Enhanced Conditional GAN-Augmented Deep Neural Network with Self-Attention for Diabetic Foot Ulcer Classification

Tejal Khade*, Chandra Prakash†
Computational Intelligence and Smart Motion Research Group (CISMR)
Department of Computer Science and Engineering
Sardar Vallabhbhai National Institute of Technology, Surat, India - 395007
*p23cs019@coed.svnit.ac.in, †cprakash@coed.svnit.ac.in

***Abstract*—Diabetes is one of the most prevalent diseases globally, significantly impacting the quality of life and often leading to severe complications. Among these, Diabetic Foot Ulcer (DFU) is one of the most severe complications, potentially leading to infections, amputations, or even death if not properly managed. Automated DFU classification remains a critical challenge due to class imbalance and limited dataset diversity. Traditional augmentation techniques often fall short in capturing the complex and diverse variations present in medical images. In this paper, we propose a novel classification framework that combines data augmentation using an enhanced Conditional Generative Adversarial Network (cGAN) with a self-attention-based dense neural network for accurate DFU classification. The proposed method addresses class imbalance in the DFUC2021 dataset by generating high-quality synthetic images for the minority class. Our enhanced cGAN demonstrates superior performance in image generation, achieving a Fréchet Inception Distance (FID) of 32.14, Peak Signal-to-Noise Ratio (PSNR) of 18.32, and Structural Similarity Index Measure (SSIM) of 0.451. The classification model attained 92.57% accuracy outperforming state-of-the-art models. This methodology aims to effectively tackle class imbalance and improves classification robustness.**

***Index Terms*—Diabetic Foot Ulcer (DFU), Medical Imaging, Class Imbalance, Data Augmentation, Conditional Generative Adversarial Network (cGAN), Self-Attention, Deep Learning, Convolutional Neural Network (CNN)**

I. INTRODUCTION

Diabetic Foot Ulcer (DFU) is a very serious health concern that can develop in people with diabetes, which currently affects over 537 million individuals around the world [9], frequently leading to infections, gangrene, and lower-limb amputations. These chronic wounds result from peripheral neuropathy, poor vascular circulation, and compromised immune response, representing a significant burden on healthcare systems worldwide. Early detection and classification of DFUs are crucial for preventing severe outcomes, yet manual screening remains time-consuming and requires specialized expertise that may be unavailable in resource-limited settings. Statistics reveal that individuals with diabetes face a 15-25% lifetime risk of developing foot ulcers, with approximately 85% of diabetes related amputations preceded by ulceration [8]. However, medical imaging datasets typically suffer from class imbalance and limited diversity in samples, which significantly impact the performance of deep learning models. Traditional augmentation methods offer limited diversity, as they only generate minor variations without introducing truly novel data for robust model training. Generative Adversarial Networks (GANs) have emerged as a powerful solution for synthetic data generation, comprising two competing neural networks: a generator that creates realistic synthetic samples and a discriminator that distinguishes between real and generated data[10]. This adversarial process drives continuous improvement in both networks, ultimately resulting in the generation of highly realistic synthetic data. Conditional GANs (cGANs) extend this framework by enabling controlled generation based on class labels, making them particularly suitable for addressing class imbalance in medical imaging applications[7, 10]. This study presents an enhanced Conditional Generative Adversarial Network (cGAN) architecture for DFU image augmentation combined with a novel CNN classification model integrated with self-attention mechanisms. The primary contributions include development of an enhanced cGAN architecture optimized for medical image synthesis, integration of synthetic data augmentation with a self-attention based CNN for improved classification accuracy and validation of generated synthetic images using state-of-the-art evaluation metrics.

II. RELATED WORK

Research in Diabetic Foot Ulcer (DFU) classification has advanced rapidly with the integration of deep learning and medical imaging. Hamghalam and Simpson introduced ESGAN and EnhGAN two cGAN based models for brain tumor segmentation in MRI scans [7]. ESGAN focused on class separability in limited data, while EnhGAN enhanced segmentation of complex regions. Though effective on BraTS'13 and BraTS'18, further refinement is needed for broader clinical use.

Waheed et al. and Amin et al. used GAN-based methods—ACGAN for chest X-rays and CGAN for CT images—to classify COVID-19 using the UCSD-AI4H dataset [10, 3]. While both achieved improved accuracy through synthetic augmentation, they faced limitations such as over-reliance on synthetic data and limited generalization to real clinical scenarios.

N. Bansal and A. Vidyarthi proposed DFootNet, a deep neural network integrating CNNs with residual blocks and feature fusion layers for DFU classification [4]. They applied traditional data augmentation like rotation, flipping, Gaussian noise, shearing, and translation—to boost model robustness. The model achieved 98.87% accuracy on Kaggle DFU datasets. However, the need for more diverse datasets, better hyperparameter tuning, clinical validation was a limitation.

Toofanee et al. and Qayyum et al. proposed hybrid Vision Transformer (ViT)-based frameworks with feature fusion and weighted loss to tackle class imbalance in DFU classification [9, 8]. While effective, both faced challenges like high computational cost, residual imbalance, and limited use of ensemble strategies.

Yap et al. evaluated advanced deep learning models such as EfficientDet, Cascade R-CNN with DetNet, and Faster R-CNN with deformable convolutions for DFU detection [12]. Data augmentation improved model robustness, yielding promising results on MS-COCO and DFUC2020 datasets. However, challenges like false positives and limited adaptability for remote patient monitoring were noted. Alzubaidi et al. [1, 2] proposed several hybrid CNN architectures for DFU classification using a dataset of 754 images. Their models utilized feature aggregation techniques such as Global Average Pooling (GAP) and dropout, and included DFU_QUTNet, which outperformed fine-tuned models like GoogleNet, AlexNet, and VGG16 when paired with an SVM classifier. Despite achieving improved accuracy, the approaches were constrained by the small dataset size, binary classification focus, limited generalizability, and lack of clinical validation. Yap et al. and Goyal et al. [11, 6] addressed infection and ischemia classification in DFUs using pre-trained CNNs with data augmentation, but struggled with class imbalance and poor image quality, suggesting GANs as a better alternative. Goyal et al. also introduced DFUNet, a CNN for ulcer classification, which outperformed LeNet and AlexNet with an AUC of 0.961, showing strong performance even with limited data [5].

Based on the reviewed literature, current DFU classification methods face several limitations, including small and non-diverse datasets [4, 1, 2], persistent class imbalance inadequately addressed by traditional augmentation [8, 11, 6], over-reliance on synthetic data with limited clinical generalization [3], emphasis on binary rather than multiclass classification [1, 8], insufficient clinical validation [4, 2], and challenges due to visual similarity between classes and data quality issues [6]. Addressing these gaps, our work focuses on mitigating class imbalance and enhancing image quality to improve DFU classification performance.

This paper is structured into five main sections: Section I introduces the DFU classification problem and challenges of class imbalance, motivating the use of enhanced cGANs with self-attention CNNs. Section II reviews related work in DFU classification and medical image synthesis. Section III presents proposed methodology for data augmentation and DFU classification. Section IV provides experimental results and Section V concludes with key contributions and future work directions, followed by acknowledgments and references.

III. PROPOSED METHODOLOGY

The proposed methodology shown in Figure 1 for Diabetic Foot Ulcer (DFU) classification adopts a structured pipeline encompassing dataset preparation, data preprocessing, augmentation, model training, and evaluation. This systematic workflow ensures robust and reproducible results.

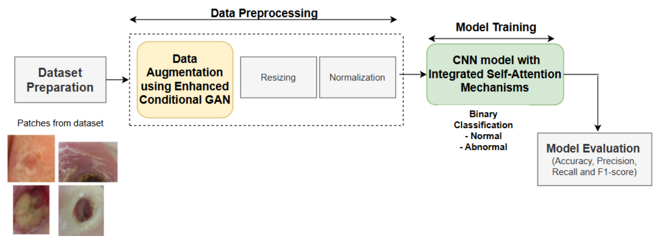

Fig. 1. Proposed Methodology Pipeline

A. Dataset

This study uses the DFUC2021 dataset, a comprehensive collection of diabetic foot ulcer (DFU) images, as the basis for classification and analysis. The dataset is divided into two subsets: Part A and Part B. This work focuses on Part A, which is used to train and evaluate models for binary classification between DFU affected skin patches and normal skin patches. As summarized in Table I, Part A includes 1,038 abnormal (ulcer) and 641 normal (non-ulcer) samples, highlighting a notable class imbalance.

TABLE I
DFUC2021(PART A) DATASET DESCRIPTION

Data Type	Internal Sub Division	Sample Count
PartA_DFU_Dataset	Abnormal	1038
	Normal	641

B. Data Pre-processing

The DFU dataset comprises image patches of varying sizes labeled as normal or abnormal. For uniform model input, images are resized to 224×224 and normalized to [-1,1]. They are then converted to tensors for training and back to NumPy arrays for visualization.

C. Data Augmentation

Data augmentation plays a crucial role in mitigating class imbalance by increasing the diversity and quantity of training samples for underrepresented classes. Severe class imbalance, as observed in the DFUC2021 dataset, poses a major challenge to deep learning models, often resulting in overfitting and poor generalization.[11] Traditional augmentation methods[4] often fail to capture the intricate variability in medical images. To

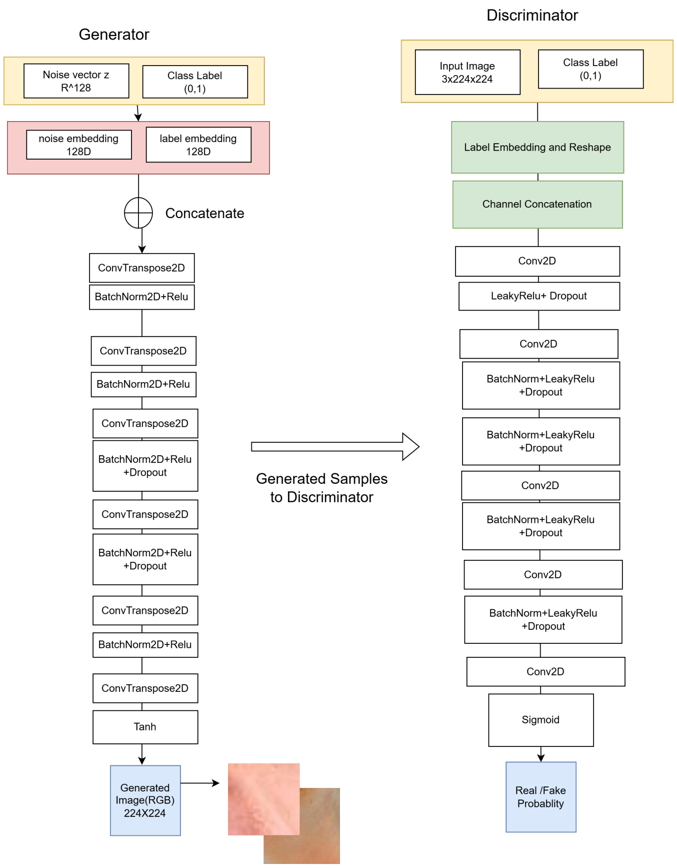

Fig. 2. Enhanced Conditional Generative Adversarial Network for synthetic image generation

address this limitation, we employ an enhanced Conditional Generative Adversarial Network (cGAN) to generate high quality images for the minority class (see Figure 2). cGAN consists of a generator and a discriminator that compete adversarially to generate realistic synthetic images [9].

1) Generator: The generator takes a 128-dimensional noise vector and a class label (0 or 1) as input. Both are embedded into 128-dimensional vectors and concatenated to form a 256-dimensional representation. This vector is passed through transposed convolutional layers that progressively upsample the spatial dimensions. Normalization and activation functions are applied at each stage. The output is a 224×224 RGB image with scaled pixel values suitable for image data.

2) Discriminator: The discriminator receives an input image of size 224x224×3 along with the corresponding class label. The label is converted into a spatial format and merged with the input image, resulting in an extended input with an additional channel. This input is then passed through a series of layers that reduce spatial dimensions while increasing the depth of feature representations. At each stage, normalization, non-linear activation, and dropout are used to ensure stable learning and prevent overfitting. The final output is a single value between 0 and 1, representing the probability that the input image is real or generated.

3) Generator Loss: The generator loss shows how well the generator is creating synthetic DFU images that look real[3]. A lower generator loss means the generated images are getting better at fooling the discriminator.

4) Discriminator Loss: The discriminator loss measures how well it distinguishes real DFU images from fake ones[3]. When both losses stabilize, it means the generator creates high-quality images and the discriminator still challenges them indicating successful model training. The loss for the generator (*G_Loss*) and discriminator (*D_Loss*) are defined as:

$$\text{G\_Loss}(z) = -\log(D(G(z))) \tag{1}$$

$$\text{D\_Loss}(x, z) = -\log(D(x)) - \log(1 - D(G(z))) \tag{2}$$

where x is a real data sample and z is a random input noise vector. The generator G aims to minimize its loss by fooling the discriminator D into classifying generated data as real, as shown in (1). The discriminator loss, defined in (2), has two objectives: correctly identifying real data x as real, $D(x) = 1$, and fake data $G(z)$ as fake, $D(G(z)) = 0$.

The adversarial training process, illustrated in Figure 2, enables the generation of high-quality synthetic samples that retain essential characteristics of medical images. By producing normal samples through class conditional synthesis, it effectively balances the dataset. The augmented dataset serves as the primary training set, allowing investigation into how synthetic augmentation mitigates class imbalance while preserving generalization to real world medical scenarios.

D. Image Quality analysis and Model Evaluation

The quality of synthetic images was evaluated using FID, PSNR, and SSIM [7]. Low FID and high PSNR and SSIM scores confirm that generated images closely resemble real DFU images, ensuring their suitability for training. Additionally, bar plots visualizing class distribution before and after augmentation highlights how cGANs effectively addressed class imbalance.

E. Proposed Deep Neural Network Architecture with Self Attention mechanism

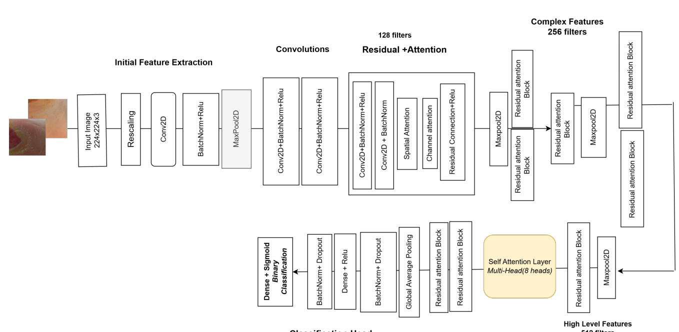

Fig. 3. Proposed CNN model with Integrated Self-Attention Mechanisms for binary classification

A novel architecture that combines a Convolutional Neural Network (CNN) with a Generative Adversarial Network

(GAN) to improve the classification of Diabetic Foot Ulcers (DFU) is shown in Figure 3. This hybrid model addresses two key challenges in medical image analysis. First, deeper CNNs can extract better features but may suffer from problems like vanishing gradients and overfitting, which reduce performance. Second, DFU datasets often have more abnormal cases than normal ones, causing class imbalance that affects learning. To solve this, the CNN is optimized for better feature extraction with average number of layers, and Enhanced cGAN generates high-quality synthetic images for the minority class. Together, they enhance both data diversity and the reliability of the classification model.

The deep neural network architecture is designed to efficiently process DFU images for classification tasks. It emphasizes deep convolutional layers, skip connections, bottleneck layers, and dense connections to optimize feature extraction while maintaining computational efficiency. Skip connections are effectively incorporated to facilitate efficient gradient flow across layers, thereby mitigating the vanishing gradient problem and enhancing the network's capacity to learn long-range dependencies. Dense connections further facilitate seamless information flow, improving gradient propagation and feature reuse. By integrating advanced components such as feature extraction blocks, residual layers, dense connections, and multi-filter operations, our proposed model captures features at multiple scales, enabling recognition of both fine-grained details and broader contextual information. Strategically placed residual layers further enhance the network's ability to extract subtle features from ulcer images, thereby improving model transparency and interpretability. The hyperparameters are fine-tuned and used during model training and testing. The details of these parameters, along with their experimental values, are presented in Table II.

TABLE II
HYPERPARAMETERS USED FOR CLASSIFICATION MODEL TRAINING

Hyperparameter	Value
Batch Size	32
Image Size (Height × Width)	224 × 224
Number of Classes	2
Epochs	200
Learning Rate	0.0001
Loss Function	Binary Cross Entropy (BCELoss)
Optimizer (Generator & Discriminator)	Adam ($\beta_1 = 0.5$, $\beta_2 = 0.999$)

The design layers of our proposed architecture is shown in Table III.

TABLE III
PROPOSED CNN-ATTENTION ARCHITECTURE OVERVIEW

Component	Layers	Key Features
Input Processing	3	Conv2D, BatchNorm, Resize
Feature Extraction	4	Conv2D blocks, MaxPool
Attention Modules	3	Spatial, Channel, Residual
Self-Attention	1	8-head attention
Complex Features	2	Residual + Attention
Classification	4	Dense, Dropout, GAP

IV. EXPERIMENTAL RESULTS

This section presents results on the DFUC2021 dataset, highlighting the impact of enhanced cGAN-based augmentation on class imbalance and evaluating the performance of our CNN with self-attention classification model.

A. *Class Distribution in Dataset after Data Augmentation*

TABLE IV
DFUC2021 DATASET: IMPACT OF DATA AUGMENTATION

Data Type	Class	Samples Before	Samples After
PartA_DFU	Abnormal	1038	1038
	Normal	641	1024

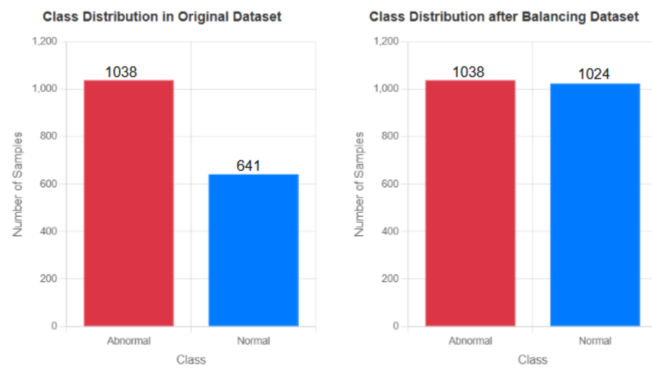

Fig. 4. Class Distribution of DFUC2021 dataset before and after augmentation

Class distribution before and after applying Enhanced cGAN for data augmentation is shown in Figure 4.To address this class imbalance problem, we employed enhanced Conditional GAN to synthesize additional normal samples for creating a balanced training dataset. The dataset was then divided into training and validation subsets where 70% of the data was allocated for training the model, while the remaining 30% was reserved for validation. This improves the model's ability to learn discriminative features for both classes equally. As a result, the classification performance particularly in terms of recall and F1-score shows significant improvement.

B. *Quantitative Evaluations for Generated Data*

1) Discriminator accuracy: Figure 5 shows discriminator accuracy which measures the ability of the GAN discriminator to correctly classify images as real or fake. Over time, as the generator improves, the discriminator's accuracy fluctuates and stabilizes around 76%, indicating a balanced adversarial training. This reflects the generator's ability to produce realistic synthetic images that effectively challenge the discriminator.

2) Image Quality Improvement: Figure 6 illustrates a visual comparison of CGAN and Enhanced CGAN augmentation. In both subfigures, the top row displays original images, while the bottom row presents the corresponding augmented samples. The Enhanced CGAN (right) demonstrates progressively improved image quality, highlighting its effectiveness in generating more realistic synthetic data compared to the standard CGAN (left).

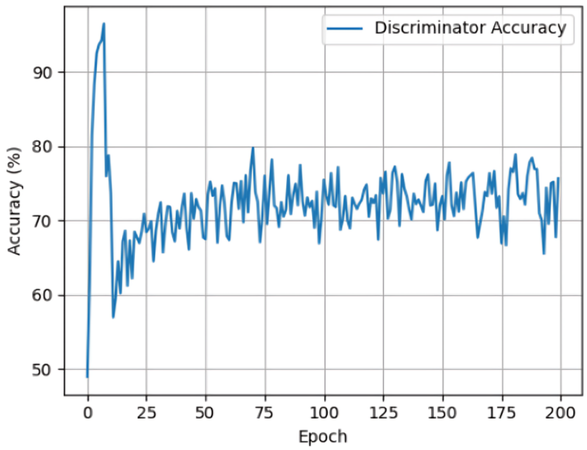

Fig. 5. Discriminator accuracy during Enhanced CGAN training.

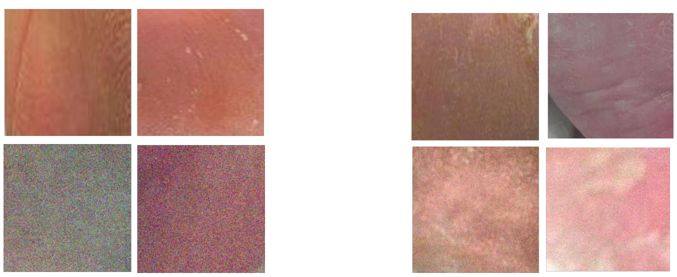

Fig. 6. Samples after augmentation using Conditional GAN (left) in second row and Enhanced Conditional GAN (right) in second row.

Table V presents the quantitative evaluation of generated image quality and diversity using standard metrics: Peak Signal-to-Noise Ratio (PSNR), Fréchet Inception Distance (FID), and Structural Similarity Index Measure (SSIM). PSNR (Equation (3)) quantifies reconstruction fidelity, FID (Equation (4)) measures distribution alignment between real and synthetic data, and SSIM (Equation (5)) captures perceptual similarity in terms of luminance, contrast, and structure.

$$\mathrm{PSNR} = 10 \cdot \log_{10}\left(\frac{MAX_I^2}{\mathrm{MSE}}\right) \tag{3}$$

where: MAX_I is the maximum possible pixel value, MSE is the Mean Squared Error between the original and generated images.

$$\mathrm{FID} = \|\mu_r - \mu_g\|^2 + \mathrm{Tr}\left(\Sigma_r + \Sigma_g - 2\left(\Sigma_r \Sigma_g\right)^{1/2}\right) \tag{4}$$

where:μ_r, Σ_r are the mean and covariance of real image features, μ_g, Σ_g are the mean and covariance of generated image features, and Tr denotes the trace of the matrix.

TABLE V
IMAGE QUALITY ANALYSIS USING CGAN AND ENHANCED CGAN

GAN Method	FID ↓	PSNR ↑	SSIM ↑
CGAN	42.40	13.01	0.176
Enhanced CGAN	32.14	18.32	0.451

$$\mathrm{SSIM}(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{5}$$

where:μ_x, μ_y are the means of images x and y, σ_x^2, σ_y^2 are the variances of x and y, σ_{xy} is the covariance between x and y, C_1, C_2 are constants to stabilize the division.

3) Performance Metrics of Classification Model: Model performance is evaluated across standard metrics such as accuracy, precision, recall, and F1-score.

- **Accuracy:** The proportion of correctly classified samples across all predictions. For the proposed model, an accuracy of 92.57% was achieved.
- **Precision and Recall:** Precision measures the proportion of true positives among all positive predictions, while recall evaluates the model's ability to correctly identify actual positives.

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives + False Positives}} = 0.9061$$

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives + False Negatives}} = 0.9204$$

- **F1-score:** The harmonic mean of precision and recall, balancing the trade-off between them.

$$\mathrm{F1} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} = 0.9373$$

4) Confusion Matrix Analysis: The figure 7 demonstrates strong performance on the validation dataset, with 287 true positives correctly identified abnormal cases. It also achieved 176 true negatives, correctly identifying normal cases. There were 17 false positives where normal cases were misclassified as abnormal, and 25 false negatives where actual abnormal cases were missed.

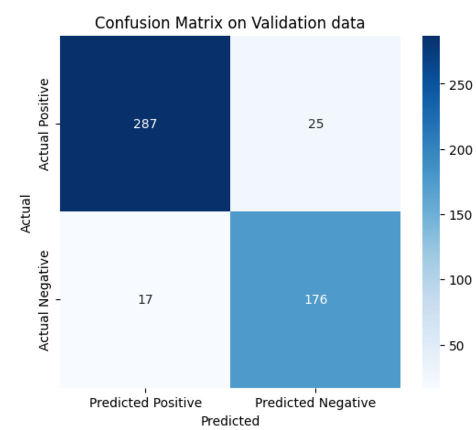

Fig. 7. Confusion Matrix for binary classification

C. Comparison with State-of-the-Art Models

Table VI presents the performance comparison between state-of-the-art and the proposed model. The proposed model achieves the highest F1 score, demonstrating its effectiveness in DFU classification.

TABLE VI
PERFORMANCE COMPARISON WITH STATE-OF-THE-ART MODELS

Model	Accuracy	Precision	Recall	F1 score	AUC
LeNet [5]	0.8752	0.8610	0.9012	0.8806	0.9541
AlexNet [5]	0.8873	0.8924	0.9028	0.8921	0.9603
GoogleNet [5]	0.9024	0.8857	0.9101	0.9073	0.9685
EfficientNetB0 [8]	0.9251	0.8923	0.9502	0.9204	0.9812
DenseNet121 [8]	0.9186	0.8881	0.9445	0.9154	0.9787
ResNet101 [11]	0.9202	0.8896	0.9470	0.9174	0.9793
InceptionV3 [11]	0.9114	0.8807	0.9362	0.9075	0.9745
VGG16 [11]	0.9069	0.8769	0.9310	0.9032	0.9723
Proposed Model	**0.9257**	**0.9061**	**0.9204**	**0.9373**	**0.9431**

D. Receiver Operating Characteristic(ROC) Curve Analysis

ROC curve shown in Fig 8 presents a comprehensive comparison of the proposed model with state-of-the-art models. It helps visualize trade-offs between sensitivity and specificity of our proposed classification model and state-of-the-art models.

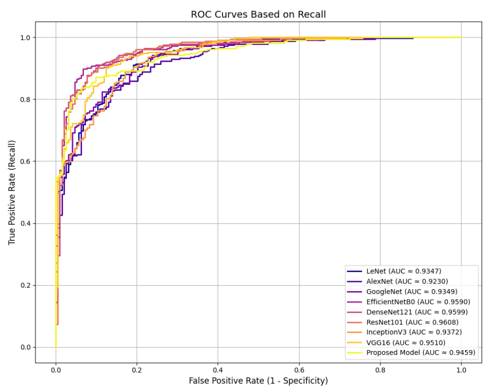

Fig. 8. Area under curve for state-of-the-art models and proposed model

V. CONCLUSION AND FUTURE WORK

This study proposes a deep learning approach combining an enhanced conditional GAN for data augmentation with a self-attention–based CNN for diabetic foot ulcer (DFU) classification. By addressing class imbalance through synthetic image generation, the method improves classification accuracy and robustness. The model achieves strong performance in binary classification, demonstrating potential for medical image analysis. While results are promising, further refinement is still possible.

In future, we plan to extend the methodology to multiclass classification and other medical imaging domains. Enhancements to the generative component will focus on producing higher-quality synthetic data across varied skin lesion categories. Furthermore, we aim to explore real-world deployment of the framework for reliable, automated image analysis in clinical settings.

VI. ACKNOWLEDGEMENT

We gratefully acknowledge the support provided by DFU Dataset Organisers and Science & Engineering Research Board (SERB), Department of Science and Technology, India, through project grant EEQ/2022/000561.

REFERENCES

[1] L. Alzubaidi et al. "Comparison of Hybrid Convolutional Neural Network Models for Diabetic Foot Ulcer Classification". In: *Journal of Engineering Science and Technology* 16.3 (2021).

[2] L. Alzubaidi et al. "DFU_QUTNet: diabetic foot ulcer classification using novel deep convolutional neural network". In: *Multimedia Tools and Applications* 79.21 (2020), pp. 15655–15677. DOI: 10.1007/s11042-019-07820-w.

[3] J. Amin et al. "Quantum Machine Learning Architecture for COVID-19 Classification Based on Synthetic Data Generation Using Conditional Adversarial Neural Network". In: *Cognitive Computation* 14.5 (2022), pp. 1677–1688. DOI: 10.1007/s12559-021-09926-6.

[4] N. Bansal and A. Vidyarthi. "DFootNet: A domain adaptive classification framework for diabetic foot ulcers using dense neural network architecture". In: *Cognitive Computation* 16.5 (2024), pp. 2511–2527. DOI: 10.1007/s12559-024-10282-4.

[5] M. Goyal et al. "DFUNet: Convolutional Neural Networks for Diabetic Foot Ulcer Classification". In: *arXiv preprint* arXiv:1711.10448v2 (2017). DOI: 10.48550/arXiv.1711.10448.

[6] M. Goyal et al. "Recognition of ischaemia and infection in diabetic foot ulcers: Dataset and techniques". In: *Computers in Biology and Medicine* 117 (2020), p. 103616. DOI: 10.1016/j.compbiomed.2020.103616.

[7] M. Hamghalam and A. L. Simpson. "Medical image synthesis via conditional GANs: Application to segmenting brain tumours". In: *Computers in Biology and Medicine* 170 (2024), p. 107982. DOI: 10.1016/j.compbiomed.2024.107982.

[8] A. Qayyum et al. "Efficient Multi-model Vision Transformer Based on Feature Fusion for Classification of DFUC2021 Challenge". In: *Diabetic Foot Ulcers Grand Challenge*. Springer International Publishing, 2022, pp. 62–75. DOI: 10.1007/978-3-030-94907-5_5.

[9] M. S. A. Toofanee et al. "DFU-SIAM: A Novel Diabetic Foot Ulcer Classification With Deep Learning". In: *IEEE Access* 11 (2023), pp. 98315–98332. DOI: 10.1109/ACCESS.2023.3312531.

[10] A. Waheed et al. "CovidGAN: Data Augmentation Using Auxiliary Classifier GAN for Improved Covid-19 Detection". In: *IEEE Access* 8 (2020), pp. 91916–91923. DOI: 10.1109/ACCESS.2020.2994762.

[11] M. H. Yap et al. "Analysis Towards Classification of Infection and Ischaemia of Diabetic Foot Ulcers". In: *2021 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI)*. Athens, Greece, 2021, pp. 1–4. DOI: 10.1109/BHI50953.2021.9508563.

[12] Moi Hoon Yap et al. "Deep learning in diabetic foot ulcers detection: A comprehensive evaluation". In: *Computers in Biology and Medicine* 135 (2021), p. 104596. DOI: 10.1016/j.compbiomed.2021.104596.