



OPEN A feature explainability-based deep learning technique for diabetic foot ulcer identification

Pramod Singh Rathore¹, Abhishek Kumar³, Amita Nandal^{2✉}, Arvind Dhaka¹ & Arpit Kumar Sharma¹

Diabetic foot ulcers (DFUs) are a common and serious complication of diabetes, presenting as open sores or wounds on the sole. They result from impaired blood circulation and neuropathy associated with diabetes, increasing the risk of severe infections and even amputations if untreated. Early detection, effective wound care, and diabetes management are crucial to prevent and treat DFUs. Artificial intelligence (AI), particularly through deep learning, has revolutionized DFU diagnosis and treatment. This work introduces the DFU_XAI framework to enhance the interpretability of deep learning models for DFU labeling and localization, ensuring clinical relevance. The framework evaluates six advanced models—Xception, DenseNet121, ResNet50, InceptionV3, MobileNetV2, and Siamese Neural Network (SNN)—using interpretability techniques like SHAP, LIME, and Grad-CAM. Among these, the SNN model excelled with 98.76% accuracy, 99.3% precision, 97.7% recall, 98.5% F1-score, and 98.6% AUC. Grad-CAM heat maps effectively identified ulcer locations, aiding clinicians with precise and visually interpretable insights. The DFU_XAI framework integrates explainability into AI-driven healthcare, enhancing trust and usability in clinical settings. This approach addresses challenges of transparency in AI for DFU management, offering reliable and efficient solutions to this critical healthcare issue. Traditional DFU methods are labor-intensive and costly, highlighting the transformative potential of AI-driven systems.

Keywords LIME, Heat Map, AI, DL, Diabetic Foot Ulcer

The curative condition known as diabetes is also known as Diabetes Mellitus (DM). Hyperglycemia, or abnormally high blood sugar levels, is the determinant of diabetes¹. Lower limb amputation, cardiovascular disease, blindness, and severe defeat are some of the severe and perhaps deadly consequences that could result from this. Diabetic foot ulcers (DFUs) are open sores that may develop on a diabetic person's foot [2²⁵. It is important to note that over 80% of people with diabetes live in weaker nations, which may be defined by a lack of resources and expertise in the field of healthcare³. Additionally, between 15 and 25% of people with diabetes may eventually develop DFU⁴, which might lead to lower limb amputation and other complications⁵.

More than one million people with diabetes who have “dangerous feet” have lost a limb each year owing to a lack of education and insufficient treatment⁶. Taking extra precautions, such as washing their hands often, taking their medication as prescribed, and seeing their doctor regularly, is essential for people with diabetes⁷. The financial burden will therefore fall heavily on families and people with DM, especially in less developed countries⁸. DFUs are on the rise and will likely continue to do so as the number of persons with diabetes increases⁹. The limited resources and expertise aimed at finding a solution will be further taxed as a result of this. An important limb, such as a foot, is lost by about one million diabetics a year as a result of inadequate treatment¹⁰.

Medical professionals need to examine patient records thoroughly to arrive at a correct diagnosis. The use of computer-aided diagnostics techniques can reduce expenses and increase efficiency, in contrast to traditional diagnostic processes that are more prone to mistakes. These days, wearable medical and mobile technology could help manage diabetes and its complications by keeping an eye on harmful foot stress and inflammatory reactions, and by making remission last longer^{11,12}. Improved quality of life for patients is another benefit. A safe, remotely deployable, very dependable, and reasonably cost-automated DFU diagnosis approach is desperately needed. Researchers have made great strides in the field of Deep Learning (DL)¹³ and visual Computer Vision

¹Department of Computer and Communication Engineering, Manipal University Jaipur, Jaipur, India. ²Department of IoT and Intelligent Systems, Manipal University Jaipur, Jaipur, India. ³Department of CSE, Chandigarh University, Punjab, India. ✉email: amita.nandal@jaipur.manipal.edu; amita_nandal@yahoo.com

(CV) applications, which have made intelligent systems, including DFU illness detection, a very active and exciting place to be. As a consequence of modern clinical practice, the DFU test now incorporates many crucial steps for early diagnosis¹⁴. The medical history of the patient is reviewed; a specialist in diabetic foot care or wounds examines the DFU thoroughly; further imaging studies, including an MRI, CT scan, or X-ray, may be necessary¹⁵.

Significant progress has been made in the area of biomedicine in recent years. With breakthroughs in areas such as protein structure identification, medical picture classification, and genomic sequence alignment standing out. The extraordinary increase achieved in highly advanced electronics may be responsible for these advancements¹⁶. The massive volume of biological data necessitates trustworthy and efficient computational systems for processing, analysis, and storage¹⁷. Methods centred on deep education can tackle these complex challenges¹⁸.

The use of DL architectures, and more especially Siamese Neural Network's (SNN's), might greatly improve the efficiency with which novice data can acquire hierarchical deep knowledge characteristics. This gives these algorithms the ability to understand intricate biological data networks and patterns. As a result of increased funding and attention, DL has grown substantially and found widespread use across industries^{19,20}. Due to its ability to examine large amounts of data, identify appropriate patterns, and create educated estimates, DL approaches have become more important in the biological domain. In the emerging field of digital healing, medical images play a crucial role in the diagnosis of various energy problems. To achieve their goals in healthcare, traditional Machine Learning (ML) and DL algorithms rely on identifying and engaging qualities that are highly affected by facial forms, colours, and sizes²¹. Several research has shown that algorithms based on ML and DL can accurately identify and classify DFUs.

Many researchers have looked at these complicated problems, but thus yet, no satisfactory solutions have emerged. However, due to the complex and opaque nature of many DL and ML approaches, it may be difficult to include the meaning behind the model's indications or assessments and to discover certain errors or biases. The problem with black box models is that they are too sensitive to hyperparameter choices, which makes them difficult to tune and adapt to new data. Black box models may be impacted by overfitting, which leads to poor performance when applied to fresh data²². There is a rationale for the limited interpretability and transparency of black box models, even though they have achieved significant gains in numerous applications. By making black box methods more transparent and clearer, Explainable Artificial Intelligence (XAI) could solve these problems.

Despite their continued explicability, DL models have mostly been used in the judgement of DFU. Problems with bias, explainability, and honesty in DL models, data isolation when testing DL models, and security and burden when using DL models for DFU reasoning are all examples of these challenges²³. This is something that XAI can handle. Utilising the technology housed at XAI, black box DL models have the potential to improve white box models. White box DL models like this one are becoming more popular as a result of their success in Artificial Intelligence (AI) based systems. These options are limited since robots cannot communicate with humans or explain their results. Younger medical professionals and physicians benefit from XAI-based DL models because they make these methods easier to understand, use, and engage with. The suggested XAI-based DFU_XAI framework solves the issue of comprehensible DL models for applications like DFU analysis.

To facilitate the diagnosis and classification of DFUs, this study established a method known as DFU_XAI. After testing many AI models, they discovered that SNN accurately identified ulcers at 98.5%. Afterwards, the researchers used specialised methods to see inside the artificial intelligence's "black box" and observe its decision-making process. Weighted Gradient Activation Mapping for Classes (Grad-CAM) technology, Shapley Additive Explanations (SHAP), and Locally Interpretable Explanations and Model-Independent (LIME) are three ways that assist in illustrating which areas of a picture the AI employs as its primary emphasis for ulcer identification. Instead of developing a ML system that only provides answers without any context, the aim was to develop one that physicians could trust and comprehend. It helps physicians diagnose patients more accurately by illuminating the AI's thought process²⁴. This method is not without its difficulties, however, such as the AI models intricacy and the volume of intricate data they handle.

Endocrinologists and AI experts work together in the DFU_XAI outline to further understanding of DFUs. AI models that include explanations enhance the accuracy of forecasts. Because it is easier for physicians to understand the model's decisions, explainability helps to develop trust among them. Compared to the present dataset, DFU_XAI will collect more data throughout the broader medical imaging community²⁵. With this information, the DFU_XAI system can refine its predictions for the early detection of DFU²⁶. The model and its predictions may be better understood and trusted with the help of the following key contributions²⁷.

The paper is organized into several key sections, each addressing specific aspects of the research. Following the introduction, a comprehensive literature review examines existing studies on DFU detection and classification, presenting diverse perspectives from various researchers. The methodology section outlines the core components of the study, including Convolutional Neural Network (CNN) models, image annotation techniques, data augmentation strategies, and the proposed XAI framework for enhancing DFU classification accuracy. To provide transparency in DL model predictions, three XAI approaches are detailed: LIME, SHAP, and Grad-CAM. The paper then describes the experimental design and presents a thorough analysis of the results, comparing them with current benchmarks. The final sections offer a discussion of the findings and their implications, concluding with a summary of the research contributions and potential future directions.

Literature review

DFUs may now be categorised in several ways. The existing corpus of research is examined in a literature review. A synopsis of the main sources that back up the argument is given in this section. Da Costa Oliveira et al.¹² presented the advantages and disadvantages of utilizing AI to diagnose ulcers in remote patients in a thorough literature review. Patient physiology and sensor characteristics informed their choice of imaging modalities and

optical sensors for DFU detection. According to the research, the monitoring technique is determined by the data source, which in turn influences the AI algorithms used.

In their most recent publication, Toofanee et al.¹³ researcher led them to propose DFU_SPNet, a CNN architecture that significantly improves DFU classification performance metrics. The three convolution layers in this network are layered and run in parallel, and they use kernel sizes that vary. Between each convolutional layer, this design employs a batch normalisation and leaky RELU activation transition layer. A limitation of the research is that the DFU_SPNet model did extremely well, although it employs the identical filter in parallel processing blocks. Changing the number of filters might improve the results.

For DFU classification, a hybrid Network was used, as stated by Tulloch et al.¹⁴. By merging conventional convolutional layers with multi-branched concurrent ones, their approach produced four DNS with six simultaneous convolution modules. Depending on the array of convolutional layers in a parallel network, each network was categorised accordingly. Combining branches with varying filter widths might provide a more accurate feature map. The model works well, but the processing time and expenses are raised because of all the parallel convolution blocks that are needed.

Alzubaidi et al.²⁰ developed a specialized deep-learning model called DFU_QUTNet. This CNN was specifically created to categorize images of DFUs. The model aims to assist in the classification and management of DFU imagery, potentially aiding in diagnosis and treatment planning for this serious complication of diabetes. The network's architecture prioritised expanding in width while maintaining depth, similar to current networks. To train SVM and KNN classifiers, DFU_QUTNet features are used. As a result, the F1 score rose significantly. One drawback of this method is that all of the parallel convolutional layers must have the same kernel size (3×3). Increasing precision is possible with convolutional kernels of varying sizes.

One deep CNN approach was created by Lo et al.¹⁶ by using parallel convolution blocks, DFUNet was able to combine data from both individual convolution stages and whole convolution processes. The linde-buzo-gray method was also used to derive low-level features for DFU prediction precision. Classical LeNet, GoogleNet, and AlexNet are examples of high-level extractors. ML algorithms were trained to identify DFUs using these extracted properties. Without transition layers amongst parallel units, this design could not work. With appropriately built transition layers, a multiple-layer concatenation feature map has the potential to extract additional features. In their two-step process for determining the DFU region. Hela et al.¹⁷ utilized a picture-capturing box in conjunction with SVM classification. The picture was segmented using superpixels and retrieved characteristics in a two-stage classification process.

To separate skin and deep foot units in full-foot pictures, Manu et al.¹⁸ created a two-scale convolutional network design for transfer learning. The DSC against neighbouring skin was $0.851 (\pm 0.148)$, against DFU it was $0.794 (\pm 0.104)$, and against both, it was $0.899 (\pm 0.072)$ in this new design. Though effective, the method is not without its drawbacks. Problematically, it has trouble handling really big datasets. The practice of patients placing their foot on a box during data collection is strictly forbidden in hospitals due to infection control concerns. Although more precise, the earlier proposed studies are less transparent, interpretable, and explicable. Bilal et al.³⁴ proposed NIMEQ-SACNet, a self-attention-based precision medicine model that uses deep learning and image data for the detection of vision-threatening diabetic retinopathy, which can improve diagnostic precision. Wei et al.³² used Mendelian randomization strategies to explore causal relationships between type 2 diabetes and neurological disorders, showing how useful AI-based models may be in understanding complex comorbidities. Hu et al.²⁹ worked on medical device product innovation choices, presenting visual explanations for CNN-based predictions. This technique helps highlight critical areas in medical images, which helps improve the decision-making process of clinicians. Bilal et al.³⁶ developed EdgeSVDNet, a 5G-enabled model for the automatic detection of diabetic retinopathy using U-Net. This model demonstrated how explainable AI approaches can make medical imaging models more reliable by increasing transparency.

Bilal et al.³⁸ highlighted the role of AI-based U-Net architectures in diabetic retinopathy detection, which provides visual explainability and increases clinicians' confidence in AI systems. Cheng et al.³⁰ proposed a promising therapy for diabetic encephalopathy that helps classify diabetic foot ulcers by inhibiting hippocampal ferroptosis. This research emphasizes the role of hybrid models, which help improve explainability and classification accuracy.

The problem that current DL frameworks do not have the explainability and reliability that doctors require to trust DL algorithms during clinical operations is what prompted this study. Table 1 summarises the article.

Methodology

This study employs a model that has been previously trained to extract visual data from diabetic foot photos and uses DL to distinguish between healthy skin and ulcers. CNN, the algorithm uses heat maps or segmentation of regions to forecast ulcers. The DFU_XAI classification system is shown in Fig. 1. It incorporates three XAI approaches, six fine-tuned deep CNN models for model training, data augmentation, and performance measures. XAI-based transparent approach aids in finding biases and reducing AI model biases. A lack of feature explainability, an imbalanced dataset, and a poor supply of DFU datasets are some of the challenges. Here are several solutions: an augmentation method, a dataset balancer using SMOTE, and a visual attribute explanation using a combination of XAI and ResNet50 algorithms. Using the SNN technique, compare two photos of the same foot, and then use contrastive loss to identify which two images are similar and which are different. To improve the model's ability to make decisions and enhance confidence and interpretability in clinical situations, explainability methodologies like ResNet, Grad-CAM technology, and SHAP values are included.

In contrast to LIME, which finds perturbation, SHAP employs a gradient explanation. Grad-CAM generates scores using gradients. Disparities in feature description, imbalance in the dataset, and lack of access to the DFU dataset were the three main problems in implementing the DFU_XAI framework. To address these concerns, add more data, use SMOTE to even out the dataset, and combine ResNet50 with three XAI algorithms—one

Paper	Year	Strength/Outcome	Weakness
Lohaj et al. ⁴	2024	Transfer learning approach to classify DFU with improved accuracy	Requires large annotated datasets for training
Goyal et al. ⁸	2018	Decrease the model's depth while enlarging the PCB filter dimensions.	Insufficient TL's between PCB
Nagaraju et al. ¹⁰	2023	Multi-scale CNN approach for DFU localization and identification	Lack of robustness across varied ulcer presentations
Elmannai et al. ¹⁷	2023	Integrated image segmentation and classification with improved real-time accuracy	Limited validation on diverse datasets
Das et al. ¹⁹	2023	Arrange convolutional layers in parallel, tailored for transfer learning	Ensure uniform filter sizes (32, 64, and 128) across each stacked PCB.
Alzubaidi et al. ²⁰	2021	Integrate multi-branch parallel with traditional convolutional layers	Computation time
Arumuga Maria Devi et al. ²¹	2024	Enhance the model's width while minimizing computational expenses	Apply consistent 3 × 3 filter dimensions across the PCB
Wang et al. ²²	2020	Detect the wound on a sample obtained from a boxed capture of ulcerated skin.	Little running speed
Shichai, et al. ²³	2024	Predict segmentation of DFU samples at the pixel level.	Small data set
Kim et al. ²⁸	2023	Hybrid CNN-LSTM model for DFU detection, achieving high sensitivity and specificity	Increased training time due to model complexity

Table 1. Summary of existing DFU detecting methods. PCB: Parallel, TL: Transfer Learning; CNN: Convolutional Neural Network; DFU: Diabetic foot ulcers; CNN-LSTM: Convolutional Neural Network-Long Short Term Memory.

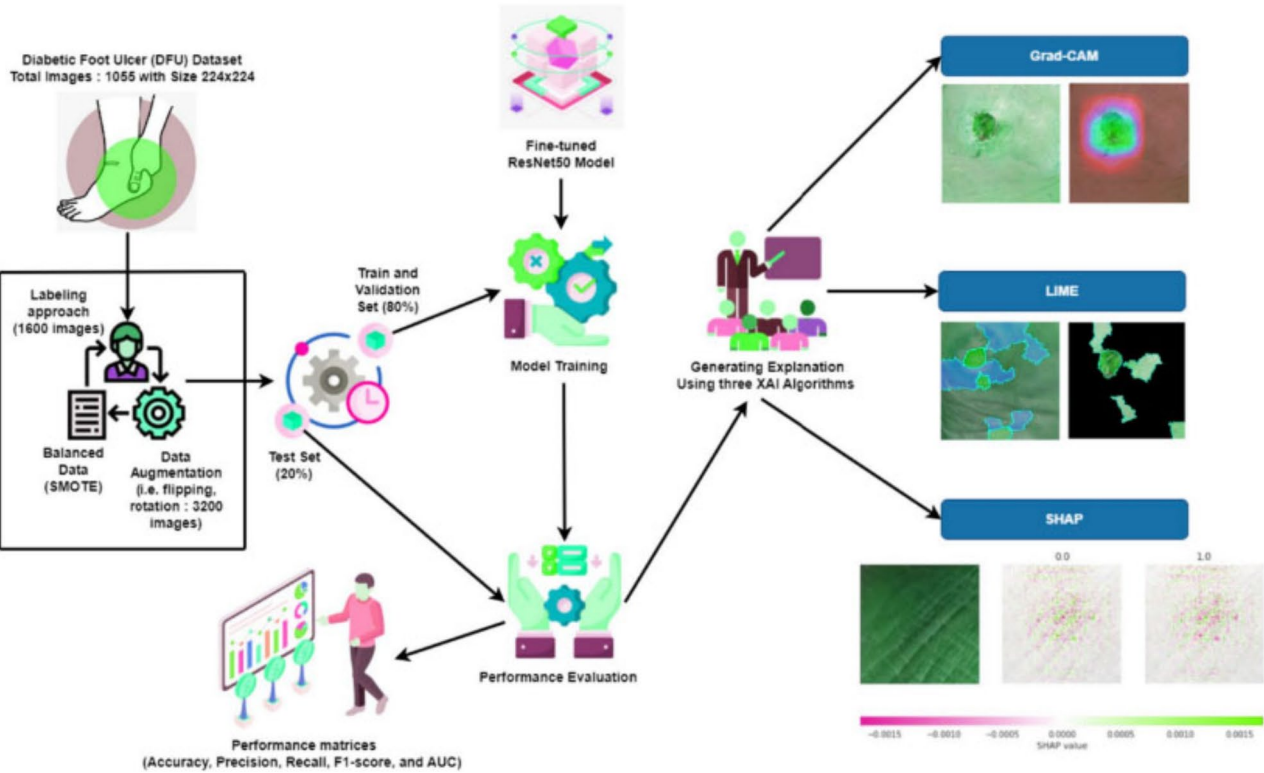


Fig. 1. DFU_XAI Framework.

of which is SNN—to make visual features more understandable. The experimental design process and outcome computation is shown in the DFU_XAI framework displays the sequential categorisation in Algorithm 1.

The sequential categorisation process of Algorithm 1 in the DFU_XAI framework explainability techniques like SHAP, LIME, and Grad-CAM within the framework. Describe how these tools enhance the interpretability of model predictions and contribute to clinical decision-making.

Explainability techniques SHAP, LIME, and Grad-CAM

SHAP (Shapley Additive Explanations)

- Objective: Quantifies the contribution of each input feature to model predictions, providing local explanations.

- **Clinical Impact:** Helps clinicians validate predictions by indicating which regions of the image are influencing the classification. Positive/negative contributions help align model predictions with clinical expectations and build trust.

LIME (Locally Interpretable Model-Agnostic Explanations)

- **Objective:** Breaks the image into superpixels, then perturbs them to see their effect on predictions.
- **Clinical Impact:** Highlights relevant regions (such as ulcer sites) that are useful for verification by clinicians. This helps identify biases or errors and increases clinicians' confidence in the model.

Grad-CAM (Gradient-weighted Class Activation Mapping)

Purpose: Generates heatmaps that indicate which areas of the image are critical for the model's prediction, by backpropagating gradients.

Clinical impact: Provides intuitive visualizations that indicate where the model is focusing. This aids in accurate ulcer localization, which is useful for diagnosis and treatment planning.

Objective of the paper

The main objective of this paper is to use deep learning models integrated with explainable AI (XAI) techniques to improve the diagnosis and localization of diabetic foot ulcers (DFUs). By incorporating XAI methods such as Grad-CAM, SHAP, and LIME, this framework brings transparency to the decision-making process of the model, allowing clinicians to understand and trust the predictions.

In addition, the use of data augmentation significantly improves model performance, which overcomes the challenges of limited and imbalanced datasets. Techniques such as rotation, scaling, and flipping reduce overfitting by increasing the diversity of training samples and improve the generalization ability of the model.

The combination of these state-of-the-art deep learning architectures, explainability tools, and data augmentation creates a robust, accurate, and interpretable system for DFU detection that helps clinicians make informed and reliable decisions.

Input:

- DFU dataset with size 224x224

Variables:

- η = Learning rate
- δ = Epochs
- γ = Batch size
- σ = Number of images covered in one batch size
- ω = Pre-trained CNN method weight

Output:

- Predictions with explainability (Grad-CAM, SHAP, LIME)

Steps:**1. Preprocessing:**

- Resize and normalize images.
- Augment dataset with rotation, scaling, and contrast adjustments.

2. Model Training:

- Initialize pre-trained CNN (e.g., ResNet50, SNN).
- Fine-tune layers with:
 - Binary cross-entropy loss.
 - Adam optimizer with $\eta = 0.0001$.
 - Batch size $\gamma = 32$ and epochs $\delta = 50$.
- Use SMOTE to address class imbalance.

3. Explainability Integration:

- Compute Grad-CAM heatmaps for ulcer localization.
- Generate SHAP values for feature importance.
- Apply LIME for superpixel-based analysis.

4. Post-Processing:

- Validate predictions using metrics (accuracy, precision, recall, F1-score, AUC).
- Visualize heatmaps for clinician interpretation.

Algorithm. Planned DFX_XAI Outline for DFU Finding

Loss function: For the SNN, the contrastive loss function can be defined as.

$$L = (1 - y) \cdot \frac{1}{2} \cdot D^2 + y \cdot \frac{1}{2} \cdot \max(0, m - D)^2$$

Where: y is the label (1 for dissimilar, 0 for similar), D is the Euclidean distance between embeddings, m is the margin for dissimilar pairs.

The description of used dataset in this methodology as follows:

The diabetic foot ulcer dataset

To evaluate the model, data was gathered from 1050 skin patches that were donated by patients. Only 540 of the patches were determined to be normal, suggesting healthy skin; 510 were determined to be aberrant and classified as ulcers. On the other hand, a large amount of labelled data is necessary for the proper operation of a DL method. It could take a lot of time and money to gather a lot of medical data. Therefore, DL models may be improved and overfitting reduced by using image labelling, data enhancement, transfer learning, and regularisation. A solution to this problem is the use of patch labelling to enhance the size of the dataset. It sorts data by selecting the most important parts of a big sample and assigning them to the appropriate category. The first research used a 224-pixel-high by 224-pixel-wide sliding window to recover the Region of Interest (ROI) from each sample. The window was moved from top left to bottom right. According to the areas of normal and ulcer skin, patches are categorised as either healthy or ulcer. In contrast, the DFU_XAI architecture enhanced the efficiency of DL models by concentrating on individual DL traits rather than the whole sample. The parts associated with ulcers were marked in larger samples. Important DL information is extracted for classification using this method. When training a DL model, using appropriate DL data helps to prevent overfitting by reducing memory allocations and compute demands. This method yielded 1680 image patches, 830 of which were normal and 850 of which were ulcer related. Picture examples that have been cropped are shown in Fig. 2.

By employing new colour spaces to enhance contrast, resizing, relocating, and randomly scaling data, expand both its size and diversity. In this study, 1680 skin patches were rotated at angles of 90° and 180°. Following this process, the DFU data set was created by the researchers. It contains 3100 skin patches, out of which 1710 photographs depict aberrant ulcers, and 1490 pictures show healthy, usual skin. The DFU dataset is divided into two sets, one for training and one for testing, via the “train_test_split” algorithm. 10% of the information from the training set is re-split using the same procedure for validation. The dataset for this investigation is summarised in Table 2. Using SMOTE, this approach tackles imbalanced DFU datasets. Overfitting caused by



Fig. 2. Ulcer Skin normal and abnormal Images [Available at: [google.com/search/images](https://www.google.com/search/images)].

Dataset	Label	Original image	After cropping	After augmentation
DFU	Abnormal (ulcer)	510	850	1710
	Normal (Healthy)	540	830	1490
	Total Images	1050	1680	3100

Table 2. The dataset.

random oversampling is, however, addressed by this strategy. Due to imbalances in the datasets, DL models might be biased. SMOTE distributes data evenly to reduce bias. Model performance is impacted by efforts to minimise bias and collect minority label characteristics. The results of models may not always be improved by SMOTE. It can generate synthetic samples that aren't always realistic. The distribution of minority labels may not be correctly reflected by these hypothetical samples. When samples from the majority and the minority overlap, this can cause information from the dominant label to seep out.

1. Dataset Composition and Size

- **Source:** Data was gathered from **1050 skin patches** donated by patients.
 - **Healthy skin (Normal):** 540 patches.
 - **Ulcerated skin (Abnormal):** 510 patches.
- **Enhanced Dataset:**
 - After preprocessing and data augmentation, the dataset expanded to **3100 patches**:
 - **Healthy skin:** 1490 images.
 - **Ulcerated skin:** 1710 images.

2. Preprocessing Techniques

- **Patch Labeling:**
 - Images were divided into 224×224-pixel patches.
 - Patches were labeled as healthy or ulcerated based on visual features.
- **Sliding Window Approach:**
 - Regions of interest (ROI) were extracted from samples by sliding a window across images.
 - This helped segment and focus on the most relevant areas for classification.

3. Data Augmentation

- Augmentation was applied to increase the dataset size and diversity while reducing overfitting:
 - **Techniques Used:**
 - Rotation (90° and 180°).
 - Contrast enhancement.
 - Resizing.
 - Relocating and random scaling.
 - These methods generated additional training samples, improving model generalization.

4. Addressing Dataset Imbalance

- **SMOTE (Synthetic Minority Oversampling Technique):**
 - Synthetic samples were generated to balance the number of healthy and ulcerated samples.
 - This reduced bias and improved the model's ability to learn minority class features.
- **Challenges with SMOTE:**
 - While it helped address imbalances, synthetic samples may not always reflect real-world data distributions.

5. Train-Test Split

- The dataset was split into training, validation, and testing sets:
 - **Training Set:** Majority of the data used to train the models.

- **Validation Set:** 10% of the training data used for hyperparameter tuning.
- **Test Set:** Held out for evaluating final model performance.

6. Significance of Preprocessing and Augmentation

- The applied preprocessing and augmentation methods enhanced model training by:
 - Mitigating overfitting through diversified data.
 - Improving the model's robustness to variations in real-world data.

Reducing computational demands by focusing on key image features.

Strong hardware and time are required for a CNN model with many training parameters. These concerns may be resolved by transferring variables and weights from existing dataset techniques to the new model. Levels may be bridged and more levels added using Transfer Learning and transferred parts. This method is less computationally demanding and faster even with little datasets. Data is expensive, time-consuming, and difficult to get by, yet traditional AI must train its model. This issue is addressed, DFU_XAI used ImageNet weights derived via transfer learning. First, the DFU_XAI framework's training time is reduced using this method. Second, it improves the success rate of the framework. Third, it allows for the insertion of more data and the adjustment of parameters during model training, which improves the framework.

This research uses six state-of-the-art CNN methods for classification: ResNet50, DenseNet121, InceptionV3, MobileNetV2, Xception, and SNN. Architecture, classification accuracy, and explainable prediction were the deciding factors in the selection of these cutting-edge models. Structures and modules differ from model to model. The following networks provide outputs from their siamese layers: Xception, ResNet, DenseNet, InceptionV3, MobileNetV2, and InceptionV3 inverted. Xception makes use of depthwise separable convolution blocks. When it comes to ImageNet, these networks shine. The input image's DL data was obtained using pre-trained weights to improve the model's performance. Big datasets are no match for these models. These CNN models used transfer learning with pre-set weights. By exchanging data and weights, models may train more quickly and efficiently. Performance is improved by combining pre-trained models with fine-tuned layers. The data and split ratio were used to train and evaluate six pre-trained models. These models have six fine-tuning layers to enhance performance. Layers that altered the model's pre-trained weights enhanced DFU_XAI. Refinement layers categorise DL characteristics that have been retrieved. The optimised layers enhanced the model's functionality.

Residual network

Robust architecture for DL Residual Network (ResNet), short for Residual Network, which has shown promise in many picture classification tasks, one of which is the detection of DFUs. ResNet is a DL system that uses residual learning to address the vanishing gradient problem and capture intricate patterns in medical images. ResNet learns complex features from high-resolution images, allowing it to reliably distinguish ulcerous from non-ulcerous regions in DFU identification. The use of ResNet in conjunction with explainability techniques such as Grad-CAM and saliency maps allows for the visualisation of the method's procedure for making decisions. With the help of this partnership, healthcare professionals are better able to accept and evaluate the model's indications, which in turn improves their ability to make impartial decisions²⁰.

Convolutional neural networks

Tasks requiring object recognition, such as image segmentation, detection, and classification,

are well-suited to DL techniques like CNNs, often called "ConvNet". One kind of DL that can find patterns in program and image data is CNNs. Spiral algorithms exploit a data leak to trick the recommendation system into giving an inaccurate result by pushing the recommendation system over the data.

A feature map depicting the input data's structure as a function of features is the end product. Using the principle that nearby pixels have stronger correlations than distant ones, CNN implements²⁴. In this case, Fig. 3 shows the pooling process that the Max Pooling layer employs, which involves sliding a two-dimensional filter over each feature map channel. The feature maps are reduced in size by the pooling layer, which condenses all features within the coverage region of the filter. As a result, the network does not have to learn as many parameters, and its burden is mitigated. There are several types of pooling layers, including max, average, and global.

DL Method is used for the detection of DFUs; CNNs and similar models excel at processing images. CNNs can identify the intricate patterns and textures seen in DFUs because, via convolutional processes, they dynamically learn the spatial arrangement of features from input photos. By stacking several convolutional layers, CNNs accurately identify DFUs by capturing edges, textures, shapes, and lesions. To find the parts of the foot picture that have the most impact on the model's predictions, CNNs combined with explainability techniques like Grad-CAM technology and saliency maps like DFU_XAI may be employed. By exposing the model's decision-making process, increases diagnostic accuracy, empowers clinicians to trust AI-driven judgements, and ultimately leads to better patient outcomes. A strategy called XAI is used to incorporate the most significant elements of the trained method. This article describes in detail the three most used XAI methods for visual analysis: Grad-CAM, SHAP, and LIME.

Locally interpretable explanations and model-independent

The goal of the LIME method is to provide detailed justification for each calculation made by a black box method. A key idea behind LIME is to simplify and make more transparent a "glass box" model to approximatively

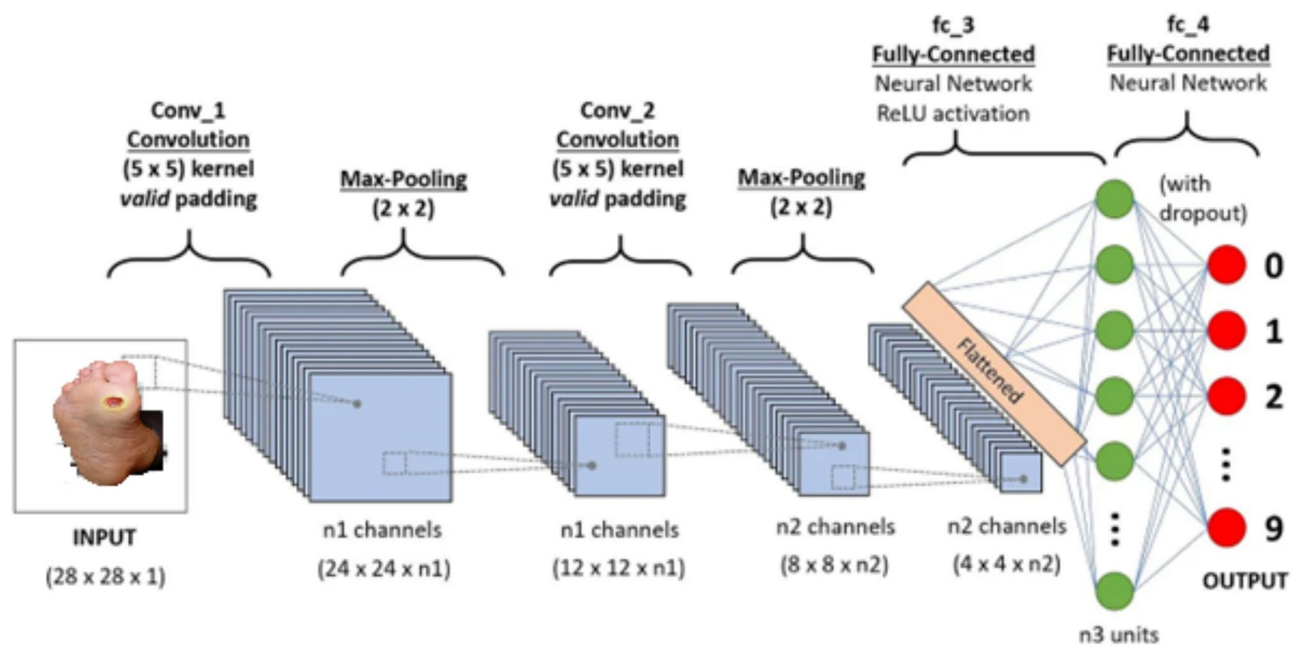


Fig. 3. Framework for number recognition with CNN's.

simulate the local behaviour of a “black box” model. This will make interpretation much easier. By activating or deactivating super pixels selectively, LIME distorts visuals. This technique determines the significance of continuous super pixels in the output class's original image and then disseminates that information. To make ML systems more trustworthy, LIME shows how the input features of CNN models affect predictions, which increases model interpretability and transparency. To use LIME, the first step is to divide an image into super pixels²⁵. The count of super pixels determines the region's segmentation. Hyper pixels are adjacent pixels that share the same hue and position. This method produces more thorough and precise segmentation, which aids in locating zones that may forecast the output class. LIME approximates the black box model with an easier interpretable model to explain model predictions locally. Super pixel activation and deactivation perturb images, revealing model decision-making areas.

Weighted gradient activation mapping for classes

In typical CNN models several blocks are combined for feature extraction and categorization. A fully connected layer in the Classification block calculates probability scores from the SoftMax layer using the recorded features. In the end, the classification results of the model are generated with the likelihood score, which has the most value. By selecting the category that most closely matches the input image, the accuracy and performance of the model can be improved. Without modifying the network architecture or training, Grad-CAM provides visual explanations. This technique finds important areas of the image and uses the gradient of the feature map of the final convolutional layer to highlight the components that have the greatest impact on predicting the outcome. Grad-CAM pinpoints important areas of a picture to make the model more transparent and easier to comprehend, which in turn makes the model's forecast clearer to the user. Huge gradients impact the accuracy of Grad-CAM picture predictions. Visualisation methods such as Grad-CAM and GradCAM++ assess a convolutional layer of a CNN by extracting quantitative and qualitative information about the layer's inner workings via the identification of distinct image characteristics. You can fix Grad-CAM's low heatmap resolution with Grad-CAM++. Based on the results, Grad-CAM and GradCAM++ have the potential to improve the network's ulcer analytics capabilities by accurately locating the ulcer images. Grad-CAM produces pictorial clarifications by drawing attention to crucial regions of a picture that impact the model's forecast. By providing higher heatmap resolution, Grad-CAM++ enhances Grad-CAM and makes it easier to localise ulcers in photos¹⁸.

Shapley additive explanations

Specifically, the SHAP method offers local explainability for text, picture, and tabular datasets. Subtractive feature importance is determined by averaging the feature space's marginal contributions. SHAP values illustrate the impact of every input attribute on the model's anticipated output. To improve the interpretability and transparency of decision-making in DL models, the DFU_XAI outline employs XAI methods such as GradCAM, SHAP, and LIME. The superpixel is retrieved from the expected samples using LIME in this method. The DL model's transparency is enhanced by the acquired pixels. SHAP may provide both positive and negative numbers to describe the judgements made by DL models in different samples. These features make DL models more interpretable and transparent in their decision-making process¹⁸. To create the heatmap, Grad-CAM employs the last convolution layer of the DL model. The methods judgement is then made clear by highlighting these heat maps. The decision-making process may be fully understood by looking at the SHAP values, which show how

each feature contributed to the model's output. This method explains how each input characteristic affects the predictions by producing positive and negative values.

Siamese neural network

Two comparable networks that examine syntactic networks between different classes make up the SNN, a sign verification technique. They work side by side to create header graphics and divide up the weights. Knowledge regarding parallel processing, useful for classification or similarity measure, may be acquired using this CNN architecture. Improved heading representations are a result of the Siamese effect on animate nerve organ networks' evaluation of heading correspondences. To make positive pairings more similar and negative pairings less similar, that are employed contrastive loss, penalises dissimilar samples if they are too near to the given margin. Medical image analysis makes use of SNNs to compare and develop functions of similarity between input items. Achieving symmetric learning is made possible by these networks' identical subnetworks, which share parameters and weights³⁵. They find the distance metric between embeddings and use it to create a fixed-size vector that represents the higher-dimensional representation of the input picture. Improved classification efficiency, extensive feature learning, and ulcer development/healing tracking are all possible using SNNs. They may be added to other deep-learning models to boost accuracy, used to evaluate the network's performance in classification and matching, and tested on a separate dataset. The structure of an SNN is shown in Fig. 4.

The categorisation and analysis of DFU is facilitated by the use of SNN's, which exhibit strong performance in similarity identification and feature learning. By incorporating them into DFU processes, diagnostic accuracy and patient outcomes are enhanced.

Comparative Performance comparative performance of the six deep learning models.

The DFU_XAI framework integrates six advanced deep learning models—Xception, DenseNet121, ResNet50, InceptionV3, MobileNetV2, and Siamese Neural Network (SNN)—to improve the diagnostic accuracy and explainability of diabetic foot ulcers (DFUs). Each model offers a distinct strength with its unique architectural features.

Xception uses depthwise separable convolutions, which significantly reduces parameters and increases computational efficiency. It is suitable for large datasets but is weak in capturing complex DFU feature interactions.

DenseNet121 uses dense connectivity, in which every layer is connected to all previous layers. It improves gradient flow and helps to robustly extract intricate DFU features.

ResNet50 introduces residual connections that solve the vanishing gradient problem. It helps to train deeper networks and can accurately distinguish high-resolution medical images.

InceptionV3 uses multiple filter sizes that capture multi-scale features. It is effective in detecting DFU characteristics at different resolutions but requires more computational resources.

MobileNetV2 is an efficiency-focused model that uses inverted residuals and linear bottlenecks. It achieves moderate diagnostic accuracy and performs best in resource-constrained environments such as mobile diagnostics.

The Siamese Neural Network (SNN) uses two identical subnetworks and a contrastive loss function that optimize the capability to compare image pairs. It achieves superior performance metrics in ulcer localization and classification.

By using pre-trained weights and fine-tuning layers, the DFU_XAI framework ensures a balanced approach that gives equal importance to accuracy, computational efficiency, and clinical interpretability. This integrated architecture provides a robust foundation for DFU classification and facilitates explainable AI-based clinical decision-making.

SNN transforms each input image (I_1, I_2) into high-dimensional feature embeddings ($f(I_1), f(I_2)$) using two identical subnetworks with shared weights. Mathematically:

$$f(I_1) = g(I_1, W), f(I_2) = g(I_2, W).$$

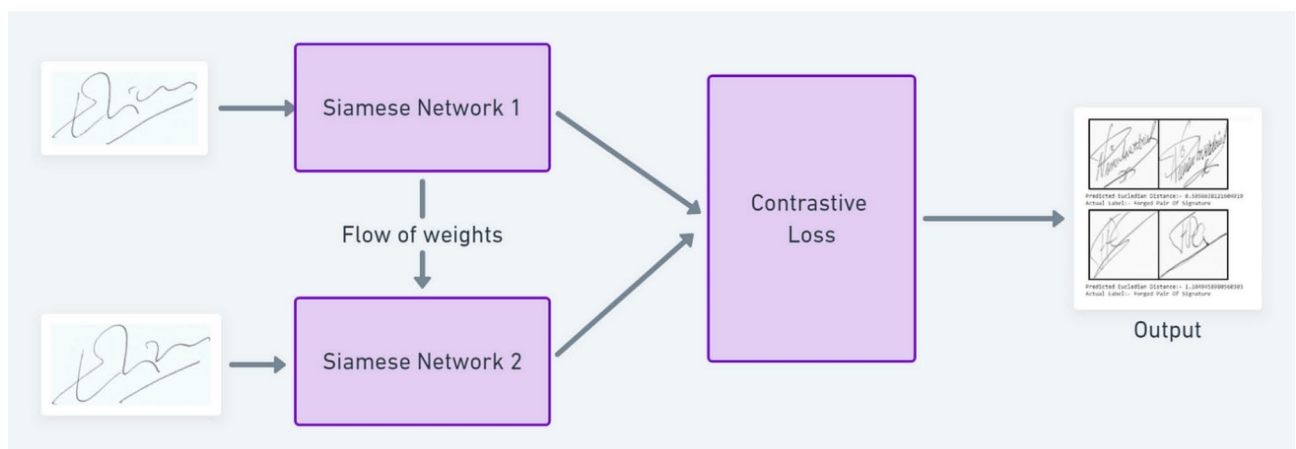


Fig. 4. SNN Architecture.

where $g(I_1, W)$ is the feature extraction function (e.g., a CNN), and W represents the shared weights of the network.

Distance Metric: A distance metric is used to quantify the similarity or dissimilarity between two embeddings, which is usually the Euclidean distance (D).

$$D(f(I_1), f(I_2)) = \sqrt{\sum_{i=1}^n (f_i(I_1) - f_i(I_2))^2}$$

Contrastive loss function: Siamese Neural Network (SNN) uses a contrastive loss function so that embeddings of similar pairs stay close to each other and dissimilar pairs move away. The loss function is defined in a way that helps to train the model accurately.

$$L = (1 - y) \cdot \frac{1}{2} \cdot D^2 + y \cdot \frac{1}{2} \cdot \max(0, m - D)^2$$

Where:

- y is the label (1 for dissimilar, 0 for similar),
- D is the Euclidean distance between embeddings,
- m is the margin for dissimilar pairs.

Optimization objective: The goal of training is to minimize LLL, which.

- Reduces D for similar pairs ($Y = 0$).
- Ensures D for dissimilar pairs ($Y = 1$) exceeds the margin m .

Symmetry and parameter sharing: The shared weights of twin subnetworks ensure symmetry.

$$g(I_1, W) = g(I_2, W)$$

This simplifies the training, reduces the number of parameters, and ensures consistent feature learning, which maintains uniformity among the input images.

Results and discussion

In this section the hyperparameters, working conditions, objective and subjective factors, and assessment outcomes of the DFU_XAI outline are explained in detail. Apart from this, the framework that has been suggested has been compared with existing methods. First, an efficient Learning Rate (LR); doosra, optimal number of epochs; third, an effective optimizer; chautha, moderate group size; And fifth, the ideal train/test divided ratio was considered when parameters were chosen for the learning, training, testing, and validation phases in the experimental environment of the DFU_XAI framework.

Experimental setup

Utilizing Xception, InceptionV3, DenseNet121, MobileNetV2, ResNet50, and SNN to train the DL model with updated layers. Lots of resources are required for this simulation. DFU_XAI was constructed using these components. The models were updated to use pre-trained ImageNet weights instead of random ones. To distinguish between samples with ulcers and those without, tweaked the last layers of the applicable models. The performance of the new layer model was enhanced using the softmax activation method. The LR, batch size, and number of epochs are the three parameters that are first tuned.

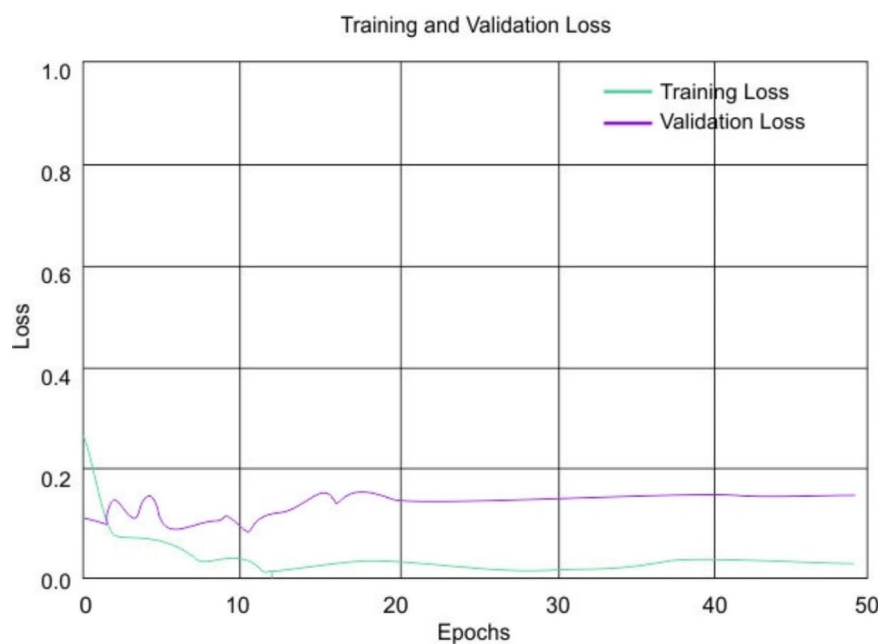
This model was trained using an Adam optimiser with a LR of 0.0001, a binary cross-entropy loss that lasted 50 epochs, and 32 batch sizes. The model's speed can drop if someone utilise LR. So, choose the optimal number of epochs, and the Y-axis shows the accuracy achieved using LR, which trains the model. Adam stands head and shoulders above the competition when it comes to optimisers because of its remarkable adaptive LR capabilities. To reduce memory use and training time, choose the ideal batch size. A smaller batch size shortens the training time, whereas a larger batch size requires more memory. Overfitting occurs when there are too many epochs and underfitting occurs when there are too few. Hence, to tackle both issues, they chose a small number of epochs.

To choose ResNet50 as the best model according to performance indicators, the aforementioned criteria are considered throughout the selection process. This is it, the best ResNet50 model ever. After the first epoch, the model's training loss was 0.2866 and its validation loss was 0.1108. This training loss from the ResNet50 method is shown on the Y-axis, while the X-axis shows the total number of epochs. Performance criteria were used to pick the best model.

Based on performance indicators after training, the best model was chosen. To create the DFU_XAI framework, CNN models that had been previously trained were fine-tuned. The SNN, MobileNetV2, ResNet50, DenseNet121, InceptionV3, and Xception are some of the improved models that make use of ImageNet weights and data to save training time. A SNN model was the most effective. The healthcare industry places a premium on clarity and openness, making it particularly challenging to integrate and scale the DFU_XAI architecture. For several classification tasks, this article discusses the Receiver Operating Characteristic (ROC) curve and Confusion Matrix (CM) of the final pre-trained ResNet50 model. Using recall and specificity as evaluation criteria, the ROC curve was constructed. The X-axis represents fall-out (1-specificity), while the Y-axis represents the recall score. The highest AUC score was 0.98535, achieved by the ResNet50 model (blue highlighted). The DFU test set was used to generate the final CM, where the X-axis represents the anticipated class labels and



(a)



(b)

Fig. 5. Results of the training simulation using the DFU dataset for the proposed DFU_XAI framework (a) model trained without overfitting, and (b) trajectory of the loss function.

the Y-axis represents the actual class labels. The credibility of medical trial investigators is assessed by this CM. While training an AI model, this method can detect and correct any biases that may exist.

The results of the DFU dataset training simulations using the proposed DFU_XAI framework are shown in Fig. 5. We have fine-tuned each of the hyperparameters used to train CNN's framework. Two hyperparameters that are vital for training the model are the optimiser and the gradient descent loss function. Adam was selected as the optimiser function because to its ability to manage sparse gradients on large datasets and its integration of

the best features of AdaGrad and RMSProp. Since the DFU dataset is used for binary classification in this work, the binary cross-entropy loss function is a good choice. It is possible to generalise the model using a batch size of 32. The suggested framework was trained using data from the 50th epoch. Over 98% validation accuracy and 97% training accuracy were reached after just 29 training epochs by the model. Avoid over-optimism by using cross-validation with oversampling. This technique provides an accurate evaluation of the model's efficiency because the model was trained without ever seeing or oversampling test data. Training the model without overfitting is seen in Fig. 5 (a). The loss value is dropping at a quick pace, as shown in Fig. 5 (b), which is the loss function curve.

Evaluation metrics

Classification problems often use confusion metrics as an evaluation tool for DL models. The relationship between the predicted and real labels is shown by it. Two yes/no labels for a maximum of four possible outcomes are allowed in a binary classification model. These are a few: If a model can accurately identify an ulcer in a picture marked "ulcer," we say that the model is True Positive (Tp). A False Positive (Fp) occurs when a picture that does not include an ulcer label is mistakenly thought to be an ulcer. When a picture with an ulcer label is mistakenly thought to not have an ulcer, this is called a False Negative (Fn). Pictures that are accurately identified as ulcer-free but do not include ulcer labelling are called True Negative (Tn). Equation 1 measures the overall effectiveness of the model in correctly identifying both positive (ulcer) and negative (healthy) cases. Equation 2 is used to correctly identify positive cases (ulcers) out of all cases that were predicted to be positive. Equation 3 measures the model's ability to correctly identify all positive cases (ulcers). It is particularly important in medical diagnoses, as it ensures that all actual positive cases are identified, minimizing the chances of missing any ulcers (false negatives). Equation 4 is the harmonic mean of precision and recall, providing a balanced measure of the model's accuracy when there is an uneven class distribution (e.g., more healthy cases than ulcer cases). These definitions apply to many assessment measures:

$$Accuracy = \frac{Tp + Tn}{(Tp + Tn + Fp + Fn)} \quad (1)$$

$$Precision = \frac{Tp}{(Tp + Fp)} \quad (2)$$

$$Recall = \frac{Tp}{(Tp + Fn)} \quad (3)$$

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

Explainability: Describe the Grad-CAM computation mathematically:

$$L_c^k = \frac{\partial Y^c}{\partial A^k} \quad (5)$$

Heatmap = $\text{ReLU} \sum_k w_k A^k$, w_k = Global Average Pooling (L_k^c)

The DFU_XAI framework's efficacy is evaluated using certain criteria, which are used for DCFU classification. The DFU_XAI framework can provide reliable forecasts because of its high accuracy. Minimising Fp allows for precise prediction of foot ulcers. Recall should represent the majority of foot ulcers to avoid missing any serious cases, hence minimising Fn is important. The DFU_XAI framework uses the F1-score to detect ulcers while keeping Fp and Fn intact. The AUC indicates that the model correctly gives ulcer cases a higher probability than healthy ones.

Analysing the result

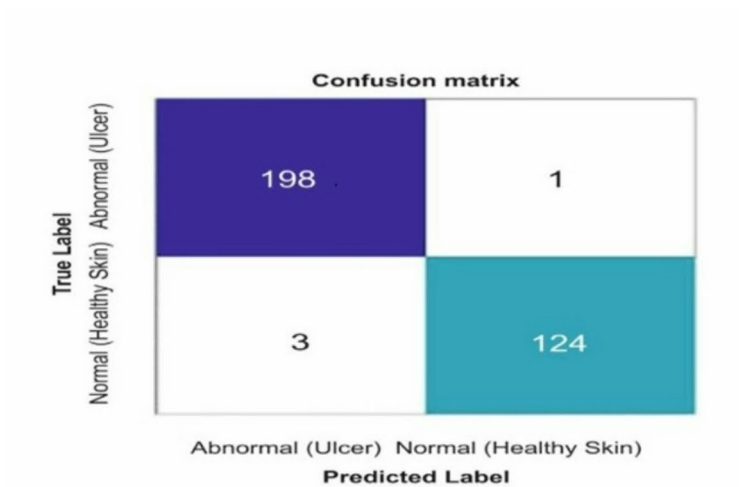
The receiver operating characteristic curve and matrix of confusion for the DFU dataset are shown in Fig. 6 of the DFU_XAI. The ResNet50 transfer learning model forms the basis of the framework. The use of explainable methodologies allows for the detection of normal or ulcerated foot tissue using CNN. Alternately, Fig. 6 (a) displays normal and ulcer images from 124 to 198 DFU samples, respectively. The suggested approach erred in its classification of one ulcer sample. The DFU_XAI framework's explainability is its strongest suit as it ensures the model's dependability. The consistent and stable structure of this framework is shown by its surface value of 0.986, as shown in Fig. 6 (b).

All of the transfer learning models are run separately on the DFU dataset to assess the DFU_XAI framework.

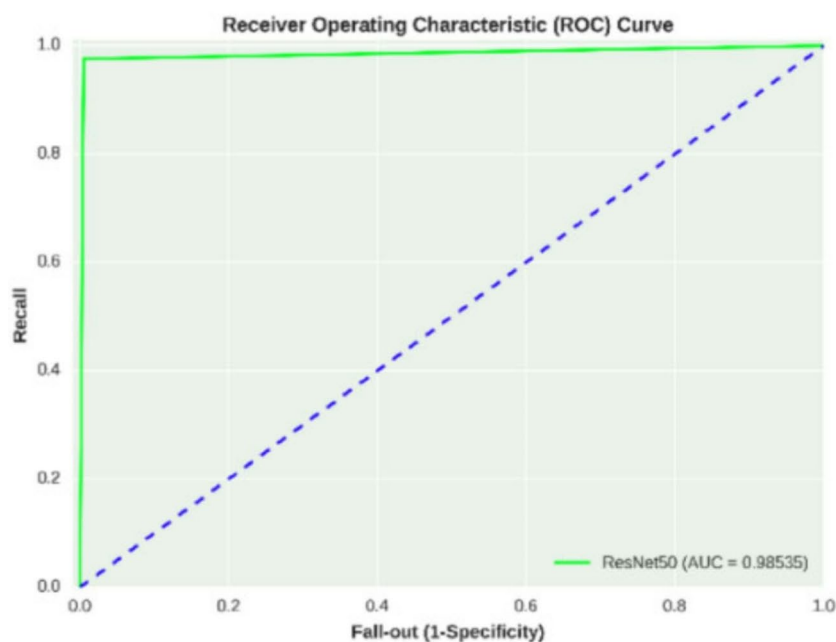
The Fig. 7 clearly shows the transition from raw images to heat maps and overlays, making it easy to understand the localization process. The overlaid heat maps provide actionable insights for clinicians, highlighting ulcerous regions for further examination or treatment.

Figure 8 displays the results of evaluating all of the pre-trained CNN models used in this experiment. With values of 99.3% for accuracy, 98.5% for precision, 98.7% for recall, and 98.6% for AUC, SNN is the most effective model. Based on SNN, the DFU_XAI framework achieved the best results.

Skilled medical professionals can confirm these XAI-localized zones. These specialists may check the highlighted regions for correctness and clinical relevance to validate model predictions and align with clinical interpretations. Patient data may demonstrate the therapeutic efficacy of XAI. Decision-making using ResNet50 models in clinical contexts may be visualised and analysed using Grad-CAM, LIME, and SHAP. ResNet50 can



(a)



(b)

Fig. 6. (a) Predicted table, and (b) receiver operating characteristic.

identify ulcers in medical photos. Reasons for the model's predictions are laid forth here. It might look into the DFU's sloping design and undiscovered constraints.

In terms of DFU analysis, SNNs is the best method. When compared to traditional neural networks, SNNs are superior at prediction because they mimic the way the brain operates. The interpretability and reliability of DFU predictions might be enhanced using SNN and XAI. The stimulation mappings from the final convolutional layer are used by Grad-CAM to produce heat maps that emphasise picture regions with class labels. The heat maps generated by the last convolutional layer of ResNet50. Figure 9 displays the outcomes of three gradient-

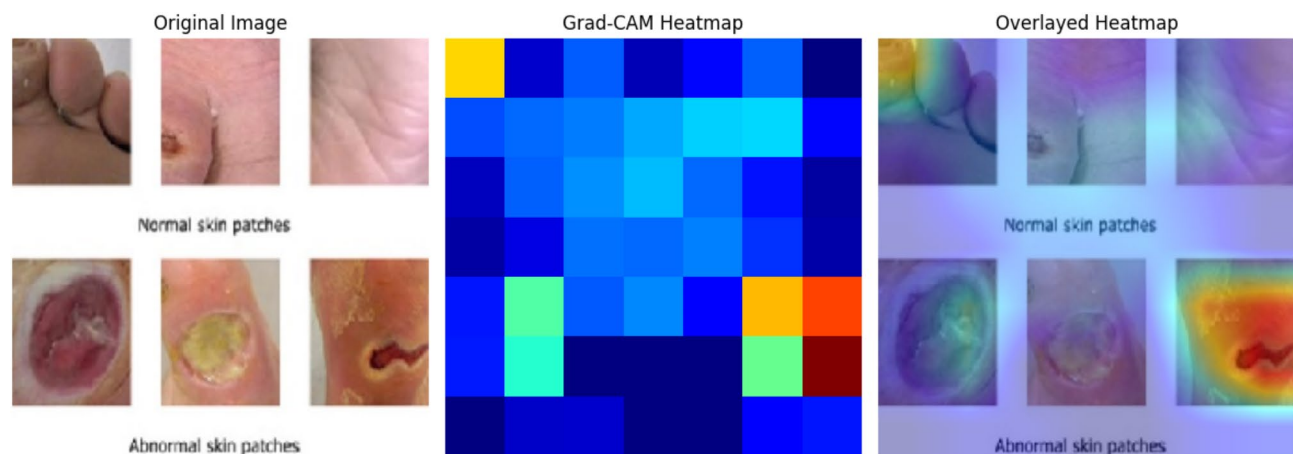


Fig. 7. Heat maps for the diabetic foot ulcers.

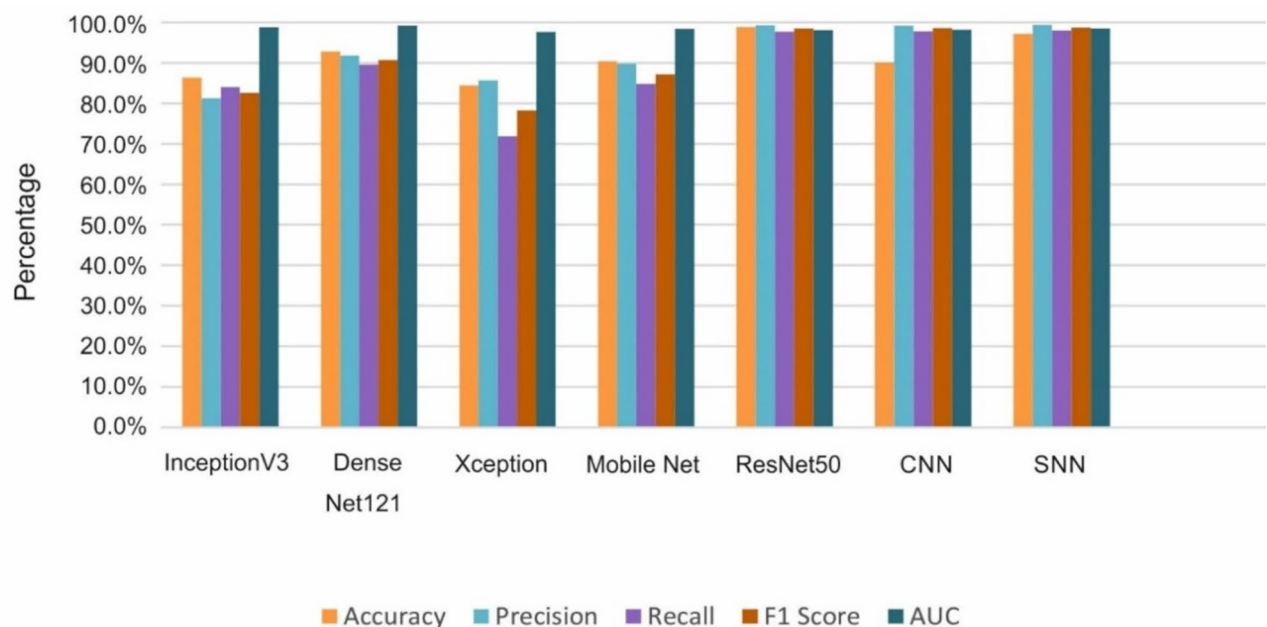


Fig. 8. DFU_XUI framework compared with Ultra Modern Network.

based approaches, which were applied to a large number of test pictures by comparing heat map data to DFU specialist ulcer markers. That way, the doctor may use the picture to predict where the ulcer is.

In clinical contexts, XAI algorithms like Grad-CAM, LIME, SHAP, and SNN may be used to display and study The ResNet50 model's decision-making mechanism. Medical practitioners may rely on the heatmaps generated by these XAI algorithms for diagnosis, treatment, and prognosis. These technologies augment the clarity of ResNet50's prognostic reasoning, enabling physicians to authenticate it. Maintaining transparency and integrity ensures that the model's forecasts and therapeutic suggestions are precise and applicable to the context of the clinical environment. Enhancing the interpretability and reliability of predictions in the analysis of DFUs may be achieved by combining SNN with XAI methods.

In terms of performance, SNN outperforms ResNet. An assessment of the performance of the DFU_XAI outline was conducted utilizing the ROC curve and the AUC. Figure 9 displays the ROC curve of six widely used CNN models, namely DenseNet121, ResNet50, InceptionV3, Xception, MobileNetV2, and SNN. These models use identical learning parameters and sample splits. When comparing CNNs with SNNs, the latter showed superior performance in categorising patients with clean DFUs and ulcers.

Figure 10 describes the CNN and ROC curves comparison. To diagnose DFUs, DFU_XAI integrates XAI algorithms with the traditional AI model, in contrast to earlier models that solely use training data or standard artificial intelligence. Its unique attributes are the origin of its exceptional performance:

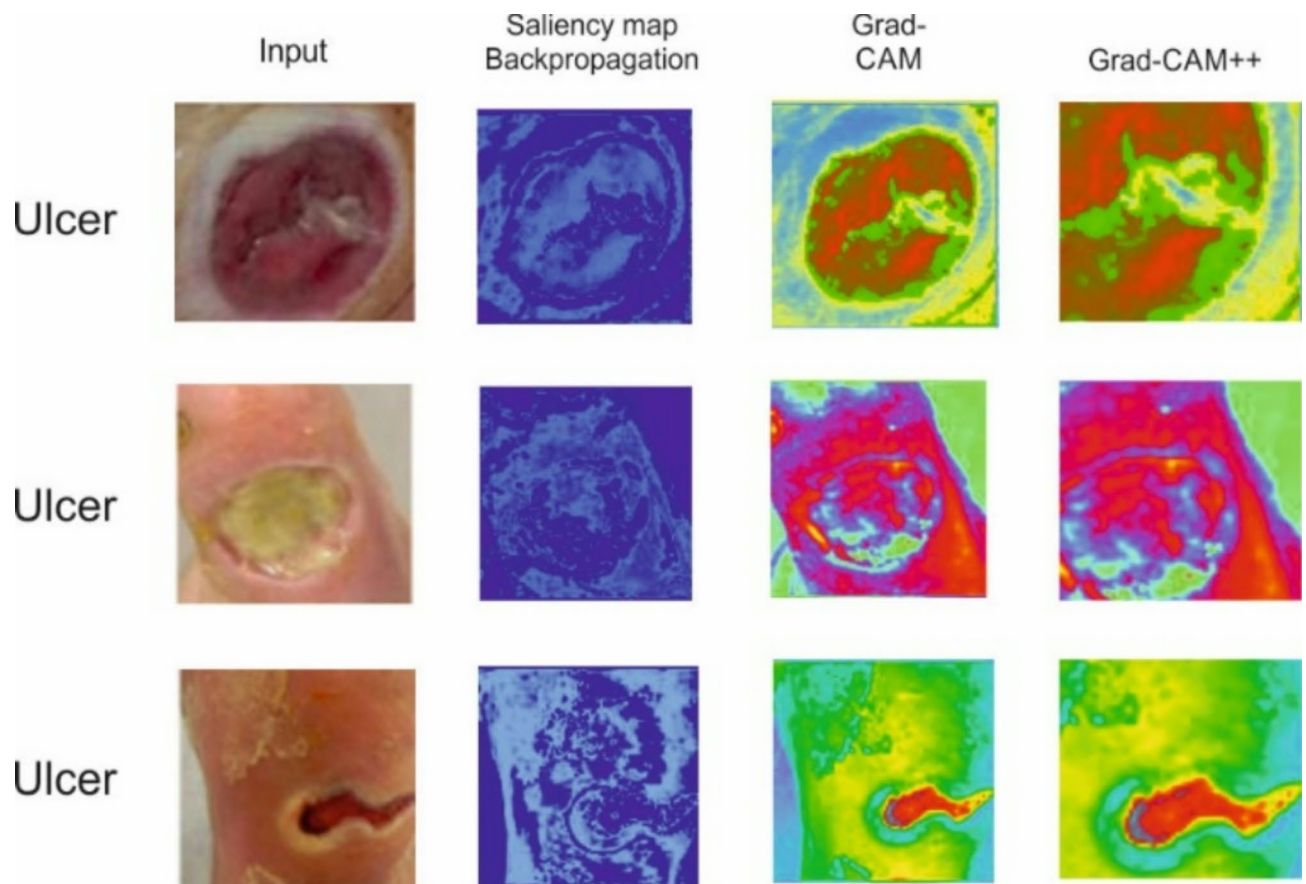


Fig. 9. Three gradient-based approach testing results.

To address the problem of “vanishing gradient,” the DFU_XAI framework utilises a fine-tuned ResNet50 model.

- To enhance explainability, XAI algorithms were included in the pre-trained CNN model.
- An automated heatmap tool using the DFU dataset and AI may be advantageous for endocrinologists and doctors in elucidating ulcers.

Further, the obtained results are compared with the existing literature as shown in Table 3. The main features of DFUs for the years 2023 and 2024 are outlined in this table, with an emphasis on historical trends and developments.

Discussions

DFU analysis is the sole application of this specific architecture. This method is suited primarily for DFU assessment and certain medical scenarios. Following this, the SNN may emerge as the best model compared to others like ResNet50 within the same framework. However, this framework is not sufficiently adaptable to diverse datasets and medical scenarios; it needs to be more flexible. Implementing this system in real-world clinical settings has proven to be challenging. This approach may fail to explain the progression of ulcers, patient health, and diabetes. Clinically acceptable and skill-relevant XAI-based model explanations are very important in real-world clinical settings. Predictions on the reliability and safety of the framework must take patient data privacy and confidentiality into account. Clinical endocrinologists and other medical professionals may benefit from the DFU_XAI's precision in diagnosis and treatment. If endocrinologists use XAI, they may be able to understand how most AI models make judgements. To aid doctors in the prediction of ulcers, this paradigm places a focus on risk factors. Endocrinologists may find comfort in its openness and high level of accuracy (98.76%).

While AI models do increase the practical applicability, they do not improve the diagnosis or treatment of DFUs. There are still ethical and legal hurdles to overcome when bringing medical therapies to clinics. These issues may be addressed by the introduction of DFU_XAI into medical procedures. AI based architectures have become more popular due to the success of AI decision-based systems. The XAI paradigm facilitates the confidence, understanding, and use of these techniques by younger practitioners and doctors.

Due to its explainability, the DFU_XAI framework inspires confidence in its forecasts among clinicians. Medical professionals may benefit from the XAI-powered heatmap explanation while making choices about

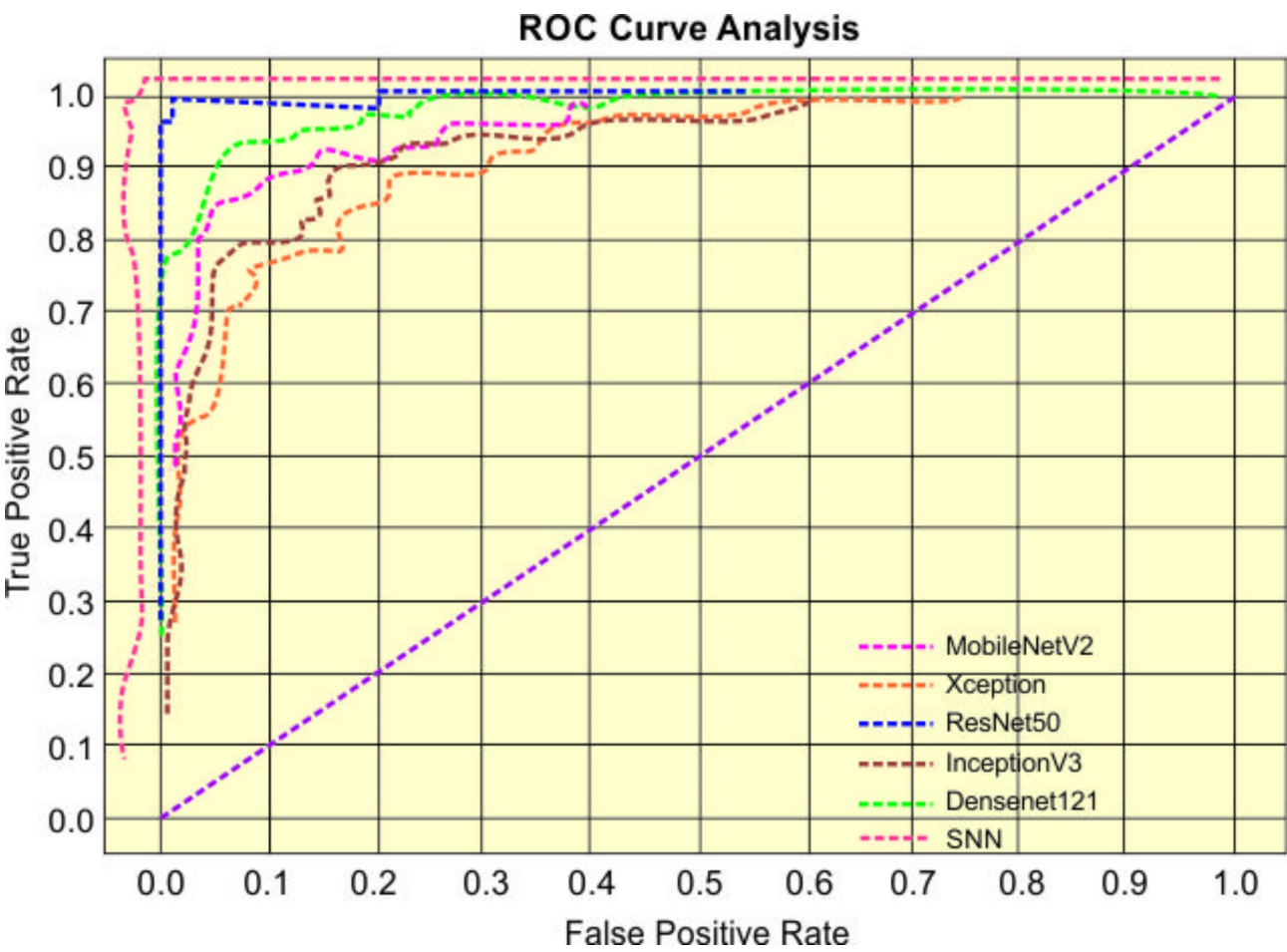


Fig. 10. CNN and ROC curves comparison.

Article	Feature	Outcome	Limitation
Kim et al. ²⁸	Cases	Approximately 18.6 million affected cases every year	Same is continued
Bethany et al. ²⁹	Risk Factor	Approximately 19% to 34% of diabetic cases	Same is maintained
Sara et al. ³⁰	Treatment	Treatment options for severe instances include outsourcing, debridement, wound dressings, medical interventions, and surgical procedures.	Options for therapy remain the same; ongoing emphasis on avoiding infections and unloading.
Chul et al. ³¹	Recommendations	Updated IWGDF recommendations for managing and preventing foot ulcers.	New policies and cooperative methods kept developing.
Dfu challenge ³²	Progress in Technology	Continuous use of machine learning and smart technologies to improve treatment and diagnosis precision.	Improvements in novel offloading strategies, diagnostic instruments, and AI-powered ulcer evaluation systems like Siamese neural network
N. Singh et al. ³³	Education Awareness	More focus on educating patients and doing regular foot exams	Continued focus on monthly checks and protective steps
Boulton et al. ³⁴	Effects of Foot Ulcers on Diabetes	Higher chance of hospitalisation and amputations, which costs healthcare services a lot of money.	Similar effects have been noticed, and attempts are still being made to lower risks through improved care.
Proposed framework	Improved DFU Detection using XAI and SNN	High accuracy in DFU detection (98.76%) with visual explanations via XAI methods (Grad-CAM, SHAP, LIME); enhances clinician trust in AI models	Limited adaptability to other skin conditions or datasets; challenges with real-world implementation

Table 3. Comparative analysis of proposed work with existing studies.

diagnosis, treatment, and prognosis. The model is now easier to understand, and physicians can confirm prognosis, thanks to this XAI-based design. Better foundation forecasts and situation recommendations are created and attainable by transparency. Employing this method, physicians may attempt to analyse lesion samples even in cases where an incorrect diagnosis is given. One may encounter clinically unpleasant responses from the DFU_XAI foundation.

Limitations

The DFU_XAI model, while effective for diabetic foot ulcer detection, may not perform well with other skin conditions or healing environments. Additionally, the framework poses risks to patient confidentiality due to data sharing and access during model training. Improving data protection and exploring Federated Learning (FL) could help address these privacy concerns and enhance the model's security.

Conclusion and future scope

This paper presents a system that utilizes DL and explainable AI methods to automatically identify foot ulcers. The system achieves a high level of accuracy especially using SNNs. Incorporating 1660 DFU image patches, the SNN model exhibited an extraordinary 98% accuracy, confirmed by outstanding F1 scores and AUC metrics. The study focused on the restricted use of skin patches and used techniques such as GradCAM, LIME, and SHAP to accurately distinguish between healthy and sick skin.

The model's exceptional precision and swift identification of ulcers have earned the confidence of clinicians, indicating prospective use in wider medical domains beyond DFUs, such as distinguishing between different skin disorders including chickenpox, shingles, and other skin abnormalities. The primary objective is to create a multiscale transfer learning framework that integrates many pre-trained models to achieve higher classification accuracy and efficiency in skin lesion diagnostics. This framework has the potential to decrease training time while simultaneously enhancing specificity and sensitivity.

Through the integration of many pre-trained networks, future research may extend the use of this system to identify a wider spectrum of dermatological problems. Unless suitable controls are implemented, this approach may encounter difficulties with data privacy. By using the distinctive features of the DFU_XAI architecture, the proposed technique aims to enhance diagnostic accuracy and comprehension in skin wound categorization, possibly revolutionising image-based diagnostics in diabetes-related comorbidities such as retinopathy and neuropathy.

Future enhancements for the DFU_XAI framework include the refinement of the model's architecture and the optimization of hyperparameters to increase the accuracy and efficiency of DFU identification. Utilizing sophisticated preprocessing and data augmentation methods may enhance the generalization capabilities of the model. The use of contemporary explainability techniques and user-friendly interfaces will provide healthcare practitioners with more profound understanding of how models impact decision-making. The use of innovative real-time techniques, especially for mobile apps, has the potential to improve healthcare procedures.

Data availability

The dataset is available at <https://www.kaggle.com/datasets/laithjj/diabetic-foot-ulcer-dfu>.

Received: 17 December 2024; Accepted: 17 February 2025

Published online: 25 February 2025

References

- Rufino, J. et al. Feature Selection for an Explainability Analysis in Detection of COVID-19 Active Cases from Facebook User-Based Online Surveys. *medRxiv* 2023-05. (2023).
- Dong, H. et al. Explainable automated coding of clinical notes using hierarchical label-wise attention networks and label embedding initialisation. *J. Biomed. Inform.* **116**, 103728 (2021).
- Gouverneur, P. J. Machine Learning Methods for Pain Investigation Using Physiological Signals. 6 (Logos Verlag Berlin GmbH, 2024).
- Lohaj, O. et al. Conceptually funded usability evaluation of an application for leveraging descriptive data analysis models for cardiovascular research. *Diagnostics* **14**(9), 917 (2024).
- Ahsan, M. et al. A deep learning approach for diabetic foot ulcer classification and recognition. *Information* **14**(1), 36. (2023).
- Thotad, P. N., Geeta, R., Bharamagoudar & Basavaraj, S. Anami. Diabetic foot ulcer detection using deep learning approaches. *Sens. Int.* **4**, 100210 (2023).
- Yap, M. et al. Deep Learning in diabetic foot ulcer detection: A comprehensive evaluation. *Comput. Biol. Med.* **135**, 104596 (2021).
- Goyal, M. et al. Robust methods for real-time diabetic foot ulcer detection and localization on mobile devices. *IEEE J. Biomedical Health Inf.* **23**(4), 1730–1741 (2018).
- Zhang, J. et al. A comprehensive review of methods based on deep learning for diabetes-related foot ulcers. *Front. Endocrinol.* **13**, 945020 (2022).
- Nagaraju, S. et al. Automated diabetic foot ulcer detection and classification using deep learning. *IEEE Access.* **11**, 127578–127588 (2023).
- Rai, M. et al. Early detection of foot ulceration in type II diabetic patient using registration method in infrared images and descriptive comparison with Deep Learning methods. *J. Supercomputing.* **78**(11), 13409–13426 (2022).
- da Oliveira, C. & Leandro, A. André Britto de Carvalho, and Daniel Oliveira Dantas. Faster R-CNN Approach for Diabetic Foot Ulcer Detection. *VISIGRAPP (4: VISAPP)*. (2021).
- Toofanee, M. S. et al. Dfu-Siam a novel diabetic foot ulcer classification with Deep Learning. *IEEE Access.* (2023).
- Tulloch, J., Zamani, R. & Akrami, M. Machine learning in the prevention, diagnosis and management of diabetic foot ulcers: A systematic review. *IEEE Access.* **8**, 198977–199000 (2020).
- Cruz-Vega, I. et al. Deep learning classification for diabetic foot thermograms. *Sensors* **20**(6), 1762 (2020).
- Lo, Z. et al. Development of an explainable artificial intelligence model for Asian vascular wound images. *Int. Wound J.* **21**(4), e14565 (2024).
- Elmannai, H. et al. Polycystic ovary syndrome detection machine learning model based on optimized feature selection and explainable artificial intelligence. *Diagnostics* **13**(8), 1506. (2023).
- Selvaraju, R. R. et al. Grad-CAM: Visual explanations from deep networks via gradient-based localization. *Proceedings of the IEEE international conference on computer vision*, 618–626. (2017).
- Das, S. K. et al. Diabetic foot ulcer identification: A review. *Diagnostics* **13**(12), 1998. (2023).
- Alzubaidi, L., Fadhel, M. A., Al-Shamma, O. M. R. A. N. & Zhang, J. Comparison of hybrid convolutional neural networks models for diabetic foot ulcer classification. *J. Eng. Sci. Technol.* **16**(3), 2001–2017 (2021).

21. Arumuga, M., Devi, T. & Hepzibai, R. Diabetic foot ulcer classification of hybrid convolutional neural network on hyperspectral imaging. *Multimedia Tools Appl.* **83**(18), 55199–55218 (2024).
22. Zhang, Y., Wang, J., Xu, S., Liu, X. & Zhang, X. MLIFeat: Multi-level information fusion based deep local features. In *Proceedings of the Asian conference on computer vision*. (2020).
23. Hong, S. et al. Personalized prediction of diabetic foot ulcer recurrence in elderly individuals using machine learning paradigms. *Technol. Health Care Preprint* 1–12. (2024).
24. Arnia, F. et al. Towards accurate Diabetic Foot Ulcer image classification: Leveraging CNN pre-trained features and extreme learning machine. *Smart Health* 100502. (2024).
25. Aldoulah, Z. A., Malik, H. & Molyet, R. A novel fused multi-class deep learning approach for chronic wounds classification. *Appl. Sci.* **13**, 11630 (2023).
26. Mayya, V. et al. Applications of Machine Learning in Diabetic Foot Ulcer diagnosis using Multimodal images: A review. *IAENG Int. J. Appl. Math.* **53.3**, 23 (2023).
27. Obayya, M. et al. Explainable artificial intelligence-enabled Teleophthalmology for diabetic retinopathy grading and classification. *Appl. Sci.* **12**, 8749 (2022).
28. Kim, J., Komondor, O., An, C. Y., Choi, Y. C. & Cho, J. Management of diabetic foot ulcers: A narrative review. *J. Yeungnam Med. Sci.* (2023).
29. Hu, F., Qiu, L. & Zhou, H. Medical device product innovation choices in Asia: An empirical analysis based on product space. *Front. Public. Health.* **10**, 871575. <https://doi.org/10.3389/fpubh.2022.871575> (2022).
30. Cheng, X. et al. Quercetin: a promising therapy for diabetic encephalopathy through inhibition of hippocampal ferroptosis. *Phytomedicine* **126**, 154887. <https://doi.org/10.1016/j.phymed.2023.154887> (2024).
31. Li, W. et al. The signaling pathways of selected traditional Chinese medicine prescriptions and their metabolites in the treatment of diabetic cardiomyopathy: A review. *Front. Pharmacol.* **15**, 1416403. <https://doi.org/10.3389/fphar.2024.1416403> (2024).
32. Wei, Y. et al. Exploring the causal relationships between type 2 diabetes and neurological disorders using a mendelian randomization strategy. *Medicine* **103**(46), e40412. <https://doi.org/10.1097/MD.00000000000040412> (2024).
33. Bilal, A., Liu, X., Shafiq, M., Ahmed, Z. & Long, H. NIMEQ-SACNet: a novel self-attention precision medicine model for vision-threatening diabetic retinopathy using image data. *Comput. Biol. Med.* **171**, 108099 (2024).
34. Bilal, A. et al. DeepSVDNet: a deep learning-based Approach for Detecting and Classifying Vision-threatening Diabetic Retinopathy in Retinal Fundus images. *Comput. Syst. Sci. Eng.* **48**(2). (2024).
35. Bilal, A. et al. Improved support Vector Machine based on CNN-SVD for vision-threatening diabetic retinopathy detection and classification. *Plos One* **19**(1), e0295951 (2024).
36. Bilal, A., Liu, X., Baig, T. I., Long, H. & Shafiq, M. EdgeSVDNet: 5G-enabled detection and classification of vision-threatening diabetic retinopathy in retinal fundus images. *Electronics* **12**(19), 4094 (2023).
37. Bilal, A., Zhu, L., Deng, A., Lu, H. & Wu, N. AI-based automatic detection and classification of diabetic retinopathy using U-Net and deep learning. *Symmetry* **14**(7), 1427 (2022).
38. Bilal, A., Sun, G., Mazhar, S., Imran, A. & Latif, J. A transfer learning and u-net-based automatic detection of diabetic retinopathy from fundus images. *Comput. Methods Biomech. Biomed. Eng. Imaging Visual.* **10**(6), 663–674 (2022).

Author contributions

Author Contributions: Conceptualization, P.S.R., A.K.; methodology, P.S.R., A.N., A.D., A.K.S.; software, P.S.R., A.N., A.D., A.K.S.; validation, A.N., and A.D.; formal analysis, A.K. and A.K.S.; investigation, A.K., A.N., and A.D.; resources, A.N., A.D., A.K.S.; data curation, P.S.R., A.K., A.N.; writing—original draft preparation, P.S.R., A.K.; writing—review and editing, P.S.R., A.N., A.D., A.K.S.; visualization, A.N., A.D.; project administration, A.N. A.D., and A.K.S.; Funding, A.N. All authors have read and agreed to the published version of the manuscript.

Funding

Open access funding provided by Manipal University Jaipur.

Declarations

Competing interests

The authors declare no competing interests.

Ethical approval

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

Additional information

Correspondence and requests for materials should be addressed to A.N.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025