linear Unit (ReLU) activation function as a transitional phase. The ReLU function bridges the high-dimensional feature space to the final output layer. This output layer, consisting of 5 units with a sigmoid activation, corresponds to the classification of the distinct railway defect types.

The training hyperparameters were dynamically selected, with an initial learning rate of 0.001 conducive to gradual convergence and the implementation of a cosine decay schedule to refine the model's weights over successive epochs. These parameters, alongside chosen data augmentation techniques, such as random cropping and color adjustments, aim to boost the model's invariance to operational variances in defect images. We performed 80:10:10 Training:Testing: Validation split. Consequently, we fine-tuned this model using pre-trained weights, employing 440 training images, 55 validation images, and 55 testing images, all distributed among 5 distinct classes derived from the CPR synthetic dataset.

### Classifier Performance

In Fig. 12, we illustrate the actual and predicted labels obtained from the classifier trained on a single batch of images representing one of the defect classes. This batch includes test samples from both synthetic and original images. Notably, none of these test samples have been used in the training set the actual labels are predicted with high precision, yielding nearly 0% loss and 100% accuracy. Such high accuracy holds significant importance in the context of digital twining diagnostic processes [6]. Therefore, these outcomes substantiate the efficacy of leveraging generative AI to address issues related to limited defective samples and to build an accurate defect classifier.

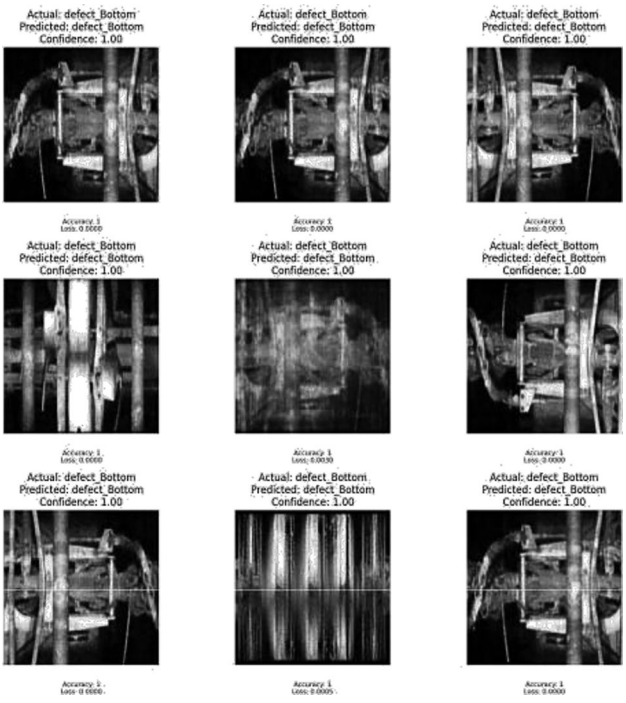


**Fig. 12** Actual and predicted labels are categorized by the trained classifier derived from a single batch of images, encompassing both synthetic and original images

Furthermore, we evaluated the defect classifier trained on synthetic data for each class, focusing on precision, recall, and accuracy. The results showed that we achieved 98%–99% accuracy for precision, recall, and F1-score, as illustrated in Fig. 13. The question of overfitting may arise when a model exhibits high accuracy with a relatively small sample size. Nevertheless, as evidenced by our qualitative output in Fig. 12, both the synthetic test defect samples and original test defect samples were accurately classified, providing further validation of the model's effectiveness.

### Summary of Findings

In summary, our comprehensive evaluation yields the following key findings:

- Our synthetic data-trained defect classifier consistently exhibits high precision, recall, and F1-score metrics, indicating its strong performance across all classes.
- The high precision values underscore the model's capacity to accurately identify and classify defects with minimal false positives. This precision is particularly vital in domains like railway quality control, where avoiding unnecessary interventions or alarms is paramount.
- The substantial recall scores signify the classifier's effectiveness in capturing a significant proportion

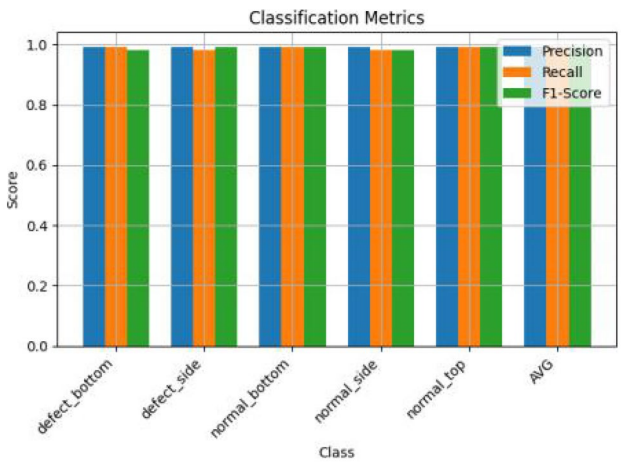


**Fig. 13** Comparison of classification metrics per class

of true positives. In safety–critical industries such as railways, this high recall rate implies the detection of a minimal number of defects that might otherwise ignored by the model.
- The achievement of a high F1-score highlights the model's efficiency, as it consistently maintains a balanced trade-off between precise defect identification and false positive reduction.

In summary, the results from batch predictions and accuracy analysis show that our model performs effectively in classifying both synthetic and original defects, indicating good generalization.

## Discussion

The following discussion elaborates on our approach, findings, and implications for the field of AI-driven railway defect detection.

- Integration with current AI and ML trends: The integration of few-shot learning with the VAE model represents a significant advancement in addressing data scarcity, particularly for rare defect types in railway systems. Our approach was tested on unseen samples, including those with as few as five images for specific defect classes. The outcomes, as determined through both qualitative and quantitative analysis, have demonstrated decent generalizability. The potential of few-shot learning to enhance the capability of the VAE model has been realized by enabling accurate defect classification with a significantly low number of training samples. The successful application of few-shot generalizability, supported by the generation of synthetic samples from a minimal number of seed images, indicates the robustness of our approach in overcoming the challenges posed by scarce data in specific defect types.
- Scope for no-sample adaption: While the VAE model has proven effective in classifying the CPR dataset with high accuracy, the journey toward zero-shot learning presents a new frontier. The potential to classify entirely unseen defects and their classes through the integration of visual-linguistic components, such as BLIP-2 [27], for grounded image captioning can be considered for future research. This approach could advance defect classification in regions where certain defects are uncommon, leveraging textual descriptions and correlations between defect features and generated captions to classify these rare instances.
- Addressing imbalanced datasets: The challenge of imbalanced datasets was met in this study for CPR dataset. Through the application of the proposed regularization technique, we not only achieved better fidelity but also significantly reduced reconstruction loss. This was particularly evident in the equalization of sample sizes across classes, a feat accomplished by our VAE-based synthetic data generation algorithm. By generating a desired number of samples for underrepresented classes, we addressed the imbalance. Therefore, we could achieve multi-class classification in this study. This work shows an improvement over our previous studies on the CPR dataset [6, 7], showcasing the adaptability of our model to diverse classification scenarios.

## Limitations and Future Work

Although our proposed approach outperformed conventional rule-based data augmentation strategies and demonstrated better performance compared to GAN-driven techniques in data quality and adaptability, there are a few limitations of this paper to mention.

### Limitations

First, while our approach performs better at generating precise replicas of defects, it may not be ideal for tasks demanding higher diversity in synthetic data, as our primary goal was ensuring accurate representation per class rather than promoting variation.

Second, the differences between certain defect types mean that occasionally, two almost identical defects might be reconstructed with varying quality.

Finally, the approach was tailored for railway defect detection, and its applicability in other contexts remains to be seen. Moreover, the Zero-shot classification for unseen samples and unseen defect types needs to be explored further. However, incorporating such features is still challenging for specialized domains. Because there might not be widely available textual descriptions about rare defects.

### Future Work

Although the model was primarily designed for railway defect detection, its potential for applications in other fields, where acquiring genuine training samples poses significant challenges. For instance,

- Aerospace and Aviation: Where simulated crashes or malfunctions are essential for safety research but physically creating such scenarios is prohibitively expensive and dangerous.
- Automotive Testing: Modern vehicles with advanced driver-assistance systems require vast amounts of data to ensure safety, and it's not feasible to induce real-world accidents or defects for data collection.

- Seismology: Creating datasets for earthquake prediction models would require waiting for real-world earthquakes, which are not only unpredictable but could be catastrophic.
- Medical Imaging for Rare Diseases: Obtaining sufficient samples for rare diseases without resorting to synthetic data is a major bottleneck.

In a nutshell, future studies could modify and fine-tune the proposed approach to produce more diverse synthetic data sets, particularly beneficial in scenarios where data generation is complex and challenging; and creativity and variation in data generation are crucial.

## Conclusion

This study investigates the potential of integrating generative techniques, such as VAE, within ICS to enhance learning and decision-making processes in areas where real-world data is scarce or difficult to procure. By employing a VAE-based approach for synthetic data generation, we demonstrated how ICS can emulate human-like analysis even with limited data. Our approach out-performed traditional rule-based data augmentation techniques and showcased improved robustness compared to GAN-based methods in terms of data quality and generalization. The application of our proposed approach in railway defect detection paves the way for further exploration and adoption of generative AI in ICS across various domains including health, climate, and heavy freights facing similar challenges.

**Supplementary Information** The online version contains supplementary material available at https://doi.org/10.1007/s12559-024-10283-3.

**Authors Contribution** Rahatara Ferdousi: Conceptualization, Methodology, Writing—Original Draft Preparation, Visualization, Validation, Software. Chunsheng Yang: Review & Data Collection, Validation, Project administration. M. Anwar Hossain: Writing—Review & Editing, Validation, Supervision. Fedwa Laamarti: Writing—Review & Editing. Abdulmotaleb El Saddik & M. Shamim Hossain: Discussion, Review, Funding acquisition, Supervision.

**Funding** The authors extend their appreciation to the researchers supporting project number (RSP2024R32), King Saud University, Riyadh, Saudi Arabia. This research is also supported in part by collaborative research funding from the National Program Office under the National Research Council of Canada's Artificial Intelligence for Logistics Program. The project ID is AI4L-123.

**Data Availability** We have signed a business agreement with the data provider. Accordingly, the datasets analyzed during the current study are not publicly available. However, we have consent for publishing the result.

## Declarations

**Competing interests** The authors declare that there are no conflicts of interest regarding the publication of this paper.

## References

1. Zheng Y, Tuan LA, Novel A. Cognitively Inspired, Unified Graph-based Multi-Task Framework for Information Extraction. Cogn Comput. 2023;15:2004–13. https://doi.org/10.1007/s12559-023-10163-2 
2. Gudivada VN, Pankanti S, Seetharaman G, Zhang Y. Cognitive computing systems: Their potential and the future. Comp. 2019;52(5):13–8.
3. Tabernik D, Šela S, Skvarč J, et al. Segmentation-based deep-learning approach for surface-defect detection. J Intell Manuf. 2020;31:759–76. https://doi.org/10.1007/s10845-019-01476-x.
4. Tabernik D, Sela S, Skvarˇc J, Skoˇcaj D. Deep-learning-based com-puter vision system for surface-defect detection. In: In Computer Vision Sys-tems: 12th International Conference, ICVS 2019, Thessaloniki, Greece, September 23–25, 2019, Proceedings 12. Springer; 2019. p. 490–500.
5. Ghaboura S, Ferdousi R, Laamarti F, Yang C, El Saddik A. Digital twin for railway: A comprehensive survey. IEEE Access. 2023;11:120237–57.
6. Ferdousi R, Laamarti F, Yang C, El Saddik A. Railtwin: a digital twin framework for railway. In 2022 IEEE 18th International Conference on Automation Science and Engineering (CASE). IEEE; 2022. p. 1767–72.
7. Yang C, Ferdousi R, El Saddik A, Li Y, Liu Z, Liao M. Lifetime learning-enabled modelling framework for digital twin. In: In 2022 IEEE 18th International Conference on Automa-tion Science and Engineering (CASE). 2022. p. 1761–6.
8. Cui S, Wang H, Zhang M, Zhang X. Defect classification on lim-ited labeled samples with multiscale. Appl Intell. 2021;51(6):3911–25.
9. Alqudah R, Al-Mousa AA, Hashyeh YA, Alzaibaq OZ. A systemic comparison between using augmented data and syn-thetic data as means of enhancing wafermap defect classification. Comput Ind. 2023;145:103809.
10. Xiao Y, Huang Y, Li C, et al. Lightweight Multi-modal Representation Learning for RGB Salient Object Detection. Cogn Comput. 2023;15:1868–83. https://doi.org/10.1007/s12559-023-10148-1.
11. Abufadda M, Mansour K. A survey of synthetic data generation for machine learning. In: In 2021 22nd international arab conference on information technology (ACIT). 2021. p. 1–7.
12. Jain S, Seth G, Paruthi A, et al. Synthetic data augmentation for surface defect detection and classification using deep learning. J Intell Manuf. 2022;33:1007–20. https://doi.org/10.1007/s10845-020-01710-x.
13. Zhang G, Cui K, Hung TY, Lu S. Defect-gan: High-fidelity defect synthesis for automated defect inspection. In: In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2021. p. 2524–34.
14. Tevosyan A, Khondkaryan L, Khachatrian H, Tade-vosyan G, Apresyan L, Babayan N, Stopper H, Navoyan Z. Improving vae based molecular representations for compound property prediction. Journal of Cheminformatics. 2022;14(1):69.
15. He X, Chang Z, Zhang L, Xu H, Chen H, Luo Z. A survey of de-fect detection applications based on generative adversarial networks. IEEE Access. 2022;10:113493–512.
16. Lu Y, Shen M, Wang H, Wang X, van Rechem C, Wei W. Machine learning for synthetic data generation: a review. arXiv preprint arXiv:2302.04062, 2023.
17. Endres M, Mannarapotta Venugopal A, Tran TS. Syn-thetic data generation: a comparative study. In: In Proceedings of the 26th International Database Engineered Applications Symposium. 2022. p. 94–102.
18. Pinheiro Cinelli L, Ara'ujo Marins M, Bar-ros da Silva EA, Lima Netto S. Variational autoencoder. In: In Varia-tional Methods for

Machine Learning with Applications to Deep Networks. Springer; 2021. p. 111–49.
19. Mak HWL, Han R, Yin HHF. Application of variational autoencoder (vae) model and image processing approaches in game design. Sensors. 2023;23(7):3457.
20. Kumar T, Mileo A, Brennan R, Bendechache M. Image data augmentation approaches: A comprehensive survey. arXiv preprint arXiv:2301.02830, 2023.
21. Wang R, Hoppe S, Monari E, Huber MF. De-fect transfer gan: Diverse defect synthesis for data augmentation. arXiv preprint arXiv:2302.08366, 2023.
22. Zhang G, Cui K, Hung TY, Lu S. Defect-gan: High-fidelity defect synthesis for automated defect inspection. arXiv preprint arXiv:2103.15158, 2021.
23. Jadon A, Kumar S. Leveraging generative ai models for synthetic data generation in healthcare: Balancing research and privacy. In: In 2023 International Conference on Smart Applications, Communications and Networking (SmartNets). IEEE; 2023. p. 1–4.
24. Wang Z, Healy G, Smeaton AF, Ward TE. Use of neural signals to evaluate the quality of generative adversarial network performance in facial image generation. Cogn Comp. 2020;12:13–24.
25. Alpaydin E. Introduction to machine learning. MIT press, 2020.
26. Shang H, Sun C, Liu J, Chen X, Yan R. Defect-aware transformer network for intelligent visual surface defect detection. Adv Eng Inform. 2023;55: 101882.
27. Li J, Li D, Savarese S, Hoi S. Blip-2: Boot-strapping language-image pre-training with frozen image encoders and large language models. In: In International conference on machine learning. PMLR; 2023. p. 19730–42.

# Terms and Conditions

Springer Nature journal content, brought to you courtesy of Springer Nature Customer Service Center GmbH (“Springer Nature”).
Springer Nature supports a reasonable amount of sharing of research papers by authors, subscribers and authorised users (“Users”), for small-scale personal, non-commercial use provided that all copyright, trade and service marks and other proprietary notices are maintained. By accessing, sharing, receiving or otherwise using the Springer Nature journal content you agree to these terms of use (“Terms”). For these purposes, Springer Nature considers academic use (by researchers and students) to be non-commercial.
These Terms are supplementary and will apply in addition to any applicable website terms and conditions, a relevant site licence or a personal subscription. These Terms will prevail over any conflict or ambiguity with regards to the relevant terms, a site licence or a personal subscription (to the extent of the conflict or ambiguity only). For Creative Commons-licensed articles, the terms of the Creative Commons license used will apply.
We collect and use personal data to provide access to the Springer Nature journal content. We may also use these personal data internally within ResearchGate and Springer Nature and as agreed share it, in an anonymised way, for purposes of tracking, analysis and reporting. We will not otherwise disclose your personal data outside the ResearchGate or the Springer Nature group of companies unless we have your permission as detailed in the Privacy Policy.
While Users may use the Springer Nature journal content for small scale, personal non-commercial use, it is important to note that Users may not:

1. use such content for the purpose of providing other users with access on a regular or large scale basis or as a means to circumvent access control;
2. use such content where to do so would be considered a criminal or statutory offence in any jurisdiction, or gives rise to civil liability, or is otherwise unlawful;
3. falsely or misleadingly imply or suggest endorsement, approval , sponsorship, or association unless explicitly agreed to by Springer Nature in writing;
4. use bots or other automated methods to access the content or redirect messages
5. override any security feature or exclusionary protocol; or
6. share the content in order to create substitute for Springer Nature products or services or a systematic database of Springer Nature journal content.

In line with the restriction against commercial use, Springer Nature does not permit the creation of a product or service that creates revenue, royalties, rent or income from our content or its inclusion as part of a paid for service or for other commercial gain. Springer Nature journal content cannot be used for inter-library loans and librarians may not upload Springer Nature journal content on a large scale into their, or any other, institutional repository.
These terms of use are reviewed regularly and may be amended at any time. Springer Nature is not obligated to publish any information or content on this website and may remove it or features or functionality at our sole discretion, at any time with or without notice. Springer Nature may revoke this licence to you at any time and remove access to any copies of the Springer Nature journal content which have been saved.
To the fullest extent permitted by law, Springer Nature makes no warranties, representations or guarantees to Users, either express or implied with respect to the Springer nature journal content and all parties disclaim and waive any implied warranties or warranties imposed by law, including merchantability or fitness for any particular purpose.
Please note that these rights do not automatically extend to content, data or other material published by Springer Nature that may be licensed from third parties.
If you would like to use or distribute our Springer Nature journal content to a wider audience or on a regular basis or in any other manner not expressly permitted by these Terms, please contact Springer Nature at

onlineservice@springernature.com