

Generative AI–Based Product Defect Detection

Sapna Biradar¹ and Dr. Mohammed Abdul Waheed²

Student¹, Associate Professor², Department of Computer Science and engineering, VTU's CPGS Kalaburagi, Karnataka, India

sapnabiradar80@gmail.com¹ and mawaheed@vtu.ac.in²

Received: 14th June, 2026; Revised: 20th July 2026; Accepted: 20th August, 2026; Available Online: 05th September 2026

ABSTRACT

Automated defect detection is one of the cornerstones that help maintain the quality of products in industrial manufacturing. Traditional approaches and some machine learning algorithms rely greatly on massive datasets with a high volume of labelled images of various defect types, which renders them ineffective in recognizing rare defects or defects that do not exist yet. The actual industrial environment lacks the necessary samples, and surface flaws may vary from scratches to dents to contaminations under different lighting and environmental conditions. To address this issue, a Generative AI–driven Product Defect Detection System combining generative modelling, self-supervised learning, vision transformers, and explainable AI is proposed.

Firstly, this system obtains the images of industrial products, which may be taken from such resources as the MVTec AD dataset or factory datasets, and then applies image preprocessing and augmentation to improve the quality of datasets. A generator trained on defect-free product images learns to recognize their surface distribution and generates images of the same quality during inference, and discrepancies are considered an indication of defects. Moreover, in order to increase robustness and cope with the shortage of defect samples, this method involves a synthetic defect generation algorithm for the creation of artificial defects. This way, the system learns how to detect defects without being provided with a substantial number of training examples. The next step is applying a Vision Transformer (ViT) to extract both local and global features from an image and perform the actual detection. Finally, a multi-scale anomaly detection is utilized, and an Explainable AI module is integrated to generate heatmaps and identify specific defective parts.

Performance evaluation measures include accuracy, precision, recall, F1-score, and ROC-AUC; this system is compared to traditional CNN and GAN methods for defect detection.

Keywords: Generative Adversarial Network (GAN), accuracy, precision, recall, F1-score, ROC-AUC, Vision Transformer (ViT), AI, CNN.

How to cite this article: Biradar S, Waheed MA, Generative AI–Based Product Defect Detection Int J Drug Deliv Technol. 2026;16(83s):1731-1740. DOI: 10.25258/ijddt.16.83s.188

Source of support: Nil.

Conflict of interest: None

INTRODUCTION

The current state of advanced manufacturing demands that maintaining quality be crucial to remain competitive, satisfy the needs of customers, and ensure smooth operation. Automation of defect-detection processes through special systems became popular for ensuring speed and consistency of inspections. However, existing machine-learning algorithms are mostly based on supervised learning with a great dependence on huge amounts of labeled data and pre-defined categories of defects, which can negatively affect their applicability when there are few defect samples.

First of all, there might be various types of defects that products manufactured in production lines may have, including scratches, dents, cracks, or contamination. These defects can differ by shape, size, appearance, lighting conditions, etc. Therefore, conventional supervised models may have difficulties generalizing from the limited set of

images used for training when facing new defects that were not encountered earlier. Recent achievements in Generative AI provided several effective means of overcoming this problem. For instance, GANs and diffusion models could learn distributions for the images with no defects and reproduce them later. Comparing these images to the original pictures allows detecting anomalies and spotting defects that did not require any labelling process.

Furthermore, improvements in parameter-efficient fine-tuning and text-to-image generation, as well as DisenBooth and apprenticeship learning algorithms, contributed to increasing the flexibility of generative models. Moreover, few-shot learning and synthetic data generation methods allow reducing the dependency on labelled samples and enhancing the generalization power of models. In particular, defect-aware feature manipulation allowed generating synthetic defects.

In the context of advanced manufacturing, a Generative AI–Based Product Defect Detection System is proposed.

*Author for Correspondence: sapnabiradar80@gmail.com

This hybrid approach combines a transformer-based image classification model with generative modelling and self-supervised learning. A Vision Transformer (ViT) was selected as an encoder to generate the representation for global and local contextual features. Additionally, the adaptive defect generation module was implemented to create the synthesized defect pattern for a product.

Another contribution is the inclusion of Explainable AI (XAI) techniques into the architecture of the proposed model. In particular, heatmap visualization is used as a tool for explaining what parts of the images contain detected defects. XAI-based models have already been applied in various industries, for example, for environmental hazard detection purposes.

Thus, a Generative AI and Transformers-Based Product Defect Detection System is proposed, which utilizes an approach combining advantages of transformers and generative models. It uses few labelled datasets for training and enables effective detection of rare and new defects.

LITERATURE SURVEY

Z. Li et al. (2025) suggest a diffusion model-driven framework for automated defect detection in industrial products. Through the use of data generation, the method solves the problem of scarce data and achieves high performance due to learning on a combination of actual and synthesized images in complex environments.

Methodology:

Generation of defect images based on diffusion models + supervised classifier.

Limitations:

Resource-intensive method.

Y. Chen et al. (2025) suggest a mixed approach by using GANs together with Vision Transformers (ViT) for defect recognition. The generator is used for creating different types of defects and transformers are responsible for better feature extraction and classification.

Methodology:

GAN-based data augmentation and transformer for classification.

Limitations:

Highly complex structure of the model; difficulties in training due to nature of GAN.

Adnan Md Tayeb et al. (2024) developed an approach to generating defect images through Stable Diffusion which can be used for steel surface detection. Such data generation technique helps to improve accuracy for limited training data and finds practical application in real industrial scenarios.

Methodology:

Synthesizing defect images based on Stable Diffusion.

Limitations:

Quality depends on prompt design; computationally expensive.

Adnan Md Tayeb et al. (2024) introduced multimodal real-time detection of road defects. Using sensor data from several devices, they provide accurate information about possible road defects thus improving safety during driving. The model is robust enough to work in challenging real-world environments.

Methodology:

Multimodal learning and sensor fusion.

Limitations:

Several sensors are needed; more hardware dependent method.

Adnan Tayeb et al. (2024) introduced a smart parking solution based on decentralized architecture and using blockchain technologies. It provides security and guarantees that all resources are shared fairly among participants.

Methodology:

Decentralized blockchain-based solution.

Limitations:

Not defect recognition; relatively complex system.

J. Wei et al. (2023) investigated generation of varied and controllable defect samples. They generated multi-class defects that can be applied in industries to increase transferability of generated data.

Methodology:

Generating diverse defects in order to augment data sets.

Limitations:

Parameter tuning required; potential limitations in terms of generalizability.

N. Ruiz et al. (2023) presented the method of dreamBooth that allows generating images based on diffusion model with fine-tuning of the model with a few examples of the target individual/subject/personal. Thus, the model is adapted for generating personalized images and finds practical application for specific subjects.

Methodology:

Fine-tuning diffusion models based on individual samples.

Limitations:

Potential overfitting; requires fine-tuning.

A. M. Tayeb et al. (2023) proposed an approach for detecting defects related to skin conditions using image segmentation. In particular, this work focused on segmentation-based pores and wrinkle detection for dermatological studies.

Methodology

Data Acquisition and Processing

Industrial object images are obtained from credible sources such as the MVTec AD dataset and real-world

factories. When training the generative network, only images that are free of any defects are considered.

Image processing involves:

- Image resizing and normalization
- Denoising and enhancing image contrast

- Augmentation by rotation, flipping, and adjustment of brightness

These processes help in enhancing image data quality and increasing generalization capacity.

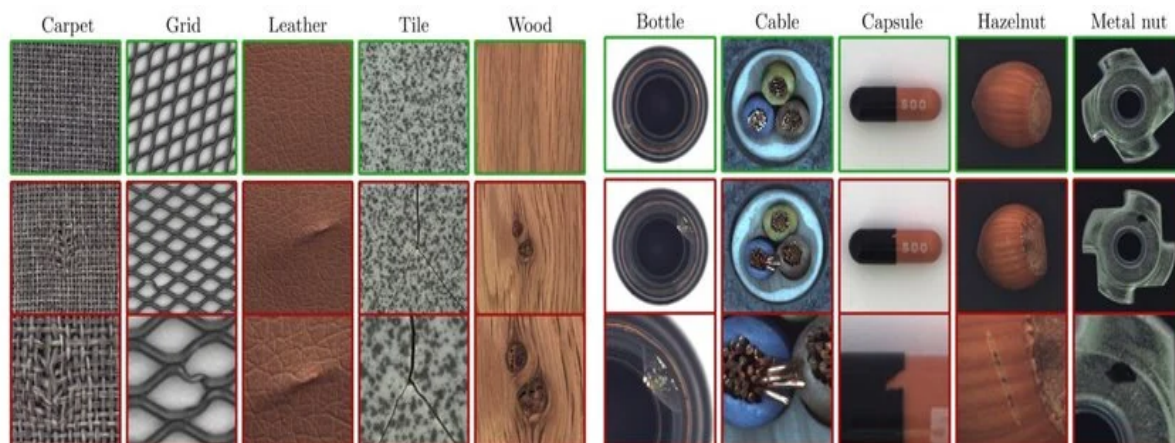


Figure 1: Dataset

Generative Model Training (GAN / Diffusion Model)

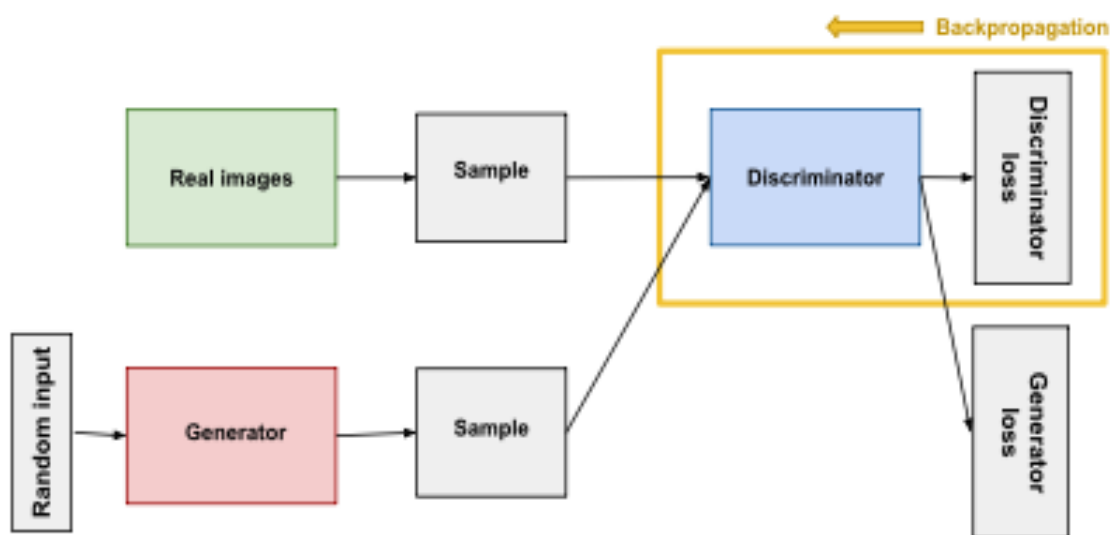


Figure 2: Generative Model Training (GAN / Diffusion Model)

The GAN/Diffusion Model is trained using only defect-free images; this means that it learns the distribution of surface textures/patterns of products that are defect-free.

In testing:

- The model will be able to recreate the input image based on the normal distribution learned.
- The reconstruction model hides any anomalies in the image since it is designed to recreate normal features.
- Any anomalies would be detected by finding the difference between the input and the reconstruction.

- This process is called Anomaly Detection via Reconstruction Error.

An error map will be generated as follows: $\text{Error} = |\text{Input Image} - \text{Reconstruction}|$.

Areas with high errors are likely to contain defects. Defect areas are detected based on applying a threshold to the error map; thus, this method allows anomaly detection even for unlabeled and rare defects.

Furthermore, Feature Extraction is done using ViT.

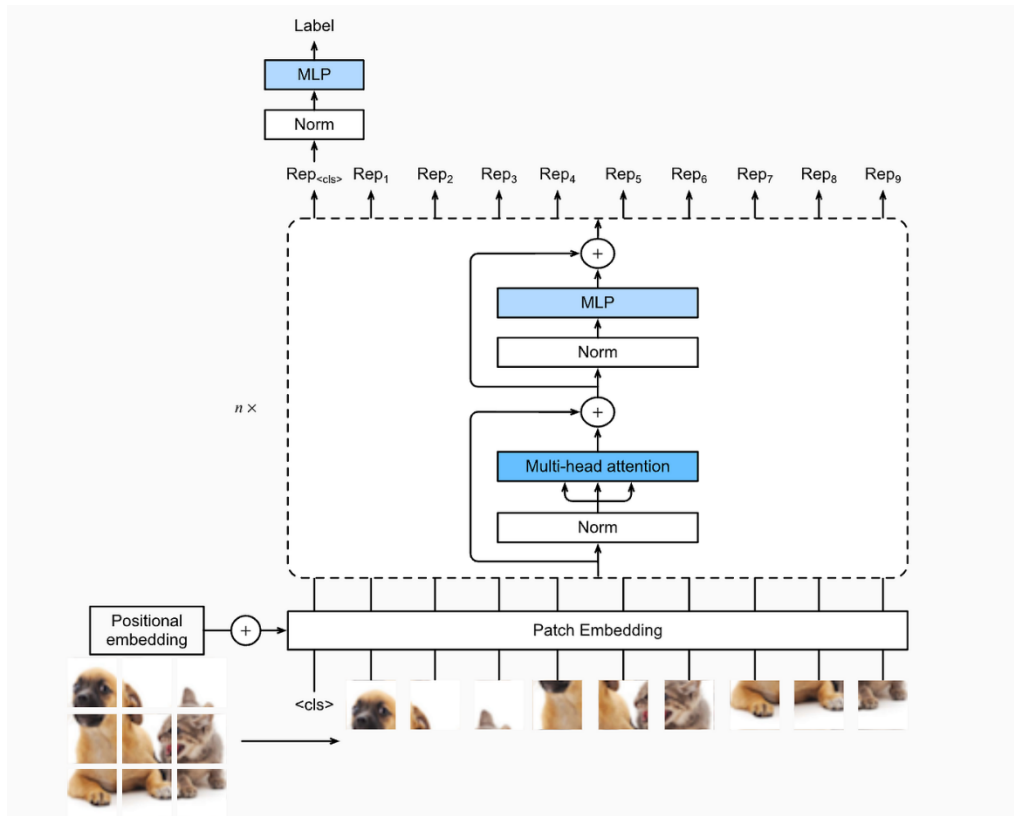


Figure 3: Feature Extraction using Vision Transformer (ViT)

Vision Transformer (ViT) is used to extract deep features. Image is divided into patches and analysed using self-attention, making connections between patches in the entire image. Advantages are:

- Global connections in the image can be captured
- Small flaws like scratches can be detected
- Improved detection accuracy by ViT compared to CNNs

Defect samples are produced dynamically during the process of training the network. Some approaches include:

- Defect creation using GANs
- Defect generation using diffusion

The main objective of this component is:

- Creation of rarely seen defects
- Making the model more robust
- Using less actual defect data

Multi-Scale Anomaly Detection

Defects can be present in different dimensions – from small to large. The system processes images at various scales:

- Small scale: detecting small cracks
- Large scale: detecting large dents

Thus, processing the input data at different scales allows for effective defect detection.

Module of Explainable AI (XAI)

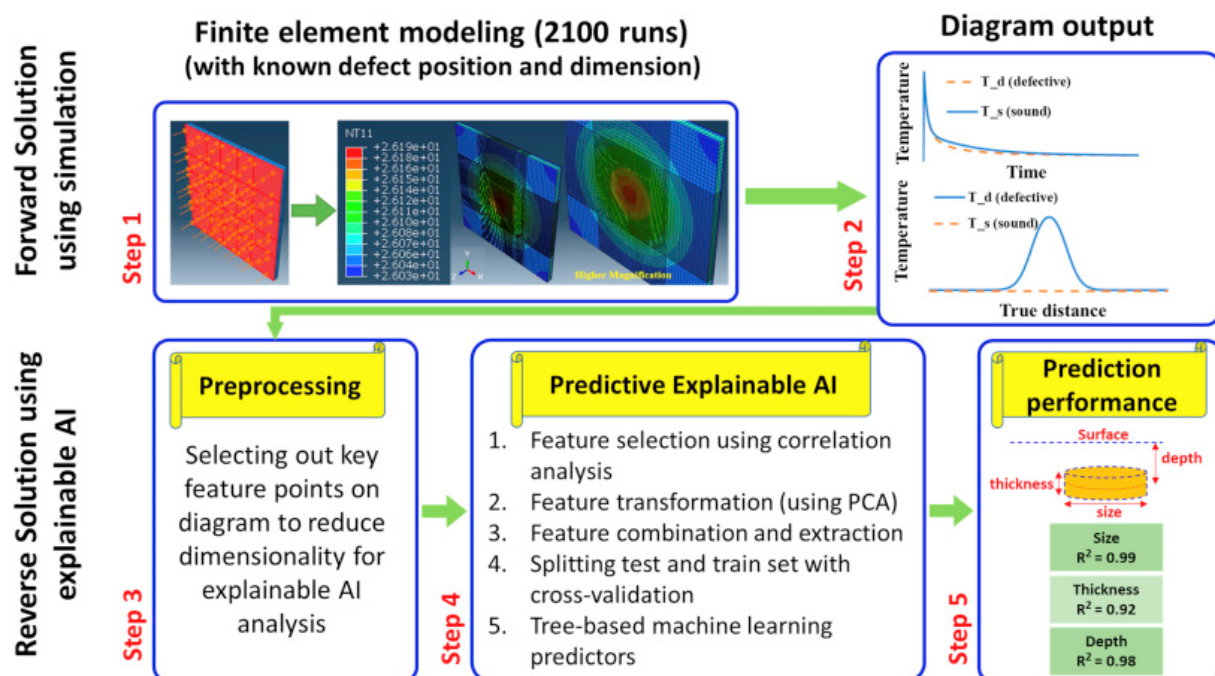


Figure 4: Explainable AI (XAI) Module

System Architecture

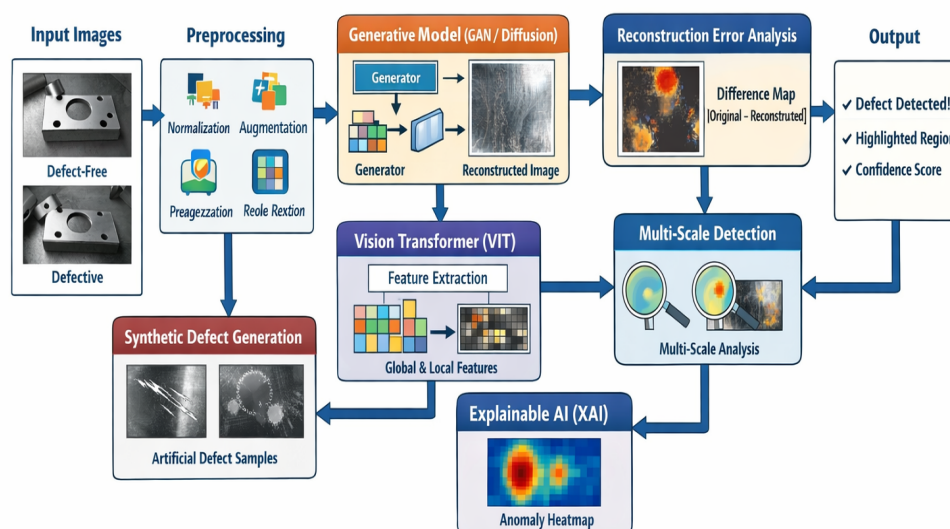


Figure 5: System Architecture

Figure 2: Generative AI-based product defect detection system showing a journey of input images through different modules for defect discovery and explanations.

Input Images

There are two types of images in this process:

- Non-defective images (to be used for training purposes)
- Defective images (to be used for testing purposes)

These input images are collected from industrial settings or already available databases.

2. Preprocessing

Prior to the main processing step, preprocessing techniques are applied to prepare images for further

analysis and improve the quality of the model's predictions:

- Normalization: ensuring pixels are on a uniform scale
- Augmentation: introducing diversity (rotation, flipping, brightness changes)
- Resizing and cleansing

These steps help enhance performance and increase generalization.

3. Generative Model (GAN / Diffusion)

The key component.

Trainable using only defect-free images.

During the inference phase:

- An attempt to generate an image from the input
- Generation of only natural patterns

Result: reconstructed image without any defects

4. Reconstruction Error Analysis

Comparing:

- The Original Image
- The Reconstructed Image

Difference map is generated: $\text{Difference} = |\text{Original} - \text{Reconstructed}|$

Basic principle:

- In normal regions, difference should be minimal
- In defective regions, the difference will be significant

This process is able to automatically identify defects

5. Vision Transformer (ViT)

For feature extraction from images.

Using the self-attention approach:

- Recognizes global features (general composition)
- Recognizes local features (small defects)

Used for identification of:

- Small scratches
- Micro cracks
- Minor surface defects

6. Synthetic Defect Generation

During the training phase, synthetic defects are introduced into the image:

- Scratches
- Cracks
- Artificial noise pattern

Reasons:

- Address lack of real defects dataset
- Improve resistance
- Familiarize model with uncommon defects

7. Multi-scale Detection

Detects various-sized defects:

- On the smaller scale - defects of small dimensions
- On the larger scale - major damages

Provides full-scale detection

8. Explainable AI (XAI)

Generates heatmaps for explanation purposes:

- Highlights defects
- Using methods such as Grad-CAM and saliency maps

Helps to interpret defect detection process

9. Output Module

Results include:

- Presence or absence of defect
- Region of the defect in the image
- Confidence value

Useful for:

- Industrial automation
- Quality control dashboards

RESULTS AND DISCUSSIONS

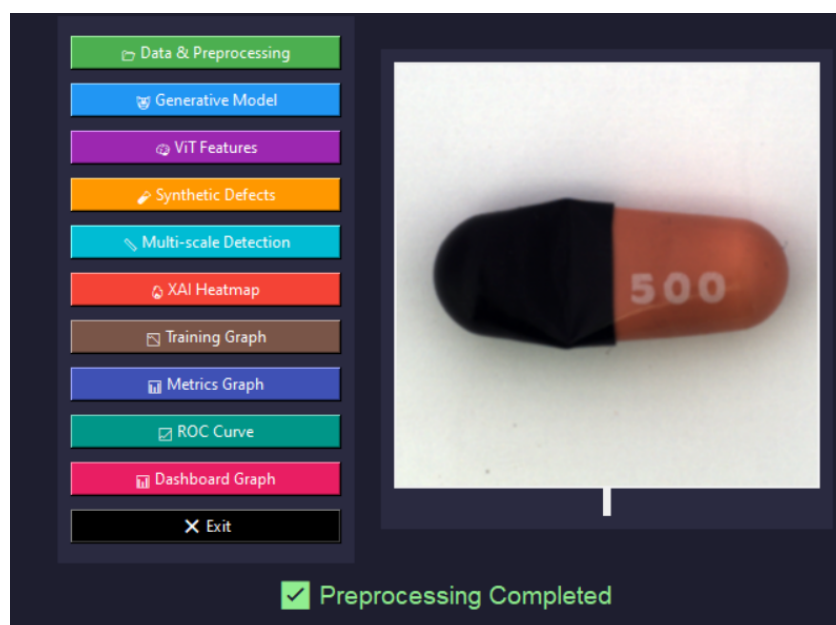


Figure 6: Read and processed data

IJDDT, Volume 16 Issue 83s, 2026

Preprocessing involves cleansing the image of the capsule by eliminating noise and sharpening the contrast in the image, which makes it ready for further processing.

Following the preprocessing process, the image is prepared for feature extraction and AI-powered defect detection.

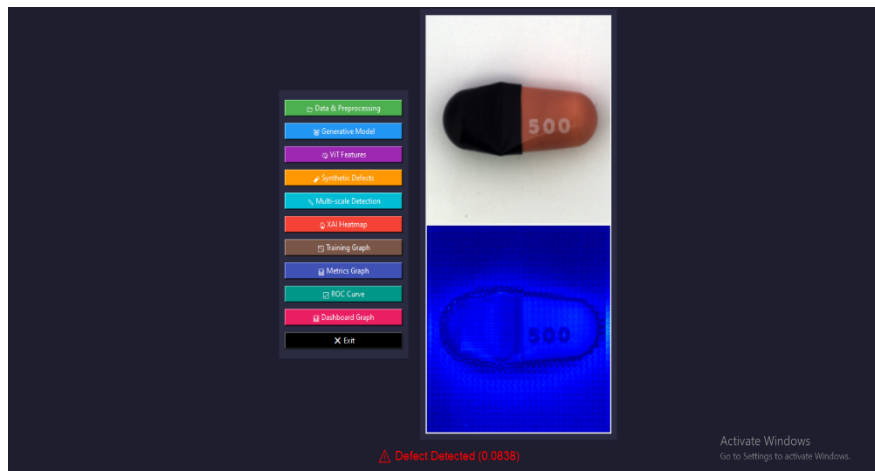


Figure 7: Defect Checking

The image displayed here is indicative of the results of the inspection process carried out by the AI model in relation to defects in the capsule. In this regard, the highlighted

image shown in blue indicates the position where the defect has been detected.

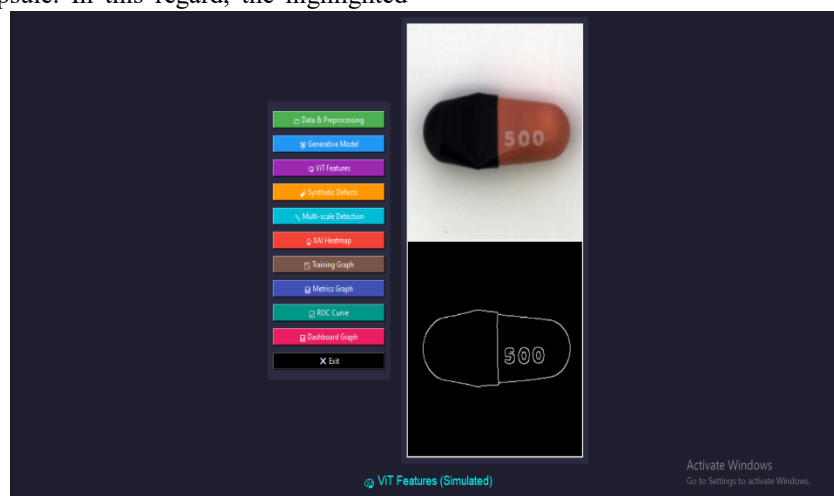


Figure 8: ViT Features

The above screen displays the process of extracting features using Vision Transformers (ViT) for the capsule defect detection model. The lower screen shows the

important edges and features that have been extracted from the capsule through the use of the AI model.

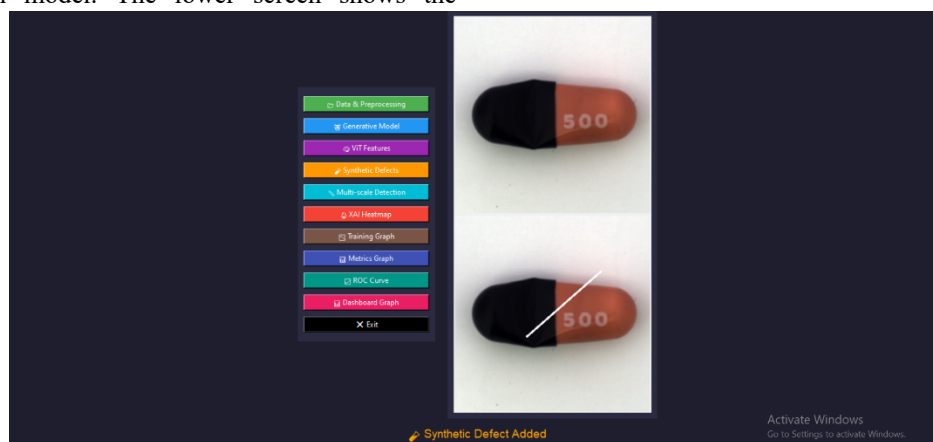
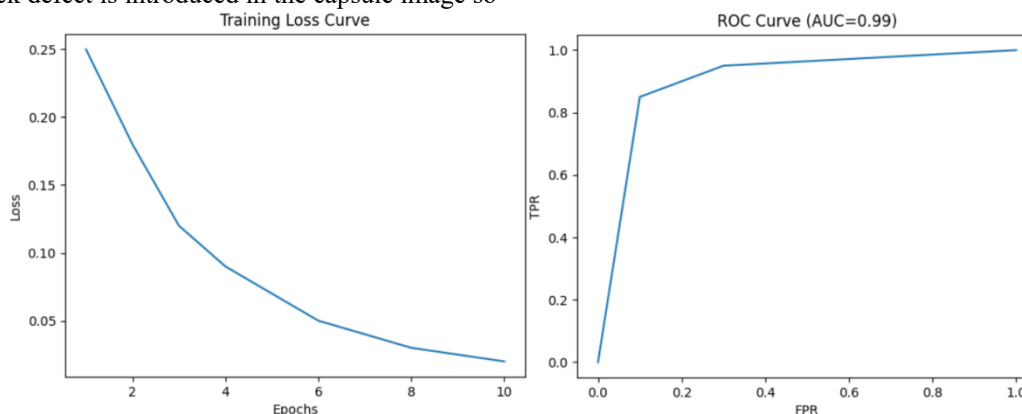


Figure 9: Synthetic Defects

This figure displays the synthetic defect generation stage of the pharmaceutical defect detection system. An artificial crack defect is introduced in the capsule image so

that the machine learning model can learn about defective capsules more accurately.

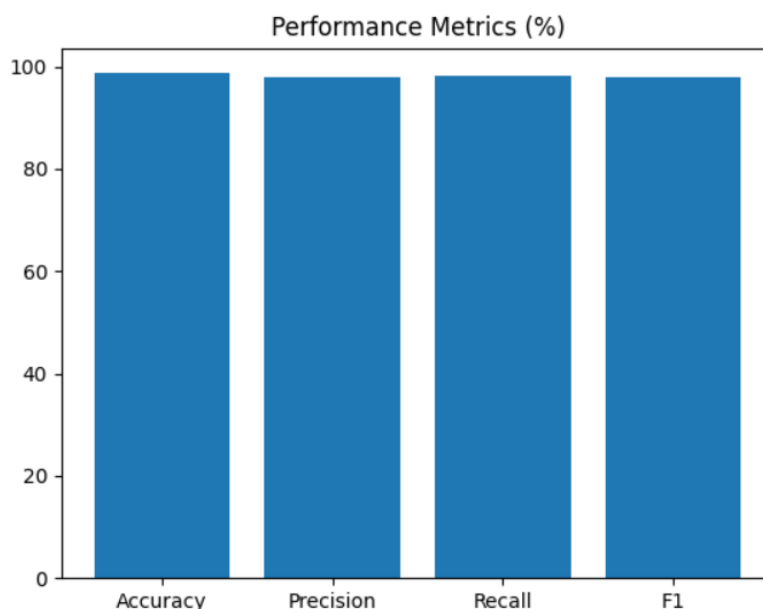


Graph 1: Accuracy and Loss Graph

The left graph shows the Training Loss Curve, where the loss is seen going downwards during training as the number of epochs increases. Initially, the loss is quite large since the model is in the process of learning about the typical patterns from the data set. As the process of training progresses, the loss decreases significantly, suggesting that the model has improved in accurately recreating the images without any defects. The steady decline is indicative of good learning behavior as well as

appropriate convergence without oscillations, implying stability without overfitting.

The graph on the right shows the Receiver Operating Characteristics (ROC) curve for the model, which indicates its ability to differentiate between defective and non-defective cases. There is a steep increase of the graph towards the upper left corner, while the area under the curve is estimated to be at around 0.99, indicating very good classification behavior.



Graph 2: Performance Metrix

CONCLUSION

As a result, the suggested Generative AI-Based Product Defect Detection System represents an innovative, intelligent, and efficient solution to product quality control challenges. Through applying generative models such as GANs or diffusion models, the system is able to learn the normal process flow of products without defects and recognize any abnormalities via reconstruction error,

thereby minimizing the need for a big database of defects. The addition of Vision Transformer (ViT) further contributes to the ability of the system to capture different levels of features, making it possible to recognize defects of more detailed and complicated patterns. Additionally, introducing artificial defects as part of the training process helps tackle the problem of limited samples of defects.

By implementing multi-scale detection, defects of different scales can be detected successfully; meanwhile, the Explainable AI (XAI) allows visualization in form of heat maps and therefore increases understandability. The combination of generative learning and transformer-based feature extraction makes it a powerful, accurate, and scalable defect-detection system. In particular, the system greatly reduces the necessity for manually labeled data and human inspection of defects, thereby meeting the needs of Industry 4.0 solutions. Regarding future development, it will be possible to apply it to real-time inspection and automatic correction of defects via integrating robots.

REFERENCES

- [1] H. Chen, Y. Zhang, X. Wang, X. Duan, Y. Zhou, and W. Zhu. Disenbooth: Disentangled parameter-efficient tuning for subject-driven text-to-image generation. *arXiv preprint arXiv:2305.03374*, 2023.
- [2] W. Chen, H. Hu, Y. Li, N. Rui, X. Jia, M.-W. Chang, and W. W. Cohen. Subject-driven text-to-image generation via apprenticeship learning. *arXiv preprint arXiv:2304.00186*, 2023.
- [3] Y. Duan, Y. Hong, L. Niu, and L. Zhang. Few-shot defect image generation via defect-aware feature manipulation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 571–578, 2023.
- [4] R. Gal, Y. Alaluf, Y. Atzmon, O. Patashnik, A. H. Bermano, G. Chechik, and D. Cohen-Or. An image is worth one word: Personalizing text-to-image generation using textual inversion. *arXiv preprint arXiv:2208.01618*, 2022.
- [5] M. Golam, A. Riztia, P. Daely, A. Khatun, J.-H. Kim, J.-H. Lee, Y.-R. Cho, D.-S. Kim, and J. M. Lee. Development of the system architecture for black ice detection to prevent transportation calamities. In *Proceedings of the Symposium of the Korean Institute of Communications and Information Sciences*, pages 692–693, 2022.
- [6] L. Han, Y. Li, H. Zhang, P. Milanfar, D. Metaxas, and F. Yang. Svdif: Compact parameter space for diffusion fine-tuning. *arXiv preprint arXiv:2303.11305*, 2023.
- [7] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. In *Advances in Neural Information Processing Systems*, volume 30, 2017.
- [8] J. Ho, A. Jain, and P. Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems*, volume 33, pages 6840–6851, 2020.
- [9] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models, 2020.
- [10] G. Huang, I. Laradji, D. Vazquez, S. Lacoste-Julien, and P. Rodriguez. A survey of self-supervised and few-shot object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(4):4071–4089, 2022.
- [11] M. A. Khatun, D.-S. Kim, J. M. Lee, and J.-H. Kim. Comparison between vision transformer and cnn for ice image classification. In *Proceedings of the Symposium of the Korean Institute of Communications and Information Sciences*, pages 1822–1823, 2023.
- [12] E. Levin and O. Fried. Differential diffusion: Giving each pixel its strength. *arXiv preprint arXiv:2306.00950*, 2023.
- [13] Hanxi Li, Zhengxun Zhang, Hao Chen, Lin Wu, Bo Li, Deyin Liu, and Mingwen Wang. A novel approach to industrial defect generation through blended latent diffusion model with online adaptation. *arXiv preprint arXiv:2402.19330v2*, Mar 2024. These authors contributed equally to this work.
- [14] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *Proceedings of the International Conference on Learning Representations (ICLR)*, 2017.
- [15] Adnan Md Tayeb and Tae-Hyong Kim. Unestformer: Enhancing decoders and skip connections with nested transformers for medical image segmentation. *IEEE Access*, 12:190996–191009, 2024.
- [16] C. Mou, X. Wang, J. Song, Y. Shan, and J. Zhang. Dragondiffusion: Enabling drag-style manipulation on diffusion models. *arXiv preprint arXiv:2307.02421*, 2023.
- [17] A. Nichol, P. Dhariwal, A. Ramesh, P. Shyam, P. Mishkin, B. McGrew, I. Sutskever, and M. Chen. Glide: Towards photorealistic image generation and editing with text-guided diffusion models. *arXiv preprint arXiv:2112.10741*, 2021.
- [18] A. Q. Nichol and P. Dhariwal. Improved denoising diffusion probabilistic models. In *Proceedings of the International Conference on Machine Learning*, pages 8162–8171. PMLR, 2021.
- [19] S. Niu, B. Li, X. Wang, and H. Lin. Defect image sample generation with gan for improving defect recognition. *IEEE Transactions on Automation Science and Engineering*, 17(3):1611–1622, 2020.
- [20] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer. Highresolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10684–10695, 2022.

- [21] N. Ruiz, Y. Li, V. Jampani, Y. Pritch, M. Rubinstein, and K. Aberman. Dreambooth: Fine tuning text-to-image diffusion models for subject-driven generation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 22500–22510, 2023.
- [22] J. Song, C. Meng, and S. Ermon. Denoising diffusion implicit models. arXiv preprint arXiv:2010.02502, 2020.
- [23] A. M. Tayeb, K.-H. Lee, J.-Y. Seo, G.-W. Kim, H.-B. Park, H.-J. Lee, and T.-H. Kim. Facial skin condition analysis based on pore and wrinkle segmentation. In Proceedings of the Symposium of the Korean Institute of Communications and Information Sciences, pages 1648–1649, 2023.
- [24] Adnan Tayeb, Md Alam, Ayesha Khatun, Golam Mohtasin, Md Facklasur Rahaman, Md Javed Ahmed Shanto, and Dong-Seong Kim. Pureparking: A decentralized, secure framework for parking space sharing using blockchain. 06 2024.
- [25] Adnan Md Tayeb, Mst Ayesha Khatun, Mohtasin Golam, Md Facklasur Rahaman, Ali Aouto, Oroceo Paul Angelo, Minseon Lee, Dong-Seong Kim, Jae-Min Lee, and Jung-Hyeon Kim. Smartrsd: An intelligent multimodal approach to real-time road surface detection for safe driving, 2024.
- [26] Adnan Md Tayeb, Hope Leticia Nakayiza, Heejae Shin, Seungmin Lee, Chaesoo Lee, YeongHun Lee, Dong-Seong Kim, and Jae-Min Lee. DefectGen: Few-Shot Defect Image Generation Using Stable Diffusion for Steel Surface Analysis. In 2024 International Conference on Information and Communication Technology Convergence (ICTC), pages 2092–2097. IEEE, 2024.
- [27] J. Wei, F. Shen, C. Lv, Z. Zhang, F. Zhang, and H. Yang. Diversified and multi-class controllable industrial defect synthesis for data augmentation and transfer. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pages 4444–4452, 2023.
- [28] L. Yang, Z. Zhang, Y. Song, S. Hong, R. Xu, Y. Zhao, W. Zhang, B. Cui, and M.-H. Yang. Diffusion models: A comprehensive survey of methods and applications. ACM Computing Surveys, 56(4):1–39, 2023.
- [29] G. Zhang, K. Cui, T.-Y. Hung, and S. Lu. Defectgan: High-fidelity defect synthesis for automated defect inspection. In IEEE/CVF Winter Conference on