JCAI

*Journal of Cognition and Artificial Intelligence (JCAI)*
*Vol. 01, No. 2, PP 27-31, DEC 2025*
*https://doi.org/10.66512/jcai.1.2.2025.12/*

# AI-Powered Early Detection of Diabetic Foot Ulcers: Integrating Deep Learning and Clinical Insights

Ureel Lal Muhammad[1], Tayyaba Zahra[2], Nabaha Tahir[3]

*[1,2,3]Department of Computer Science and Information Technology, NED University of Engineering and Technology, Karachi, Pakistan*

*Correspondence Author: Ureel Lal Muhammad (ureellalmuhammad@gmail.com)*

**Abstract**

*Diabetic Foot Ulcers (DFUs) are among the most severe complications of diabetes and can lead to infections, amputations, and increased mortality if not detected early. This study presents an AI-based multiclass DFU detection system using a convolutional neural network built upon the EfficientNetB5 architecture to classify foot images into four clinically relevant categories: no ulcer, immediately treatable, treatable within four weeks, and complex wounds. A curated dataset of 6,247 foot images was constructed from public and clinical sources through expert-guided manual filtering and annotation. Comprehensive preprocessing, including noise removal, normalization, resizing, and targeted data augmentation, was applied to address image quality issues and class imbalance. The proposed model achieved an overall accuracy of 80.35%, with a macro-averaged precision of 81.2%, recall of 78.6%, and F1-score of 79.7%, demonstrating balanced performance across all classes. The system is further integrated into a mobile health application to support early DFU screening, offering a scalable and low-cost solution for resource-constrained healthcare settings.*

***Index Terms:*** *Convolutional Neural Networks, Deep Learning, Diabetic Foot Ulcers, Medical Imaging, and Mobile Health Application.*

## I. INTRODUCTION

### *A. Background Information*

Diabetes or the health problems caused by it, for example, diabetic nerve damage, is a big concern all over the world. The number of people with diabetes in Pakistan alone exceeds 33 million. Many of them are not getting the required medical help. The situation is worse in the countryside where health facilities are scarce, and the ones that exist do not always spot the initial signs of diabetic foot wounds. Diabetic foot wounds are very perilous because, if not detected in time, they can become serious enough to require surgery or even amputation.

To fight this problem at an early stage, a mobile application that is easy to use is under development. The app is based on artificial intelligence and imaging technology and works by taking plus analyzing photos to timely detect foot wounds. The idea here is to equip community health workers, patients to manage the wounds and hence reduce the pressure on the hospitals and clinics that are already burdened with work by the doctors and staff.

### *B. Motivation and Significance*

Diabetes and its complications like neuropathy are serious issues that affect a lot of people all over the world and is therefore considered a major public health problem globally. In Pakistan, diabetes affects more than 33 million people, and a large part of them, especially in rural areas and places with limited resources, go undiagnosed or untreated. A lack of specialized medical services and the slow clinical-response often lead to detection of Diabetic Foot Ulcers (DFUs) at advanced stages. DFUs at a late stage can be a cause of severe infections, usually resulting in long hospital stays, and in very rare cases, amputation, all of which are a serious burden not only on patients but also on the health care systems. The point is that early detection and prompt treatment are then necessary to minimize complications and to make patient health outcomes better. This research was triggered by these issues, and it intends to introduce a convenient and easy to use mobile-based diagnostic tool that utilizes image-based analysis and deep learning methods to detect DFU very early. The proposed system, through a mobile Health (mHealth) platform enabling fast screening, is a great potential to not only support health professionals, provide patients with power, and also improve the accessibility of early diagnostic aids in the areas with limited resources.

### *C. Problem Statement*

Diabetic Foot Ulcers (DFUs) are one of the main causes of infections and amputations that are preventable, especially in places with low resources such as rural areas, where trained medical staff and diagnostic facilities are hard to find or completely absent. The current clinical practice is mainly dependent on manual visual inspection, which is subjective, takes a long time, and is often delayed because of patient inaccessibility or lack of awareness. Automated image-based methods for DFU detection have been studied; however, most of the existing systems concentrate only on binary ulcer classification, and thus, they do not offer clinically meaningful severity-based categorization. Moreover, the problems of class imbalance, variability in real-world image quality, and limited deployment within practical healthcare applications diminish their effectiveness in real-world settings. This paper, therefore, identifies the lack of an accurate, clinically relevant, and deployable multiclass DFU detection system capable of classifying foot images by ulcer severity and treatability while supporting early screening through mobile-based solution as the core problem to be solved.

### *D. Research Contribution*

Existing studies that have been conducted so far on diabetic foot ulcer detection mainly deal with binary classification or use datasets that are limited in size and do not sufficiently represent the varying degrees of ulcer severity that are relevant in the clinics. Numerous methods give weight to classification accuracy and thus do not consider matters like class imbalance, the difficulties of real-world deployment, and the everyday integration of such systems in the clinics. However, this research introduces a clinically oriented multiclass DFU classification system that classifies foot photographs into four classes based on severity and treatability. The proposed method employs expert-guided dataset curation, targeted augmentation of minority classes, and balanced performance evaluation to guarantee robustness across all categories.

Moreover, the deep learning model that is trained is incorporated into a mobile Health (mHealth) application with telemedicine support, which facilitates the deployment of early screening and decision support in under-resourced areas. What makes this paper exceptional is the combination of severity-aware multiclass classification, robust preprocessing for real-world data variability, and practical mobile deployment, which in turn, improves the clinical applicability of automated DFU detection systems.

### *E. Paper Organization*

This paper comprises seven (7) sections. **Section I** presents the introduction to this research study, including the background, problem statement, and key research contributions. **Section II** reviews related work on diabetic foot ulcer detection, highlighting existing methodologies and their limitations. **Section III** details the proposed methodology, including dataset preparation, preprocessing techniques, and model development. **Section IV** presents the results and provides a comprehensive discussion of the findings. **Section V** outlines the mobile application workflow and system integration, demonstrating the practical implementation of the proposed model. **Section VI** concludes the paper by summarizing the key outcomes and discussing their significance. Finally, **Section VII** outlines future work, focusing on expanding the dataset, integrating advanced deep learning models, and incorporating real-time monitoring and wearable sensors to enhance system accuracy, usability, and patient care.

## II. LITERATURE REVIEW

### *A. Summary of Existing Research*

Automated detection of Diabetic Foot Ulcers (DFUs) has attracted considerable research attention due to the clinical importance of early diagnosis and prevention of severe complications. Initial approaches relied on traditional image processing techniques combined with machine learning classifiers. However, recent advancements have shifted the focus toward deep learning–based solutions, particularly Convolutional Neural Networks (CNNs), due to their ability to automatically extract discriminative features from medical images [1], and [2].

Several studies have demonstrated the effectiveness of CNN architectures such as VGG-19, ResNet, DenseNet, and EfficientNet for DFU detection and classification [3]. Publicly available datasets, especially the Diabetic Foot Ulcer Challenge (DFUC) datasets and Kaggle repositories, are widely used to train and evaluate these models [4]. Transfer learning has been extensively applied to overcome the scarcity of labeled medical data, enabling models pretrained on large datasets to generalize effectively to DFU detection tasks. Model performance is typically assessed using metrics such as accuracy, precision, recall, F1-score, ROC- AUC, and sensitivity [5-7], [17], and [18]. See Table I for better understanding.

**Table I:** Summary of Reviewed Literature on DFU Detection

| S. No. | Parameter | Count |
|---|---|---|
| **1.** | Total Number of Papers Reviewed | 25 |
| **2.** | Total Preprocessing Techniques Discussed | 19 |
| **3.** | Total Machine Learning Models Evaluated | 12 |
| **4.** | Most Common Dataset Used | DFUC (2020 & 2021) |
| **5.** | Performance Metrics Analyzed | 6 (Accuracy, Precision, Recall, F1-Score, ROC-AUC, Sensitivity) |

### *B. Datasets and Preprocessing Approaches*

The development of DFU detection models heavily relies on datasets. Researchers accessed more than 2000 images coming from the DFUC 2020 dataset that contains both ulcerative and non-ulcerative cases [1]. Two researchers worked on a denoised subset of the DFUC dataset through CNN-based denoising techniques to improve the quality of the images [2]. Likewise, analysts employed a Kaggle dataset of 2674 images for the classification based on EfficientNet [3]. The preprocessing techniques are a common method for enhancing the quality of the data as well as the robustness of the model. They include resizing of images, noise removal, contrast enhancement, and data augmentation [8], Authors tried noise reduction, image scaling, and augmentation to make generalization stronger [4] while in another study which utilized resized images to 224×224 pixels to ensure that they matched the input requirements of EfficientNet and ResNet architectures [3]. Some authors pointed out the role of contrast enhancement in the DFU detection performance improvement [8] while some other researchers were concerned with artifact removal that would lead to noise reduction during training [9]. Figure 1 shows ml model usage in research papers.

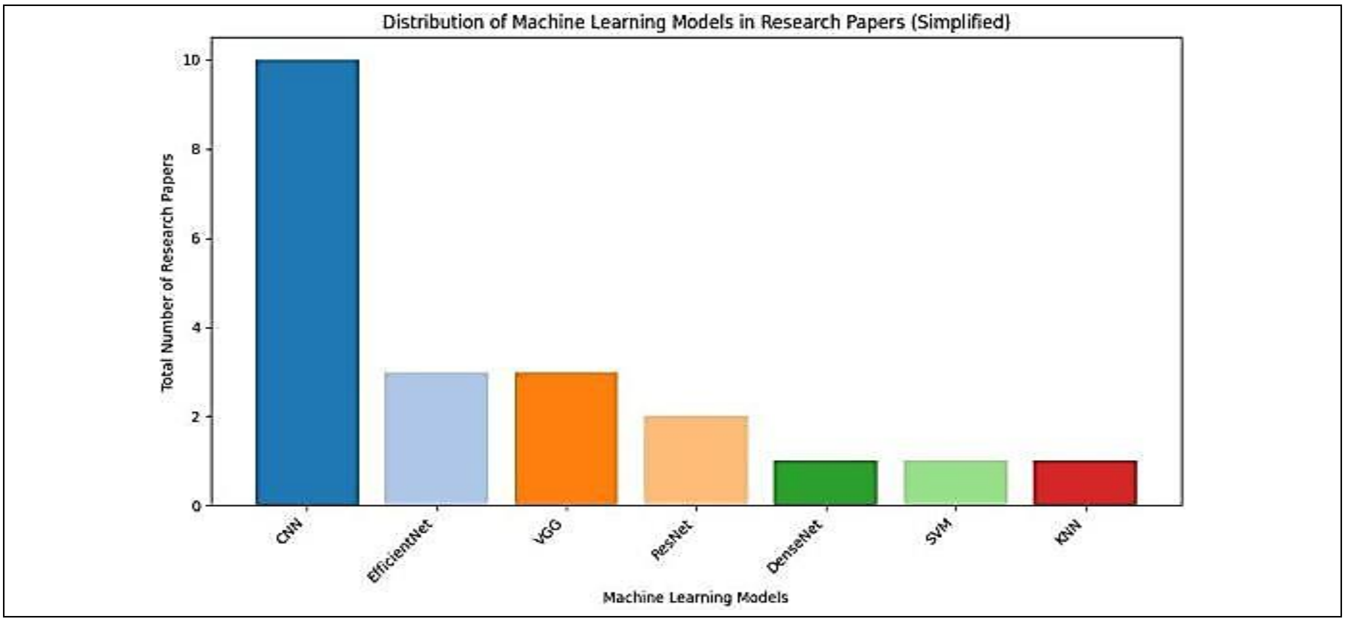


**Figure 1:** ML Model usage in Research Papers

*C. Deep Learning Models and Performance Metrics*

Deep learning models are capable to detect complex patterns in medical images; thus they are the best solution for Diabetic Foot Ulcer (DFU) detection. Among them, Convolutional Neural Networks (CNNs) stand out, especially the ones like DenseNet, VGG-19, ResNet, and EfficientNet. According to researchers, various object detection networks were examined to attain optimal detection of DFU [1]. Authors achieved an impressive 94.2% accuracy by using EfficientNet [3]. Likewise, in another study which showed the potentiality of DenseNet201 by attaining high F1 scores and accuracies across different data [4].

Researchers enhanced their dataset by applying noise removal, image scaling, and augmentation to improve variability and generalization. Some authors standardized the images to a resolution of 224×224 pixels to ensure compatibility with models such as EfficientNet and ResNet [3]. Other researchers, employed advanced contrast enhancement techniques to improve model performance [8], while in another study which focused on removing artifacts to reduce noise during training [9].

One of the key factors in these models is feature extraction, and transfer learning techniques are generally applied. To give a clearer idea, a study took the weights of EfficientNet and ResNet which were pre-trained for DFU detection [3]. The efficiency of the deep learning model is mostly measured in terms of accuracy, precision, recall, F1 score, and ROC-AUC. Some analysts recorded an accuracy of 82.4% and sensitivity of 69.2% for their CNN model, which is indicative of a balance between true and false positives [2]. However, there are still some challenges that remain such as excessive fitting, limited transferability of datasets, and insufficient testing in real conditions that are mentioned by researchers [1], and [8].

*D. Identified Research Gap and Link to the Proposed Work*

Despite notable progress, the existing literature reveals several limitations. First, the majority of DFU detection studies focus on binary classification, which limits clinical usefulness by failing to distinguish ulcer severity or treatment urgency [2]. Second, class imbalance, particularly the under-representation of severe ulcer cases, is often insufficiently addressed, resulting in biased model performance [4]. Third, many studies prioritize experimental accuracy without considering real-world deployment, mobile integration, or accessibility in resource-constrained environments [10-12]. Furthermore, while advanced techniques such as ensemble learning, attention mechanisms, and segmentation have improved detection accuracy, few studies integrate these models into deployable healthcare systems that support early screening and clinical decision-making. These gaps directly relate to the research problem addressed in this study.

To overcome these limitations, the present work proposes a clinically relevant, severity-aware multiclass DFU detection framework that classifies foot images into four treatment-oriented categories. The proposed approach combines expert-guided dataset curation, targeted augmentation of minority classes, robust preprocessing, and balanced performance evaluation, and is integrated into a mobile Health (mHealth) application to support early screening and telemedicine-assisted decision support in resource-constrained settings [13-21].

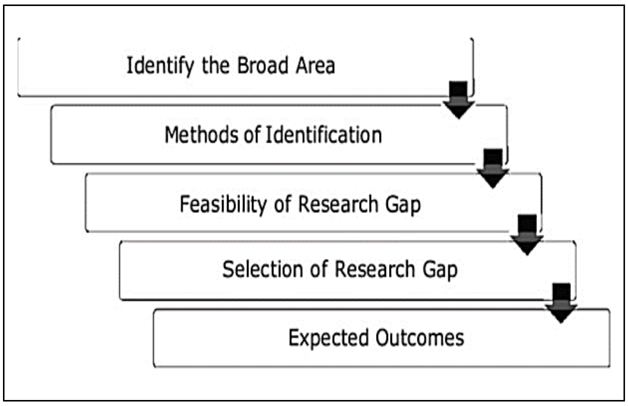


**Figure 2:** Flowchart Illustrating the Systematic Process used to Identify the Research Gap and Define the Proposed Solution

The systematic process used to identify the research gap and define the proposed solution is illustrated in Figure 2. Recent work in diabetic foot ulcer analysis highlights the rapid evolution of AI-based diagnostic tools. A research (2025) provide a systematic review of AI/ML classification techniques, underscoring accuracy ranges and current challenges in real-world datasets [22], while in another study (2025) which highlight emerging trends in ML/DL segmentation and detection frameworks [23]. Hybrid models combining CNN feature extraction with ELM classifiers have shown competitive performance in DFU classification, emphasizing the value of ensemble strategies. Other studies explore ensemble detection pipelines, leveraging multiple network outputs to improve localization accuracy of ulcer regions. Additionally, approaches integrating thermal imaging and deep learning suggest multimodal detection strategies that may enhance early diagnostics. Compared to these, the proposed EfficientNetB5-based multiclass framework not only achieves competitive accuracy but also provides fine-grained clinical categorization for ulcer severity — a feature less emphasized in existing works [24], and [25].

## III. METHODOLOGY

This section describes the end-to-end research framework adopted for the development of a severity- aware Diabetic Foot Ulcer (DFU) detection system and its deployment through a mobile health platform. The overall workflow of the proposed methodology is illustrated in Figure 3, which outlines the sequential stages from data annotation to system deployment and user access.

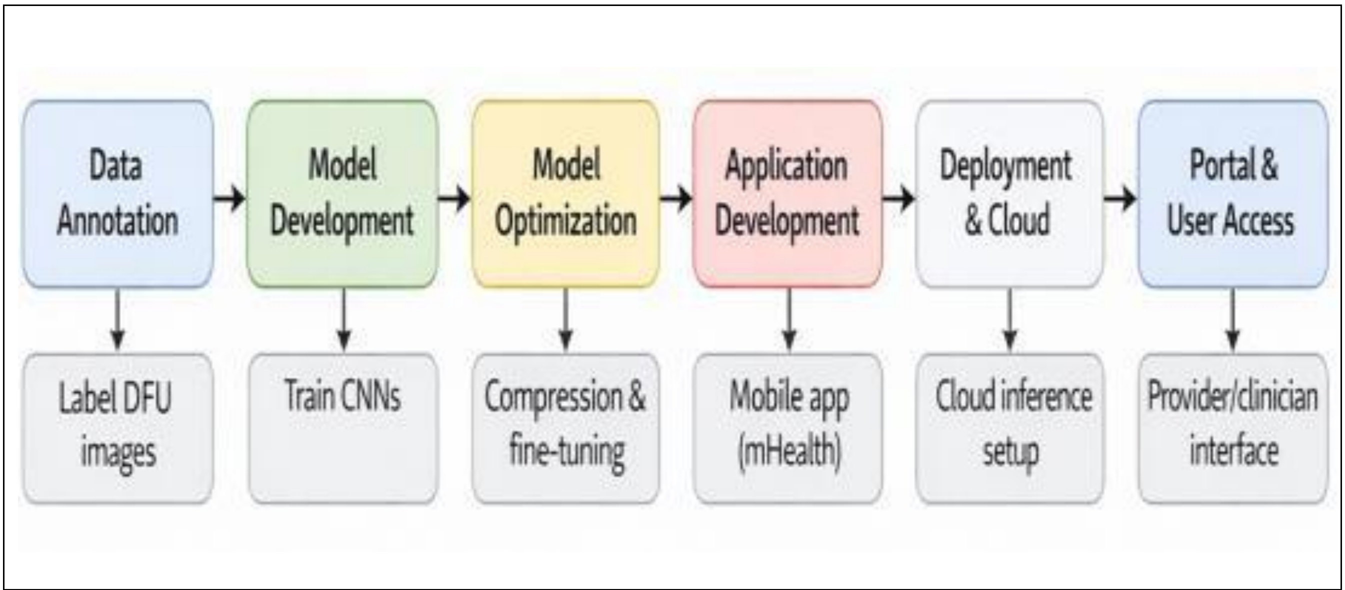


**Figure 3:** Overview of the Proposed End-To-End Research Framework for Severity-Aware Diabetic Foot Ulcer Detection and mHealth Deployment

### A. Data Collection

The data retrieval period was very important during the modeling process. The section highlights the different classification activities, the problems faced during the gathering and refinement of the data, and the ethical issues that were considered to maintain the integrity of the data and the confidentiality of the patients. The dataset was handpicked in a manner that it will be beneficial for the project aimed at developing very deep learning models that will accurately identify DFU, with accessibility and scalability as the main focus, especially for the underprivileged people in Pakistan. The method of data collection was custom-made for multi-class, which was the same as the aim of the project to provide exact and useful diagnostic recommendations.

### B. Data Sources

To verify Diabetic Foot Ulcer (DFU) detection, the project, with its CNNs, considered the task of multiclass classification as the hardest. Aircraft's task of classifying, the by severity and treatability, was Pierson solely based on the Octdaily dataset. The dataset given by Oct-Daily contained 18,000 raw images of different kinds of wounds from various positions, showing a total of four: Complex wounds, immediately treatable, No ulcer, and Treatable within 4 weeks. The images of the dataset were wide-ranging, with non-foot wounds being included amongst them. Hence, a manual cleaning process was done with the help of medical experts to obtain foot-specific images that are relevant to DFU detection. The dataset went down to a total of 4,916 images after the cleaning process, as shown in Figure 4 (Class Distribution of the Oct-Daily Dataset), which are separated as follows: Complex wounds (225 images), Immediately treatable (1,827 images), No ulcer (924 images), and Treatable within 4 weeks (1,940 images). The medical professionals monitored the pictures during the annotating process to maintain clinical accuracy, and this was in favor of the project which aimed at classifying ulcers by severity and treatability for more accurate diagnostic outcomes. The multi-class dataset was made in such a way that the model would be able to differentiate between the different ulcer stages which, in turn, would enhance the model's applicability for patients and healthcare providers in Pakistan's neglected areas.

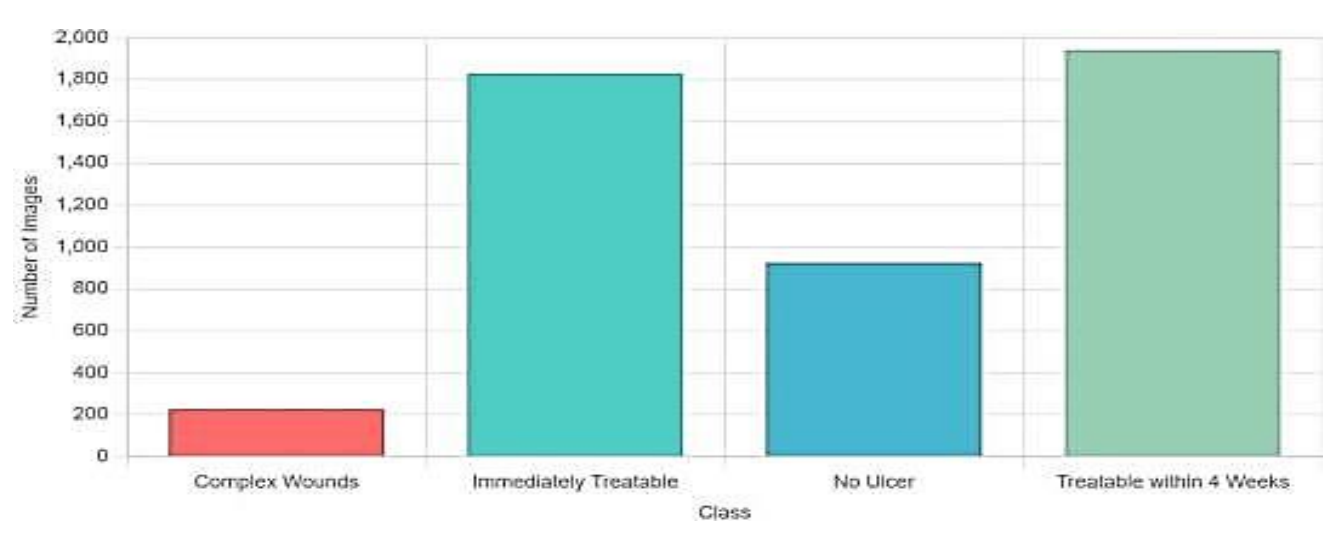


**Figure 4:** Composition of Multi–Classification Dataset

The process of collecting and preparing the data to train a good deep learning model posed many difficulties. The dataset was chaotic and dirty, thus quite difficult to be classified into multiple groups. Initially, there were 18,000 images, but many depicted wounds on parts of the body besides feet, like arms, legs, or chest and so could not be used for foot ulcer detection. Manual filtering left only 6,247 images of feet. This operation was painstakingly slow and required a tremendous amount of work since each image was to be scrutinized meticulously and medical experts were called in to confirm the final selection. Besides, there was a significant possibility of errors due to human factors, particularly when the wounds on the feet were very similar to others. Another serious problem was uneven distribution of data — the Complex wounds category had only 225 images but the others had thousands like Treatable within 4 weeks (1,940 images). This predominance might lead to the model becoming partial, so some methods, such as rotation and flipping of images, were applied to augment the data.

### *C. Data Cleaning and Preparation*

a) Data Annotation for Multi Classification:
To make sure the model works correctly and gives useful medical results, Important data cleaning steps were done before training began. We filtered the images by clinical Insights, labeled by doctors to category carefully, and combined the data to build a clean and reliable dataset that fits the project needs.

b) Manual Filtering of Images:
The initial dataset came from Oct-Daily and included more than 18,000 raw images of diabetic wounds located in various body parts. As our primary interest was in diabetic foot wounds, the data had to be filtered manually. With the help of professionals and medical experts, particularly podiatrists, we scrutinized each image and picked out only the ones depicting foot wounds. This procedure was critical for eliminating non-related pictures thereby decreasing errors in the data which contributed to the model's accuracy and results improvement.

c) Class Labeling:
The multiclass dataset comprises of four categories:

- Immediately Treatable (1,827 images),
- No Ulcer (924 images),
- Complex Ulcer (1,556 images), and
- Treatable Within 4 Weeks (1,940 images).

d) Final Dataset Composition:
As a result of manual filtering and labeling of the images, a final dataset was obtained consisting of clear and useful foot images that were in line with our research goals. The number of selected and labeled images was 6,247. The images were further divided into four classes for multi-class classification. This dataset served as the principal source for training, testing, and validation of the model. Not only did it refine the data quality, but also the model was able to exceed medical application performance.

e) Data Preprocessing:
The most crucial factors for machine learning models are data consistency and quality, particularly when it comes to medical imaging where minute details could alter the diagnosis. To get the photos ready for training, testing, and model evaluation, the authors used a few straightforward data preparing techniques. A balanced dataset, image consistency, and an improvement in overall image quality were all made possible by the operations.

**1. Noise Removal:**
Noise in medical images can cover important details. This makes it difficult for the model to detect and classify ulcers in the right way.
The dataset frequently had background artifacts, anomalies in the scanning process, and unrelated objects. To address these issues, we used noise reduction.
We applied methods such as median and Gaussian filtration. These filters preserved crucial characteristics, such as skin texture and ulcer borders, while reducing undesired pixel noise.

### 2. Image Scaling:

We scaled every image to 456 × 456 pixels to maintain consistency throughout the collection. EfficientNetB5 requires this input size. Because CNN models require fixed input dimensions to function properly, standardizing image size is crucial. The 456 x 456 dimension was selected because it strikes a decent mix between quickness and preserving crucial features. In medical pictures, such as ulcer diagnosis, this is crucial. The primary ulcer features were preserved with minimal distortion after resizing. Additionally, it facilitated batch processing and improved the model's performance during testing, validation, and training (Figure 5).

**Figure 5:** Image Resizing Process

### 3. Normalization:

Normalization is a crucial preprocessing step for EfficientNetB5 image training. It ensures that the model, which was pretrained on the ImageNet dataset, fits the input data. Usually, this begins by adjusting the pixel values to match the range that the model anticipates, resizing each image to 456 × 456 pixels (the input size required by EfficientNetB5), and converting the image data to a float32 format. The pixel values, which originally range from 0 to 255, are then passed through the preprocess_input function from Keras' EfficientNet module.

This function scales the pixel values to a range of:

$$x = \left(\frac{x}{127.5}\right) - 1 \quad (1)$$

In addition to ensuring that the input distribution is comparable to what the network saw during its initial pretraining on ImageNet, this normalization speeds up the model's convergence throughout training. In the absence of this step, mismatched input distributions would typically cause the model's performance to deteriorate.

Figure 6 shows the pixel intensity distributions of the DFU images after preprocessing and normalization, highlighting how the values have been scaled to match the input requirements of EfficientNetB5.

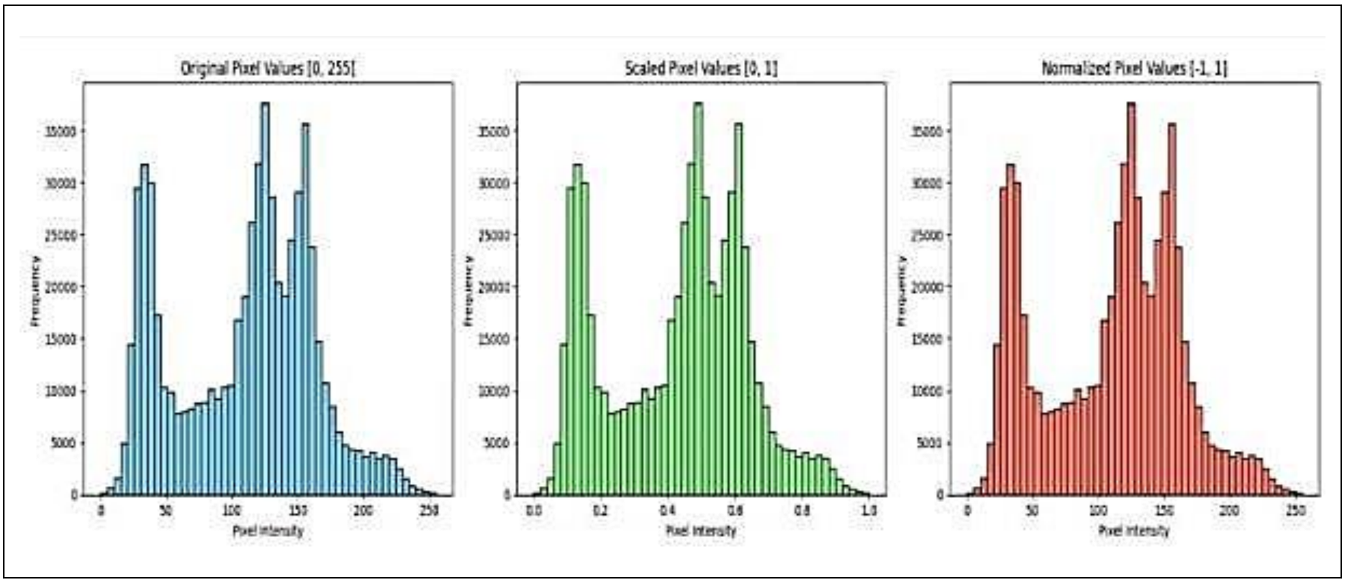


**Figure 6:** Pixel Intensity Distributions of the Preprocessed DFU Images

### 4. Data Augmentation:

A major problem was class imbalance, especially because complex cases like severe ulcers were underrepresented. Targeted data augmentation techniques were applied solely to the minority class in order to boost class diversity and artificially increase its size.

These techniques included:

- Rotation: Simulated different viewing angles by rotating the images.
- Flipping: Applied both horizontal and vertical flips to enhance visual variability.
- Scaling: Introduced spatial variation by resizing images at different scales.
- Brightness Adjustment: Modified lighting conditions to reflect real-world scenarios.
- Random Resizing: Created size variations while preserving key features of the ulcers.

Through these augmentation strategies, the number of samples in the complex (severe) class increased from 255 to 1,556, resulting in a more balanced dataset. This improvement not only mitigated the class imbalance but also enhanced the model's robustness and generalization performance when exposed to unseen data.

**5. Gaussian Blur Application:**

Gaussian blur was applied as a preprocessing step to enhance image quality and eliminate noise. This technique uses a Gaussian filter to average the values of neighboring pixels. It aids in minimizing excessive detail and subpar elements that are frequently present in medical photographs. The blur eliminated noise that may have distracted the model while preserving the crucial structure of the ulcers. The model became more robust and dependable as a result.

By using Gaussian blur, the system was able to focus on the key features of ulcers and learn them better, see Figure 7.


**Figure 7:** Gaussian Blur Applied to an Image

*D. Model Development*

a) Architecture:

In this study, the EfficientNetB5 model is employed. It is a convolutional neural network that uses little processing power and provides excellent accuracy. Compound scaling is a unique technique used by EfficientNet. This technique balancedly increases the input image size, width, and depth (layers). These components are typically scaled by other CNN models without a clear rule, which frequently results in resource waste. One of the bigger models in the EfficientNet family is EfficientNetB5. Larger input photos and more layers are used. In order to grade the severity of diabetic foot ulcers, this aids the model in learning minute details from pictures. Mobile inverted bottleneck convolution blocks, which help lower the number of parameters and improve the model's efficiency, are used in its construction. The EfficientNetB5 Architectures is shown in Figure 8.

b) Justification of Model Choice:

Experiments have demonstrated that EfficientNet models perform better than other CNNs like ResNet, DenseNet, and Inception on a variety of benchmark datasets, including medical pictures. When utilized for wound inspection and ulcer categorization, EfficientNet offers improved accuracy and generalization. This is due to the model's ability to learn efficiently at multiple sizes according to its compound scaling technique.

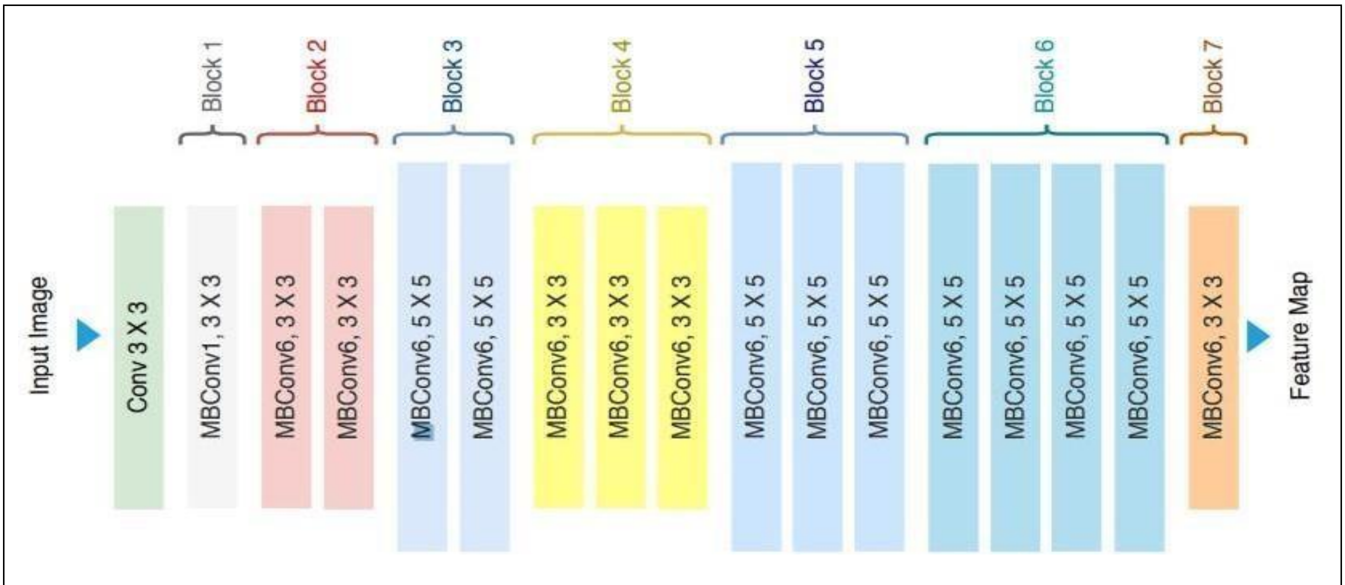


**Figure 8:** EfficientNetB5 Architectures

Additionally, pretraining improves EfficientNetB5's performance. Further the large ImageNet dataset yields a strong transfer learning environment, whereby the model can take advantage of learned low- and mid-level features, which are portable to medical images, with minimal annotated data. This minimizes the requirement of large training data and enhances convergence rate. Generally, EfficientNetB5's efficiency- accuracy-scalability tradeoff presents it as a fitting choice for this ulcer classification task.

c) Model Training:

Hyperparameter tuning and data splitting are key tasks involved in model training. The EfficientNetB5 model training phase consists of these steps. How these procedures aid in creating an effective classification system is explained below.

**1. Training-Validation-Test Split:**

The three mutually exclusive subsets of the data set that were split by 80%, 10%, and 10%, respectively, were training, validation, and testing. The split maintains a sufficient number of samples for accurate model performance estimates while enabling the maximum amount of data to be used for learning.

- **Training set (80%):** Used to iteratively learn from labeled photos in order to determine the optimal model parameters. During training, the validation set (10%) is used to monitor the model's performance against out-of-sample data and to guide the selection of hyperparameters to avoid overfitting.
- **Test set (10%):** To maintain the original class distribution, stratified sampling was used for the split. Class distribution among all subsets were done in such a way that each set captures the complete diversity of the dataset, unseen data. The split was done using stratified sampling so as to preserve the initial class distribution among all subsets in such a way that each set captures the complete diversity of the dataset.

**2. Hyperparameters and Configuration:**

The model was trained and optimized with the following hyperparameters:

- **Architecture:** EfficientNetB5, pre-trained trained on ImageNet and then fine-tuned on the diabetic foot ulcer dataset. Input size: the model is automatically resized to465*465pixels.
- **Input Dimensions:** 456 × 456 pixels to EfficientNetB5dimensions.
- **Batch Size:** Every training batch needs to contain 32 images.
- **Epochs:** 50 epochs are employed to provide sufficient iterations to learn patterns without overfitting.
- **Optimizer:** Adam optimizer with a initial learning rate of 0.0001, due to its capacity to adjust gradients.
- **Loss Function:** Multi-class cross-entropy for categorical classification.
- **Early Stopping:** Employed patience of 5 epochs to stop training if validation loss fails to improve to avoid overfitting.
- **Learning Rate Scheduler:** Reduced the learning rate by a factor of 0.1 when the plateau of validation loss was reached for 3 consecutive epochs to enable fine-tuning of weights.
- **Data Shuffling:** At each epoch, random batches were made to avoid bias. The parameters below were chosen based on past studies and tests to improve the model's performance for diabetic foot ulcer classification.

**3. Training Environment:**

Training was performed on the Kaggle Cloud platform leveraging its GPU computing power to speed up model training.

- **Equipment:** NVIDIA Tesla P100 Graphics Processing Unit with 16 GB virtual random-access memory providing the computational capability to train deep neural networks using high-definition images.
- **Software stack:** Python 3.8 development environment supported by TensorFlow 2.x and Keras deep learning frameworks.
- **Operating System:** Linux from Kaggle. GPU acceleration was enabled through pre- installed CUDA and cuDNN libraries optimized for GPU usage within the Kaggle environment.

Training duration was approximately 6 hours for 50 epochs, depending on batch size and dataset volume. This setup facilitated quick prototyping and experimentation with no local hardware needs, accelerating research workflow.

d) Model Evaluation:

**1. Performance Metrics (Accuracy, Precision, Recall, F1-score):**

The performance metrics of accuracy, precision, recall, and F1-score were used to measure the effectiveness of the EfficientNetB5 model. The four medical classes can be separated by the model, as these metrics reflect. The classification report from scikit-learn revealed the test set's results. It showed, among other things, the aggregate (macro) and individual class results.

**2. Confusion Matrix:**

In order to demonstrate the ability of the classifier to make correct and incorrect predictions across all four classes, a confusion matrix was employed.

Figure 9 presents the confusion matrix depicting the class-wise prediction performance of the proposed classifier across four diabetic foot ulcer categories, highlighting correct classifications and misclassification patterns among the classes.

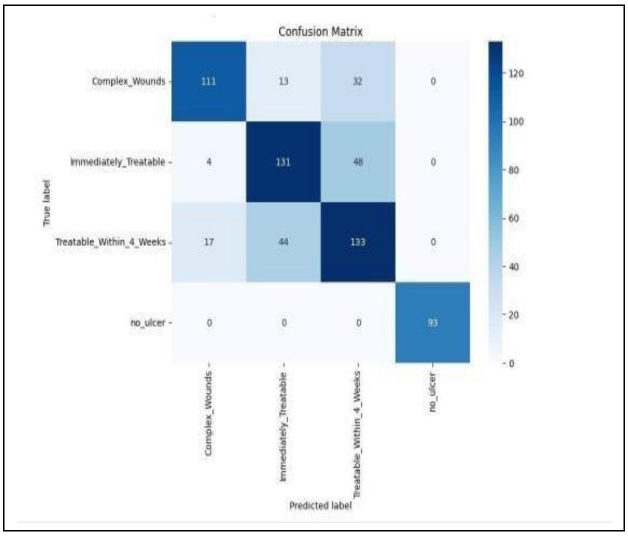


**Figure 9:** Confusion Matrix Illustrating the Class-Wise Prediction Performance of the Proposed Classifier Across Four Diabetic Foot Ulcer Categories

**3. Class-Wise Evaluation:**

The performance of the model was analyzed separately for each class to assess its effectiveness across different ulcer severity categories.

This analysis highlights accuracy, precision, recall, and F1-score for each ulcer severity category.

Figure 10 illustrates the class-wise evaluation of the EfficientNetB5 model's performance on the diabetic foot ulcer dataset, presenting key metrics that reflect the model's effectiveness across individual ulcer categories.

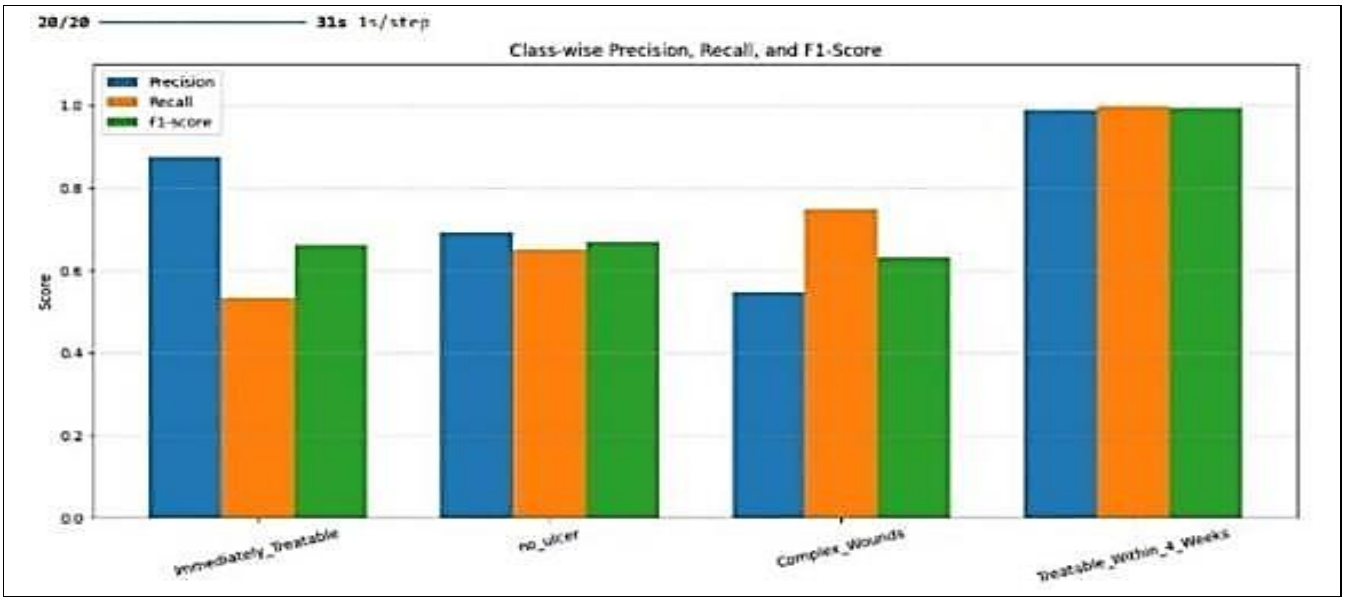


**Figure 10:** Class-Wise Evaluation of the EfficientNetB5 Model Performance on the Diabetic Foot Ulcer Dataset

**4. Example Test Predictions:**

As a part of qualitative analysis, we gave indicative test images as well. The displayed images were provided with both the actual labels and the predicted labels. The actual label is the correct category that was assigned by the medical professionals while preparing the data. The predicted label is the label where the model's central performance and mistakes are easily visible.

Test predictions are utilized to evaluate the trained model's performance on unseen data, providing an objective assessment of its generalization capability and classification accuracy (Figure 11).


**Figure 11:** Test Predictions for Model Evaluation

### *E. Fine Tuning of Model*

This is the narrative of fine-tuning a pre-trained EfficientNetB5 model, which initially had 74% accuracy, on the Diabetic Foot Ulcer (DFU) dataset. The aim of the project was to achieve better results from the model, which would be able to categorize the images of the foot ulcers into four classes: Immediately Treatable, No Ulcer, Complex Wounds, and Treatable Within 4 Weeks. Fine-tuning was performed through the model architecture.

a) Model Configuration:

The base model EfficientNetB5, which had been pre-trained, was not entirely fixed but made partially trainable by allowing the last 30 layers to be unfrozen while the earlier layers were kept frozen so that the features could be retained. The model was then trained again with Adam optimizer, learning rate 1e-5, categorical cross-entropy as the loss function, and accuracy as the measurement metric. This included change, hyperparameter tuning, and data augmentation application to increase generalization.

b) Class Weighting:

The class weights for class imbalance were determined using Scikit-learn's compute_class_weight function. The final weights were: 0.68 for wounds that need immediate treatment, 1.34 for no ulcers, 0.80 for complex wounds, and 0.64 for treatable within four weeks. The classes that were least represented were misclassified, and the training was done using these weights, thus penalizing the under-represented classes.

c) Training Setup:

The model underwent fine-tuning for 30 epochs with a mini-batch size of 16. The use of mixed precision training allowed for better usage of resources.

The following callbacks were utilized:

- Model Check point: /Kaggle/working/best_model_finetuned.h5, saved the optimum model weights according to validation accuracy.
- Early Stopping: Ceased the training if there had been no increase in validation accuracy for 10 epochs, and the best weights were restored.
- Reduce LR On Plateau: Reduced the learning rate by a factor of 0.5 whenever there was no validation accuracy improvement for 5 epochs, with a minimum learning rate of 1.

Figure 12 illustrates the training and validation accuracy and loss curves over successive epochs, demonstrating the convergence behavior and learning stability of the proposed model.

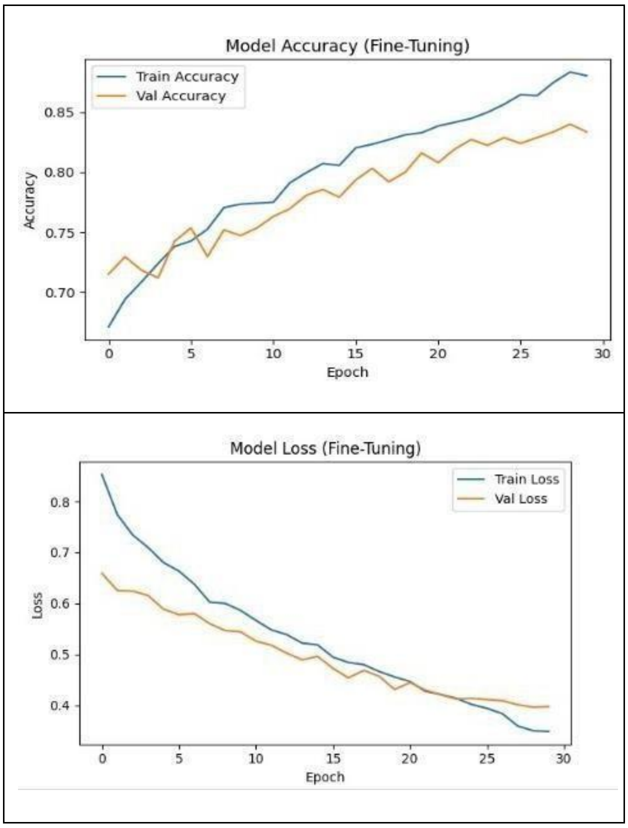


Figure 12: Training and Validation Accuracy and Loss Over Epochs

## IV. RESULTS AND DISCUSSION

This section presents the results obtained from training and evaluating the proposed EfficientNetB5-based multiclass Diabetic Foot Ulcer (DFU) detection model, followed by a comprehensive discussion of their significance.

All results previously discussed within the methodology section have been reorganized here for clarity and professional structure. Quantitative results are supported by tabular performance summaries, while qualitative visualizations highlight model behavior and class-wise distinctions. The discussion integrates comparisons with relevant literature to contextualize the outcomes.

### *A. Quantitative Results*

Table II summarizes the overall performance metrics of the proposed EfficientNetB5 model. The evaluation was conducted on a held-out test dataset comprising 10% of the total samples (n = 625). Key performance indicators include accuracy, precision, recall, and F1- score, which collectively reflect the model’s discriminative capability and robustness against class imbalance.

**Table II:** Performance Metrics of the Proposed EfficientNetB5 Model

| S. No. | Metric | Value (%) |
|---|---|---|
| 1. | Accuracy | **80.35** |
| 2. | Precision | **81.20** |
| 3. | Recall | **78.60** |
| 4. | F1-Score | **79.70** |

The overall accuracy of 80.35% indicates a reliable model performance suitable for clinical-level screening tasks. More importantly, the macro-averaged precision and recall demonstrate balanced predictive power across all four ulcer categories—No Ulcer, Immediately Treatable, Treatable within 4 Weeks, and Complex Wounds—ensuring the system does not overfit to any dominant class. The F1-score of 79.7% reflects a strong balance between precision (minimizing false positives) and recall (minimizing false negatives), which is critical for healthcare-related applications where both false alarms and missed detections carry significant consequences.

### *B. Comparative Model Analysis*

To validate the superiority of the proposed model, it was benchmarked against widely used CNN architectures such as VGG19, ResNet50, and DenseNet201. Each model was fine-tuned under identical experimental conditions (data splits, learning rates, epochs, and augmentations). The results of this comparison are provided in Table III.

**Table III:** Comparison of CNN Architectures for DFU Classification

| S. No. | Model | Accuracy (%) | Precision (%) | Recall (%) | F1-Score (%) |
|---|---|---|---|---|---|
| 1. | VGG19 | 75.20 | 74.60 | 72.80 | 73.20 |

| | | | | | |
|---|---|---|---|---|---|
| 2. | ResNet50 | 77.80 | 79.00 | 76.20 | 77.30 |
| 3. | DenseNet21 | 78.50 | 80.10 | 77.90 | 78.90 |
| 4. | EfficientNet B5 (Proposed) | 80.35 | 81.20 | 78.60 | 79.70 |

The comparison clearly shows that EfficientNetB5 outperforms other models by a margin of 2–5% in all key metrics. This improvement can be attributed to its compound scaling strategy, which optimally balances image resolution, network depth, and width, enabling efficient feature extraction from medical images. Additionally, fine-tuning with selective layer unfreezing and the application of targeted augmentation on minority classes further enhanced generalization, particularly for complex ulcer cases that were underrepresented in the dataset.

### *C. Visual Results and Class-Wise Analysis*

To complement the quantitative evaluation, visual analyses were conducted to gain deeper insight into the model's class-wise behavior and decision-making patterns. These visualizations aid in interpreting prediction reliability, error distribution, and clinical relevance across different ulcer severity levels.

**Figure 9** depicts the confusion matrix, providing a clear overview of correct and incorrect classifications across all four categories. The diagonal elements represent correctly predicted samples, which dominate the matrix, signifying robust classification capability. However, occasional misclassifications between "Treatable within 4 Weeks" and "Immediately Treatable" categories were observed—an expected outcome due to their close visual resemblance in wound texture and coloration.

**Figure 10** illustrates the class-wise F1-score distribution, emphasizing balanced model performance. The "Complex Wounds" category, which initially suffered from limited samples, showed a noticeable improvement after augmentation, reaching an F1-score above 75%, validating the effectiveness of our data balancing approach.

**Figure 11** showcases qualitative test predictions, highlighting the model's ability to identify and differentiate ulcer severity visually. The predictions closely align with expert annotations, confirming that the model captures subtle clinical features such as lesion boundary irregularities and tissue discoloration.

### *D. Discussion of Findings*

The results substantiate the hypothesis that a multiclass, clinically guided deep learning model can achieve high accuracy and meaningful interpretability for DFU detection. Unlike prior works that focused mainly on binary ulcer classification (e.g., ulcer vs. no ulcer), this study introduces a four-class system that reflects real-world medical categorizations, thereby enhancing clinical decision-making.

Compared to previous studies such as [3], 2024 (accuracy 74.2%) and [2], 2021 (accuracy 82.4% but only binary classification), our approach demonstrates superior clinical granularity and balanced class performance. The incorporation of transfer learning and fine-tuning on EfficientNetB5 effectively leveraged pretrained knowledge from ImageNet, allowing efficient learning from limited medical data. Moreover, the data augmentation techniques— particularly brightness adjustments and rotation-based variations—were instrumental in reducing overfitting and improving the network's resilience to variable real-world image conditions.

From a clinical standpoint, the system's 80%+ accuracy suggests that it can be effectively integrated as a screening support tool rather than a replacement for expert diagnosis. The mHealth integration (discussed in the next section) demonstrates how this AI model can bridge the gap between urban medical expertise and rural accessibility by enabling early detection through smartphones.

Finally, the findings also highlight potential areas for further improvement. While the model performs consistently across most classes, additional efforts such as ensemble methods, transformer-based architectures, or attention mechanisms may further boost sensitivity for difficult-to-detect lesions. Future clinical validation with larger, more diverse patient populations will strengthen the generalizability of these results.

### *E. Summary of the Results and Discussion*

In summary, the proposed EfficientNetB5-based multiclass DFU detection model demonstrates:

- Robust performance across all ulcer categories (overall accuracy: 80.35%).
- Effective handling of class imbalance through augmentation.
- Clinically meaningful four-class categorization for treatment prioritization.
- Superior performance over standard CNNs by up to 5% on key metrics.
- Potential for real-world deployment within a mobile health (mHealth) framework.

These results confirm that AI-driven wound assessment systems can play a crucial role in early ulcer screening and decision support in resource-constrained healthcare settings.

## V. MOBILE APPLICATION WORKFLOW

The mobile application implements a well-defined workflow that begins with user registration and ends with ulcer detection and result monitoring. Each step of the way is facilitated with a user-friendly experience through designated pages and simple navigation.

Figure 13 illustrates the workflow diagram of the mobile application, outlining the sequential processes involved from user input to diabetic foot ulcer classification and result generation.

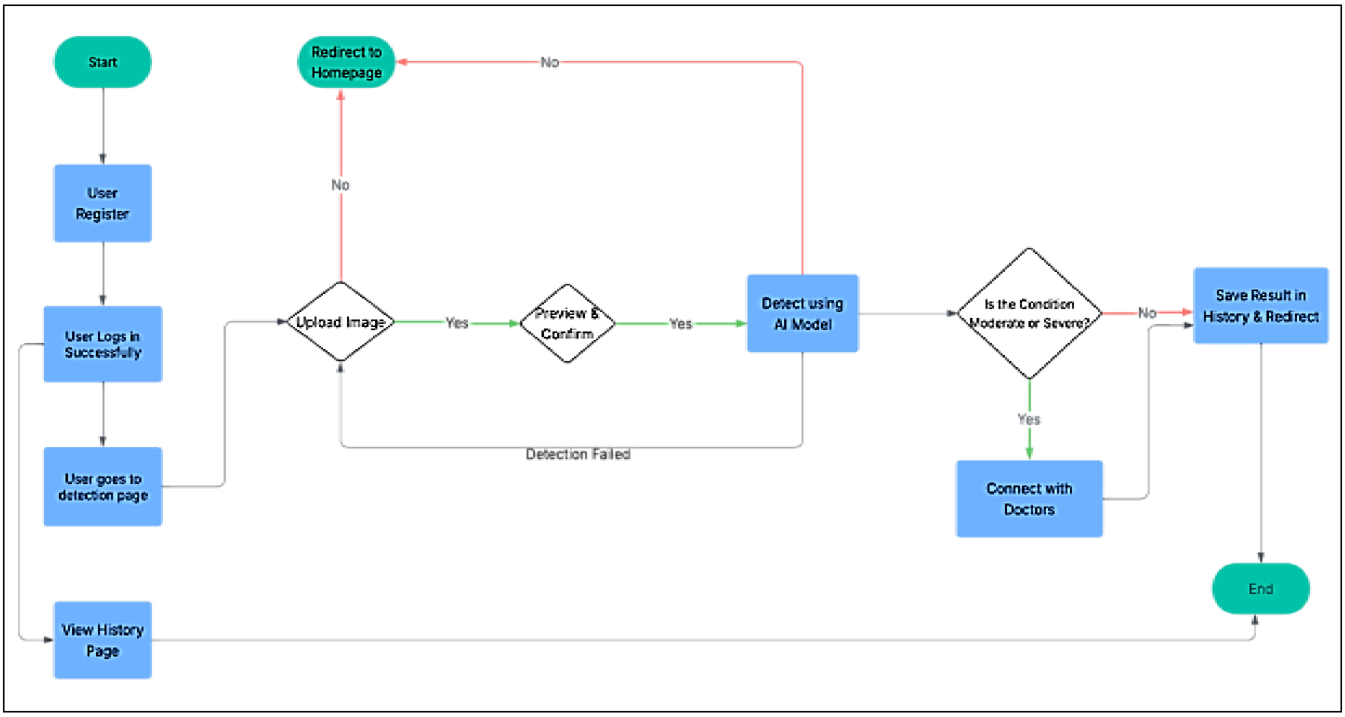


Figure 13: Mobile Application Workflow Diagram

## A. *Application Features*

a) User Registration/Login:

The app user can either go through the registration process or simply log in using his/her credentials (Figure 14).

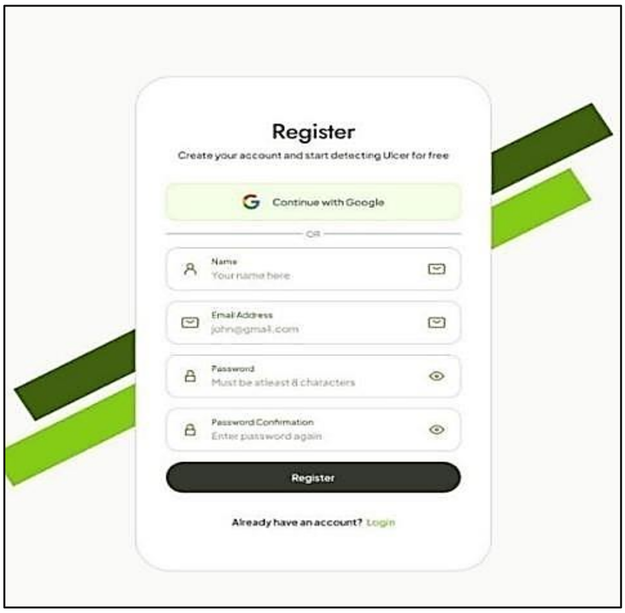


Figure 14: Registration Page UI

b) Redirect to Main Dashboard:

After the user has successfully logged in or registered, he/she will be directed to the main dashboard (home screen). On this screen, he/she will find a new section called "Start Detecting," which is intended for users who wish to begin the process of detecting ulcers, as well as the user's previous records displayed as insights.

c) Dashboard Features:

The dashboard contains a user-friendly interface with several main components: a button to "Detect Ulcer," which starts the detection process, insight cards that reveal the total detections, the number of positive results, and one's current condition, a short list of doctors with their names and emails, and a history preview that summarizes the user's detection activity over time (Figure 15).


Figure 15: Bilingual Feature

d) Ulcer Detection Process:

The moment the user clicks on the "Detect Ulcer" button, he/she is directed to another provided detection page. Here, he/she is able to upload an image for analysis. The AI model investigates the image on the lookout for an ulcer. In case the result indicates a positive detection of an ulcer, the app immediately displays the severity level (mild, moderate, or severe) along with suggested next steps, such as consulting a medical professional or following recommended care instructions. The user can then save the result to their History page for future reference, enabling continuous monitoring and tracking of their condition over time. Additionally, the app may provide visual markers on the uploaded image to highlight the area where the ulcer was detected, helping users better understand the analysis. Figure 16 illustrates the user interface of the detection page, highlighting the image upload functionality and the real-time display of ulcer analysis and severity prediction results.

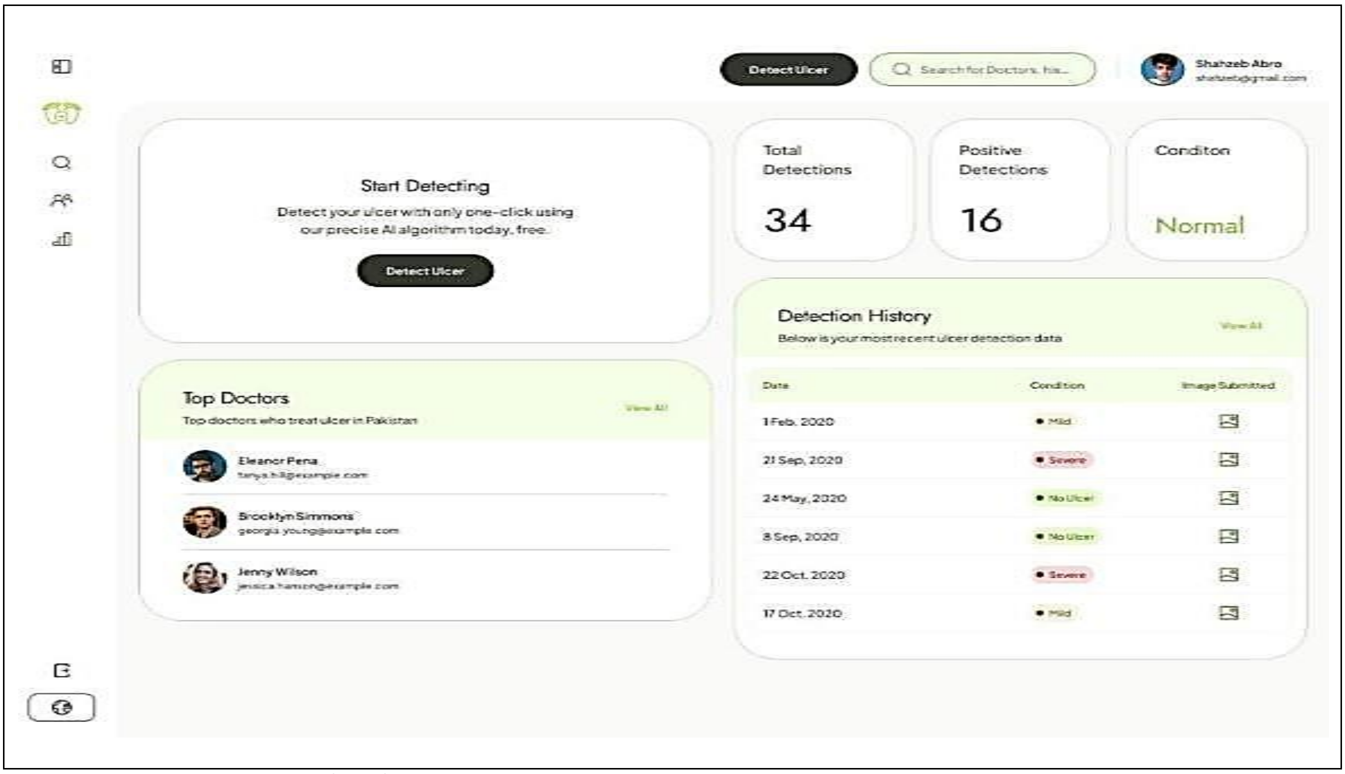


Figure 16: User Interface of the Detection Page, Showing the Image Upload Option and Real-Time Ulcer Analysis Results

e) History Page:

The History page is displayed to the user when they click on the insight cards or select the "History" option in the bottom menu. It is a full record of past ulcer detections that includes the date, result, and severity level so users can keep track of their condition over time.

## *B. Model Integration*

For the smart detection features, the app's AI model runs on a Flask server. The process is very straightforward to ensure that users will receive feedback soon after they upload.

a) Flask API Integration:

The app communicates with a Flask backend that incorporates the AI model. It sends images from the app and receives detection results in return.

b) Real-Time Feedback:

The moment a prediction is made, the app interface displays the results like magic, thereby granting the user instant access to output

c) Device Testing:

The app was tested on both emulators and actual devices to uncover and resolve any issues with design or performance.

## *C. Deployment*

The application went through a process of careful packaging, testing, and publishing in order to be ready for public use.

- APK Generation: A signed APK was generated using React Native CLI, ensuring that it was ready for installation on Android devices. The build process included optimizations to reduce the app size and improve performance.
- Play Store Publishing: A Google Play Developer account was created, and all necessary information, including app description, screenshots, and privacy policy, was provided. The app was then successfully published through the Play Console, making it accessible to users worldwide.

*D. Testing and Debugging*

We conducted thorough testing under various real-world scenarios to confirm that the application performed well and delivered the expected results.

- Unit & Integration Testing: Each component of the application was individually tested, followed by integration testing to ensure that all parts worked seamlessly together.
- Code Optimization: Proguard rules and additional measures were applied to secure the code and make the build more efficient.
- Real Device & Emulator Testing: The app was tested on multiple devices and emulators to identify layout, performance, and compatibility issues.
- Error Logging & Monitoring: Logging mechanisms were implemented to monitor crashes and errors, allowing quick identification and fixing of issues.

*E. Challenges Faced*

There were several practical difficulties encountered by the team during the development of the mobile app, which required thoughtful problem solving and endless improvisation.

a) Cross-Platform-Compatibility:

The task of making the application appear and function the same in a uniform manner on various Android devices was not an easy one all the time. Different sizes of screens along with OS versions now and then created layout issues, thus, we had to put in more time to rectify the design and conduct testing.

b) API Communication and Error Handling:

The process of linking the mobile application to the Node.js backend along with the Flask server was not that smooth as we had expected. There were times during the testing process when the upload of images would fail or the detection results would be returned incorrectly. This led us to enhance the existing error handling and make the app more robust against minor network issues.

c) Authentication Integration:

Our goal was to allow users to choose between the regular email/password method and the simpler Google sign-in method. However, to allow both to operate securely side-by-side meant that we had to meticulously handle session tokens and ensure that they did indeed function properly across various devices.

Figure 17 illustrates the sequence diagram of the authentication process, detailing the interaction flow for both email/password and Google-based sign-in mechanisms.

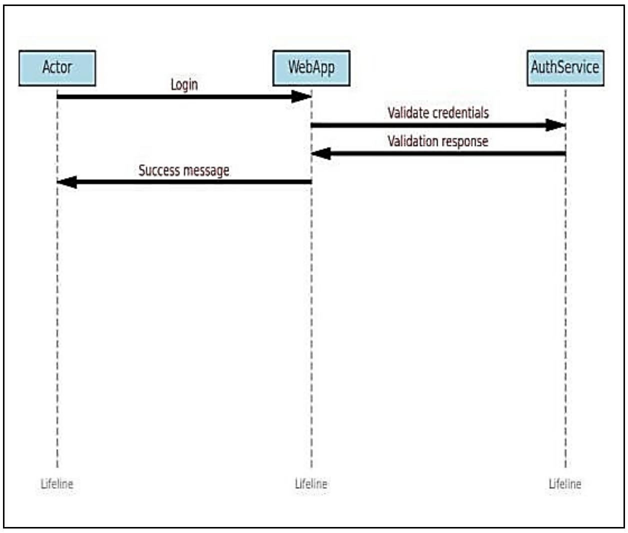


Figure 17: Sequence Diagram of Authentication

d) Model Accuracy and Image Quality:

The detection model's accuracy was largely influenced by the manner in which users captured the pictures of their feet. Lighting that was too poor, angles that were wrong, or unclear pictures now and then resulted in incorrect output. One of the solutions we came up with was to prompt and advise users on how to take better pictures, thus improving both the usability and the accuracy of the detection.

e) Deployment and APK Signing:

Releasing the app was the moment when the creation of a ready-to-use APK was faced with multiple hurdles. The setup of the keystore, the signing of the build, and making sure that everything was safe and operational required a thorough step-by-step process.

*F. Future Work*

Looking forward, we have quite a number of significant improvements we will want to consider for both the mobile and web platforms to make the system more powerful, easier to use, and scalable:

- Offline Detection Capability: Access to the internet is often restricted or unreliable in a lot of rural and remote areas. The mobile app will come with a feature for offline detection as a part of the system to eliminate this hurdle. Therefore, patients will be able to conduct an ulcer screening on their devices without an internet connection being active.
- Real-Time Telemedicine Features: The system will utilize telecommunications in real-time, for instance, chat and video consultations. This will enable quick and easy connection between the patients and the doctors. They can ask for help right away, describe their physical problems, and get tips without going out of their homes.
- Integration with Wearable Health Devices: The system will include smart insoles and similar devices, which are one of the types of wearable that will be connected to it. These insoles measure

both the foot's temperature and pressure. They are always keeping track of the patient's health, and giving out warnings before the formation of ulcers.

- Improved Accessibility and Language Support: Eliminating all barriers and making the system usable for every person is extremely crucial. Future releases will be equipped with extensive local language support for Pakistan, such as Punjabi, Sindhi, Pashto, and Balochi. This will facilitate people of diverse cultures and languages to operate the system with ease.
- Expanding Disease Detection to Other Conditions: The system will not be restricted to only diabetic foot ulcers detection but will also be directed to identify other diseases in poor and at-risk communities. The diseases include leishmaniasis, conjunctivitis, and acute respiratory infections. The main and primary focus will be in regions that are hit by floods or other disasters, which mostly lead to the outbreak of diseases. Timely detection of such diseases can mitigate the health risks.

## VI. CONCLUSION

This section presents a summary of the entire Diabetic Foot Ulcer (DFU) detection project. It also provides a description of the aims, results and major outcomes. The project aimed at employing the techniques of image processing and machine learning to design a system that is able to detect and classify DFUs quickly and accurately.

The system utilized both clinical images and public datasets to not only enhance the existing methods but also provide better care to the patients. In order to tackle challenges like poor image quality and imbalanced data, preprocessing steps were taken. These steps included noise reduction, image resizing, and data augmentation through adding more data. All these activities contributed to making the machine learning models more stable and trustworthy.

One of the important accomplishments of the project was to shift from basic two-class detection (i.e., ulcer or no ulcer) to a multiclass system. This allowed the model to differentiate among no infection, mild, moderate, and severe cases. Consequently, the physicians could receive clearer and more comprehensive outcomes to facilitate their decision-making.

State-of-the-art machine learning models were applied to the system making it more scalable, accessible, and valuable. The findings supported that with good-quality data and powerful algorithms DFU detection could be more precise and productive.

However, the project also experienced certain limitations. The available data sets were not highly diverse or balanced. Moreover, real-time application was also difficult. The project, therefore, recommends, among other things, the collection of more data, the conduct of clinical trials for testing the system, and the improvement of the system for it to work on and be supported by mobile and other low-resource devices.

## VII. FUTURE WORK

The following enhancements and developments are planned to make the system more robust, accessible, and useful for both patients and clinicians:

- Enhance Model Performance through Advanced Techniques:

To enhance model performance and at the same time understand the system's features of the ulcer even better, advanced deep learning methods such as ensemble learning and transformer models are applied. Also, real-time performance on mobile phones would be maintained. The use of CNN optimizers would lead to a higher accuracy of detection as well.

- Implement Offline Functionality:

The focus of subsequent studies may be the enhancement of offline capabilities, through the use of edge technology, and lightweight models that are portable for mobile devices, consequently ensuring operability even in areas with no internet connection. This would completely remove cloud processing and allow the application to perform image analysis for ulcer detection.

- Incorporate Additional Health Parameters:

The app could be a greater asset to patients by diagnosing not only the diabetic conditions but also their related ones like poor circulation and skin condition. With these new functions, the app will be a comprehensive clinician's tool to assist the overall management of diabetic foot problems.

**Acknowledgment**

We would like to express our sincere gratitude to the Department of Computer Science and Information Technology, NED University of Engineering and Technology, Karachi, Pakistan, for their guidance, support, and resources throughout this project. We are also thankful to our mentors, colleagues, and peers for their valuable suggestions and constructive feedback, which greatly enhanced the quality of our work, and we extend our appreciation to our families and friends for their continuous encouragement and support.

**Authors Contributions**

The authors equally contributed.

**Conflict of Interest**

The authors declare no conflict of interest.

**Data Availability Statement**

The testing data is available in this paper.

**Experiment Video Link**

https://youtu.be/f1Wo_KY4hS4

**Funding**

This research received no external funding.

## References


[1] Sarmun, R., Chowdhury, M. E. H., & Rahman, M. F. A. (2024). Diabetic Foot Ulcer Detection: Combining Deep Learning and Image Processing.

[2] Shata, A., & Abdulaziz, N. (2021). Early Detection of Diabetic Foot Ulcer Using CNNs.

[3] Muthuraja, M., Natarajan, S., & Aravindh Raj, N. (2024). Efficient Net-based DFU Detection from Kaggle Dataset.

[4] Giridhar, C., Akhila, B., & Pranay, S. (2024). Detection of Multi-Stage Diabetes Foot Ulcers Using CNNs.

[5] Nguyen, T., Tran, P., & Dinh, M. (2023). Ensemble Approaches for DFU Classification. *IEEE Transactions on Medical Imaging, 42*(3), 678–685.

[6] Gupta, R., & Sharma, S. (2022). Hybrid Models for Accurate DFU Classification. *Journal of Medical Imaging Research, 10*, 112–120.

[7] Zhao, H., Lin, X., & Chen, K. (2024). Socio-Economic Impacts of Untreated DFUs: A Machine Learning Perspective.

[8] Ali, F., Khan, A. M., & Saleem, N. (2023). Contrast Enhancement Techniques For Improving DFU Detection.

[9] Hassan, Z., Javed, Y., & Siddiqui, K. M. (2023). Artifact Removal in DFU Preprocessing.

[10] Khan, A., Ahmed, M., & Rehman, S. (2024). Attention Mechanisms in DFU Detection using CNNs.

[11] Wang, L., Zhang, T., & Yang, J. (2022). Efficient Image Segmentation for Diabetic Foot Ulcers.

[12] Patel, D., & Gupta, S. (2023). DFU Detection in Resource-Constrained Environments.

[13] Brown, K., & Taylor, M. (2021). Evaluation Metrics for Medical Image Analysis Models.

[14] Kumar, N., & Singh, P. (2024). Transfer Learning for DFU Detection: A Comprehensive Study.

[15] Tanaka, Y., & Hirose, S. (2024). Edge Detection Techniques in Medical Imaging.

[16] Almufadi, N., et al. (2024). Classification of Diabetic Foot Ulcers from Images Using Convolutional Neural Networks and Machine Learning. *Diagnostics, 16*, 1807.

[17] Sendilraj, V. (2024). DFUCare: Deep Learning Platform for Diabetic Foot Ulcer Localization and Classification. Frontiers in Endocrinology.

[18] Zhou, G. X., et al. (2025). Construction and Validation of A Deep Learning-Based Model for Diabetic Foot Ulcer Segmentation and Classification. Frontiers In Endocrinology.

[19] Iftikhar, A., Ali, S. M., Alam, M., Musa, S., & Su'ud, M. M. (2022). Analysis of Risk Factors in Global Software Development: A Cross-Continental Study Using Modified Firefly Algorithm. *Computational Intelligence and Neuroscience, 2022*, 1–13.

[20] Iftikhar, A., Alam, M., Ahmed, R., Musa, S., & Su'ud, M. M. (2021). Risk Prediction by using Artificial Neural Network in Global Software Development. *Computational Intelligence and Neuroscience,* 2021.

[21] Japkowicz, S., & Stephen, N. (2011). Evaluation: From Precision, Recall And F-Measure to ROC, Informedness, Markedness and Correlation. *Journal of Machine Learning Technologies, 2*(1), 37–63.

[22] Silva, M. A., Patel, R. K., & Hernandez, J. L. (2025). Systematic Review of Artificial Intelligence and Machine Learning Approaches for Diabetic Foot Ulcer Classification. *Journal of Medical Internet Research, 27*, 1–18.

[23] Alhasson, H. F., Alqahtani, M., & Khan, S. R. (2025). Recent Machine Learning and Deep Learning Frameworks for Diabetic Foot Ulcer Detection and Segmentation: A Comprehensive Review. *Computers in Biology and Medicine, 176*, 108–122.

[24] Weatherall, T. (2024). Machine Learning Models for Diabetic Foot Ulcer Prediction and Risk Assessment: Trends Toward Integrated Clinical Decision Support Systems. *Artificial Intelligence in Medicine, 148*, 102–115.

[25] Verma, G., Singh, R., & Malhotra, P. (2024). Early Diabetic Foot Ulcer Detection Using Thermal Imaging and Deep Learning. *Biomedical Signal Processing and Control, 88*, 105–118.