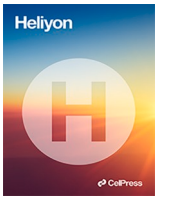
Heliyon 10 (2024) e31228

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Research article

XAI-FusionNet: Diabetic foot ulcer detection based on multi-scale feature fusion with explainable artificial intelligence

Shuvo Biswas [a], Rafid Mostafiz [b,*], Mohammad Shorif Uddin [d], Bikash Kumar Paul [a,c]

[a] Department of Information and Communication Technology, Mawlana Bhashani Science and Technology University, Bangladesh
[b] Institute of Information Technology, Noakhali Science and Technology University, Bangladesh
[c] Department of Software Engineering, Daffodil International University, Bangladesh
[d] Department of Computer Science and Engineering, Jahangirnagar University, Bangladesh

ARTICLE INFO

Keywords:
Diabetic foot ulcer
Multi-scale feature fusion
VGG19
NASNetMobile
DenseNet201
Explainable deep learning

ABSTRACT

Diabetic foot ulcer (DFU) poses a significant threat to individuals affected by diabetes, often leading to limb amputation. Early detection of DFU can greatly improve the chances of survival for diabetic patients. This work introduces FusionNet, a novel multi-scale feature fusion network designed to accurately differentiate DFU skin from healthy skin using multiple pre-trained convolutional neural network (CNN) algorithms. A dataset comprising 6963 skin images (3574 healthy and 3389 ulcer) from various patients was divided into training (6080 images), validation (672 images), and testing (211 images) sets. Initially, three image preprocessing techniques - Gaussian filter, median filter, and motion blur estimation - were applied to eliminate irrelevant, noisy, and blurry data. Subsequently, three pre-trained CNN algorithms -DenseNet201, VGG19, and NASNetMobile - were utilized to extract high-frequency features from the input images. These features were then inputted into a meta-tuner module to predict DFU by selecting the most discriminative features. Statistical tests, including Friedman and analysis of variance (ANOVA), were employed to identify significant differences between FusionNet and other sub-networks. Finally, three eXplainable Artificial Intelligence (XAI) algorithms - SHAP (SHapley Additive exPlanations), LIME (Local Interpretable Model-agnostic Explanations), and Grad-CAM (Gradient-weighted Class Activation Mapping) - were integrated into FusionNet to enhance transparency and explainability. The FusionNet classifier achieved exceptional classification results with 99.05 % accuracy, 98.18 % recall, 100.00 % precision, 99.09 % AUC, and 99.08 % F1 score. We believe that our proposed FusionNet will be a valuable tool in the medical field to distinguish DFU from healthy skin.

1. Introduction

Diabetes Mellitus (DM), generally referred to as Diabetes is a chronic condition characterized by persistent hyperglycemia or high levels of blood sugar. This enduring state leads to crucial life-threatening issues, including kidney failure, cardiovascular diseases, lower limb amputation, and blindness, which is commonly caused by Diabetic Foot Ulcers (DFU) [1]. DFU is a crucial problem of diabetes that is detected based on types of foot injuries. As per the global diabetes report, the year 2014 witnessed a notable surge in the

* Corresponding author.
E-mail addresses: it21620@mbstu.ac.bd (S. Biswas), rafid.iit@nstu.edu.bd (R. Mostafiz), shorifuddin@juniv.edu (M.S. Uddin), bikash.k.paul@ieee.org (B.K. Paul).

https://doi.org/10.1016/j.heliyon.2024.e31228
Received 24 December 2023; Received in revised form 11 May 2024; Accepted 13 May 2024
Available online 14 May 2024

number of individuals grappling with diabetes, reaching 422 million, a substantial increase from the 108 million reported in 1980. In the demographic of adults aged 18 and above, there has been a noticeable escalation in global prevalence, rising from 4.7 % in 1980 to 8.5 % in 2014 [2]. The estimation indicates that by the conclusion of 2035, the global tally of individuals living with diabetes is anticipated to climb to 600 million worldwide [3].

Notably, it is crucial to highlight that approximately 20 % of these individuals will be from economically developed nations, and the remainder will be from economically developing nations where awareness is overlooked and healthcare facilities are limited [4]. A diabetic patient has a 15%–25 % probability of developing DFU in the future, and if treatment is not received, this could lead to lower limb amputation [5]. Over a million diabetic patients lose a portion of their leg annually as a result of improper management and diagnosis of DFU [6]. A patient with diabetes who has a "high-risk" foot necessitates regular medical check-ups, ongoing costly medications, and a healthy lifestyle to prevent the previously mentioned effects. As a result, this imposes a substantial financial strain on patients and their families, particularly in developing nations where the expenses associated with treating this condition can amount to a financial burden equivalent to 5.7 years of their annual income [7]. Doctors suffer from several indecisions in the process of judging DFU, including 1) a lack of confidence in making treatment decisions; 2) a lack of transparency in the diagnosis process; and 3) the possibility of improper treatment plans and possibly negative effects for patients.

Recently, different types of artificial intelligence (AI) techniques, like deep learning (DL) and machine learning (ML), have been used in medicine to create safe, low-cost, and automatic ways to diagnose diseases like cancer, tuberculosis, diabetic foot ulcers, brain tumors, and more [8–12]. In the field of medical image processing, various DL-based segmentation algorithms like thresholding, watersheds, region growth, etc. have been demonstrating a vital role in segmenting anomalous areas from images. These algorithms can identify a specific cancer cell through particular medical data. So, such AI techniques in the medical imaging sector can be a useful tool to create a more reliable DFU framework than conventional diabetic foot ulcer tests. Employing DL frameworks helps circumvent challenges that are time-consuming to address in traditional frameworks. However, it's essential to acknowledge that these frameworks necessitate substantial volumes of well-categorized training data.

To address this challenge, the development of Transfer Learning (TL) has proven crucial. With its ability to effectively overcome the pitfalls of both reinforcement learning and supervised learning, TL is gaining widespread recognition [13]. TL encompasses four learning approaches: unsupervised, inductive, transudative, and negative. Each approach has been shown to be capable of handling the DL challenges [14]. A TL technique was applied to build up the frameworks [15] by obtaining the framework's weight through pre-trained frameworks (i.e., ImageNet [16]). The main architecture of the framework comprises three elements: a prediction classifier, a pre-trained network, and a customized head (motivated from Ref. [17]). The pre-trained networks play a crucial role in extracting high-level DL features and are seamlessly integrated with the classification head and customized network. In numerous fields, TL is indispensable for improving accuracy by leveraging its contextual and discriminative feature extraction capabilities. Examples include sentiment classification [18], medical image classification [19], applications of web scraping [20], and social media [21]. This manuscript uses TL techniques to enhance diagnostic reliability and speed up clinicians' decision-making as a result of the discussion above.

While AI-based methods have demonstrated their advantage over disease diagnosis, many of these methods lack the comprehensive interpretability of models concerning crucial features related to pathological signs in medical data. Consequently, the clinical success of these approaches remains obscure until future experiments are conducted to explain the most important information retrieved by these algorithms. Even with extremely exact experimental outcomes, it is exceedingly unlikely that real-world medical experts will accept a black-box DL model. Furthermore, the AL-based diagnosis methods now in use are not very accurate. Another pitfall of black-box models is their sensitivity to hyperparameter selection, which can pose difficulties in the generalization and optimization of novel data. Additionally, this type of model can face an overfitting issue, and as a result, black-box models can exhibit lower performance on unseen images. While black-box models have demonstrated significant success in diverse applications, a significant drawback is their lack of transparency and interpretability.

Explainable artificial intelligence (XAI) is becoming more and more popular in the medical sector, which aims to exhibit explainability and transparency for black-box classifiers that can solve the above limitations. For this reason, post-hoc approaches have gained much vogue in medical data analysis as a viable solution by presenting black-box models in a comprehensible manner for human understanding. These explanations prove valuable in assisting medical experts in uncovering potential discriminatory biases embedded within black-box models. Notably, local, model-agnostic algorithms that focus on interpreting the outcomes of provided black-box models, like SHapely Adaptive exPlanations (SHAP) [22], Gradient-Weighted Class Activation Mapping (Grad-CAM) [23], and Local Interpretable Model-agnostic Explanations (LIME) [24], are recently most popular among these algorithms. These algorithms generate perturbations of a specific sample on a medical image dataset and observe the effect of these perturbations through the black-box model's outcome to calculate the impact of individual information on a certain prediction. The above XAI algorithms will assist in creating an understandable and trustworthy network, which can contribute to enhancing prediction accuracy and confirm that the network is accountable, robust, and fair.

This paper presents a new XAI-based multi-scale feature fusion network (FusionNet) to construct a novel foot ulcer diagnosis approach from the DFU dataset and three well-known XAI algorithms—LIME, Grad-CAM, and SHAP—to explain the prediction of FusionNet. The gradient explainer library of the SHAP algorithm is employed to explain our frameworks' predictions. Grad-CAM highlights the pertinent regions of a sample by utilizing the gradient of the activation map in the last convolution layer of the CNN classifier. Lastly, the LIME algorithm is applied to explain DL features to identify DFU patients from others. Within this framework, the idea of transfer learning (TL) is applied, employing multiple pre-trained CNN networks assembled to execute the same classification challenge. A meta-tuner module utilizes all pre-trained network predictions and produces predicted outcomes. At first, we trained five well-known pre-trained CNN classifiers with fine-tuned layers called DenseNet201, VGG16, VGG19, NASNetMobile, and MobileNet

and then selected the best three among them to build the FusionNet, which is stacked utilizing a meta-tuner module to find the optimal classification outcome. Then, we conducted the proposed FusionNet and all classifiers with the help of the DFU dataset, which contains 6963 images (3574 healthy and 3389 ulcer classes). At last, to create heatmaps that strongly signify ulcer patches and healthy patches, a novel explainability approach is devised by utilizing three well-known XAI algorithms such as LIME, Grad-CAM, and SHAP. For clinical trials, the FusionNet can learn crucial DL patterns and features from the DFU dataset and provides an affordable and understandable method for identifying diabetic-caused ulcer skin. Our manuscript has made vital improvements in the following domains:

- Introduction of FusionNet, an XAI-based multi-scale feature fusion network, for early detection of diabetic foot ulcers.
- Utilization of transfer learning with multiple pre-trained CNNs addressing the same classification challenge.
- Comprehensive analysis of FusionNet's predictions utilizing XAI, demonstrating an accuracy of 99.05 %.

Section 2 presents the literature review. Section 3 demonstrates the proposed methodology. The five subsections in Section 3 are the DFU dataset, data pre-processing, building FusionNet, fine tuner, and three XAI algorithms (LIME, SHAP, and Grad-CAM). The analysis of the results is described in Section 4 which has 4 parts. These parts are environment settings, evaluation metrics, results analysis, and XAI algorithms result analysis. Finally, we conclude Section 5 and highlight future work.

1.1. Research objectives

The main objectives of this research are as follows:

- Develop an automated, explainable DL-based approach for detecting DFU and analyzing the entire dataset.
- Evaluate transfer learning (TL) on multiple pre-trained CNN models using a meta-tuner module.
- Construct a composite model by integrating various pre-trained CNN models and assess its performance metrics.
- Provide explanations for the results of the composite model using a range of XAI algorithms.

1.2. Research questions

- How can DFU be distinguished from healthy skin using DL-based methods?
- Which pre-trained CNN networks are utilized for classifying DFU skin and healthy skin?
- What XAI algorithms are employed to elucidate the classification results?

In the field of diabetic foot ulcer (DFU) detection, medical specialists require a transparent and explainable DL-based diagnosis method to accurately identify the exact location of the DFU. However, the main problem of the current research is the lack of transparency and explainability of their proposed method. The main aim in this research is to address this crucial problem by attaching various XAI-based algorithms to the proposed DL-based FusionNet. These XAI-based algorithms help the DFU specialist make their decision stronger by providing a reliable and trustworthy DFU diagnosis method.

1.3. Research weaknesses

The proposed approach is limited to binary classification, specifically distinguishing between DFU and healthy skin. When subjected to testing with other types of ulcers (such as venous foot ulcer (VFU), arterial foot ulcer (AFU), and neurotic foot ulcer (NFU)), the model tends to classify them erroneously as either healthy or DFU. Consequently, the applicability of the proposed approach is confined to binary classification scenarios and does not extend to multi-class classification.

1.4. Research strengths

The proposed work exhibits several strengths:

- Extraction of high-dimensional DL features across various scales is achieved through the utilization of multiple pre-trained CNN networks.
- Integration of a meta-tuner module aids in the selection of optimal features from the extracted set, thereby reducing the computational complexity of FusionNet.
- Incorporation of XAI-based algorithms facilitates the elucidation of the network's predictions, offering researchers insights into FusionNet's DFU prediction mechanisms.
- The proposed XAI-based diagnostic and prognostic system holds the potential to assist clinicians in making informed decisions, consequently enhancing diagnostic accuracy and patient care outcomes.

2. Literature review

The constraints outlined underscore the need for developing resilient and intelligent approaches for the automated detection of

ulcers, ensuring swift services for frontline clinicians, and encompassing both novice and seasoned experts in the realm of DFU. As a result, there has been a rise in the development of algorithms for the automated detection of ulcers. Notably, algorithms capable of extracting features from DFU images have gained prominence, offering valuable support to clinicians during the diagnostic phase through the provision of automated screening. In this context, two prevalent machine learning (ML) approaches are frequently employed. Initially, conventional classification methods rely on features derived from DFU images, a process that is often challenging and reliant on domain expertise [25]. In contrast, deep learning (DL) approaches [26], notably convolutional neural networks (CNN), demonstrate the capability to autonomously extract features from given input samples, streamlining the process of disease classification.

A multitude of works have explored the detection and classification of DFU data, aiming to differentiate between abnormal skin and normal skin. Many writers have advocated for image processing and ML methods, focusing on the analysis of diverse features such as morphological textures, hues, and patterns. The efficacy of the proposed methodologies is intricately tied to the chosen algorithms and training strategies. DL has emerged as a prevalent and widely adopted approach for the development of medical image segmentation, classification, and detection tasks [27]. Notably, DL models have demonstrated superior performance compared to traditional approaches using DFU images from diabetic-affected patients.

Manu Goyal et al. [28] introduced a conventional computer vision (CCV)-based efficient and economical approach for classifying normal and ulcerative foot skin (**2017**). They proposed a deep neural network-based classifier named DFUNet to classify healthy and ulcer patches from the DFU dataset. Their proposed DFUNet obtained maximum AUC (96.2 %) by leveraging a cross-validation technique whose outcome exhibited superior performance compared to applying conventional DL and ML methodologies.

Wang et al. [29] utilized a specialized capture box to acquire normal and healthy images from the DFU dataset (**2017**). They applied a two-phase Support Vector Machine (SVM) classifier to precisely identify the ulcer location. The two phases involved (i) segmentation, which retrieved super-pixels, and (ii) feature extraction, focusing on retrieving crucial DL features from the DFU samples. Their proposed SVM-based method attained the highest sensitivity (73.3 %) and specificity (94.6 %).

Alzubaidi et al. [30] proposed a framework utilizing a dataset of 754 feet of images encompassing both normal and ulcer skin (**2020**). They proposed a novel deep CNN-based network, called DFU_QUTNet, based on the concept of traditional CNN networks. They developed their model by increasing the width rather than the depth. This network tackled the gradient propagation issue by dispersing errors across multiple channels.

Doulamis et al. [31] proposed a non-invasive platform based on a photonic-based system for handling diabetic patients (**2021**). This innovative platform utilizes hyperspectral and thermal imaging to monitor the present condition of the ulcer, predicting biomarkers such as deoxyhemoglobin and oxyhemoglobin through imaging techniques. Furthermore, this platform was boosted by integrating signal processing techniques, leveraging DL to improve pixel quality, and reducing noise through the implementation of super-resolution techniques.

Alzubaidi et al. [32] suggested four types of hybrid networks for DFU classification (**2021**). These networks were built based on two layers: the parallel convolutional layer (PCL) and the traditional convolutional layer). Each network comprised six components of the PCL, with the number of branches in the PCL ranging from two to five. The same input samples were fed into all networks to retrieve DL features from the samples. These retrieved DL features were merged by leveraging the PCL. Among all the networks, their proposed four-branch hybrid network reached the highest F1 measure (95.8 %).

Juan et al. [33] introduced a new deep-neural network called DFU_VIRNet, designed for the automated classification of DFU skin (**2023**). Their methodology also emphasized feature maps to discern the likelihood of risky regions to detect ulcers. The proposed scheme was trained and tested with two types of samples—visible and invisible. Notably, DFU_VIRNet exhibited superior performance with the highest AUC (0.99301) and accuracy (0.97750), surpassing recent DFU classification tasks.

Das et al. [34] suggested an innovative stacked parallel (SP) framework named DFU_SPNet, which was built based on SP convolutional layers (**2022**). They used multiple diverse convolution filter sizes in DFU_SPNet to retrieve estimation maps. The DFU_SPNet was evaluated by setting the SGD optimizer and learning rate 1e-2 on the DFU dataset. They achieved maximum test accuracy (96.4 %) on this dataset, which was higher than other existing works.

Das et al. [35] proposed a feature fusion framework to extract low-level handcrafted features using ML and high-level features using DL (**2022**). They used deeper residual blocks as DL extractors. Also, they used various algorithms like artificial neural networks, logistic regression classifiers, gradient boosting, and support vector machines as ML classifiers. Among all these classifiers, LRC provided the highest results with AUC (96.50 %), F1 score (95.37 %), and sensitivity (95.23 %). However, they could improve their results by selecting the most important features among the extracted features.

Kaselimi et al. [36] provided an extensive study of the application of artificial intelligence (AI) in observing DFUs (**2022**). The study underscored the merits of these AI methods while allowing for the challenges associated with their effective implementation for faraway patient care. The analysis centered on optical sensors and imaging techniques utilized for DFU detection, considering both sensor characteristics and patient physiological aspects. Despite recommending various monitoring tactics based on image data sources, the study recognized pitfalls in the widespread application of AI algorithms.

Biswas et al. [37] presented an efficient architecture, DFU_MultiNet, for separating DFU from healthy skin (**2023**). DFU_MultiNet was developed based on the multi-scale transfer learning (MTL) concept. MTL extracted information from the input data using three sub-models at the same time. They used a concatenation layer to combine the extracted features from the sub-models. However, the combination of several sub-models complicated their architecture and slowed the training time. Apart from that, DFU_MultiNet failed to explain the results of the classification.

Thotad et al. [38] presented an innovative DL methodology called EfficientNet to detect DFU at early stages (**2023**). The study implemented the EfficientNet algorithm based on a DFU dataset where the image (ulcer and healthy) size was 844 feet. In this study,

Table 1
A summary table of all the approaches.

References	Paper summary	Performance results
Goyal et al. [28], 2017	The authors proposed an ideal framework, named DFUNet, with a very small number of CNN layers. One potential limitation of this architecture is the lack of insufficiently extracted DL features due to the fewer layers.	Accuracy (0.925), AUC (0.961), Precision (0.945), Specificity (0.911), and F1 score (0.939),
Wang et al. [29], 2017	To locate the lesion zone on ulcer samples with the help of a capture box, a two-stage SVM system was designed. The detection reliability leaned on the photo-capturing conditions.	Sensitivity (73.3 %), and Specificity (94.6 %)
Alzubaidi et al. [30], 2022	A novel approach, named DFU_QUTNet, was designed by increasing the width of the architecture without considering computational complexity.	Precision (0.954), F1 score (0.945), and Recall (0.936)
Alzubaidi et al. [32], 2021	A hybrid classifier integrates standard and multi-branch parallel convolutional layers. The parameter fine-tuning process may improve the result of the proposed approach.	Precision (97.3 %), F1 score (95.8 %), and Recall (94.5 %)
Juan et al. [33], 2023	To exhibit consistent outcomes for the locations with features of the diabetic foot through generating the activation maps, DFU_VIRNet was developed.	Accuracy (0.9775), AUC (0.993), Recall (0.982), F1 score (0.978), and Precision (0.974)
Das et al. [34], 2022	With the help of the heterogeneous filter in the middle layers and the multiple parallel convolution layers in each parallel module, unique features from the input instances were taken out. One potential limitation of this architecture is the lack of model explainability.	Accuracy (0.964), AUC (0.974), Precision (0.926), Recall (0.984), Sensitivity (0.984), and F1 score (0.954)
Das et al. [35], 2022	A feature fusion framework to extract low-level handcrafted features using several ML classifiers and high-level features using DL-based deeper residual blocks. They could improve their results by selecting the principle components among the extracted features.	AUC (96.50 %), F1 score (95.37 %), and Sensitivity (95.23 %)
Shuvo et al. [37], 2023	A novel system, named DFU_MultiNet, combines multiple CNN networks to extract discriminant features in a parallel fashion. The architecture was complex and had a very large number of training parameters due to the combination of multiple networks.	Accuracy (99.06 %), F1 score (99.08 %), and Recall (98.18 %)
Thotad et al. [38], 2023	To build up an efficient system, named EfficientNet, three crucial parameters—width, depth, and resolution—of the CNN classifier to detect ulcers A fine-tuner unit could be added to the proposed EfficientNet to improve the training time.	Precision (99 %), F1 score (98 %), Accuracy (98.97 %), and Recall (98 %)
Das et al. [39], 2023	An effective framework (AESPNet) to detect DFU by combining varying-sized kernel-based parallel convolution layers and a bottleneck attention module. One potential limitation of the AESPNet is the lack of explainability of the detection process.	Accuracy (97.02 %), Sensitivity (98.44 %), Precision (94.02 %), AUC (98.60 %), and F1 scores (96.18 %)
Shuvo et al. [40], 2024	Development of a transfer learning-based system named DFU_XAINet with the XAI method to predict DFU using the pre-trained ResNet50 model. Three XAI algorithms were used to interpret the predicted results. However, their predicted results can be improved by a combined network composed of multiple models because multiple models are able to extract more high-frequency features.	Accuracy (98.75 %), Recall (97.6 %), AUC (98.5 %), F1-measure (98.4 %), and Precision (99.2 %)
Proposed approach	The novelty of the proposed method is to extract more high-frequency features from the input image using a fusion network composed of multiple pre-trained CNN models instead of a single CNN network. A meta-tuner module is used to reduce the computational complexity of the proposed architecture because it selects the optimal features. At last, three XAI algorithms are used to provide the transparency and explainability of the predicted results that help DFU specialists make their decisions stronger.	99.05 % Accuracy, 98.18 % Recall, 100.00 % Precision, 99.09 % AUC, and 99.08 % F1 score

the authors built up a reliable framework by optimizing three important attributes (depth, resolution, and width) of the CNN algorithm to effectively identify normal feet and diabetic feet. Comparatively, their algorithm outperformed contemporary algorithms (AlexNet, VGG19, VGG16, and GoogleNet), achieving exceptional accuracy (98.97 %), F1 score (98 %), precision (98 %), and recall (99 %), respectively.

Das et al. [39] developed a robust CNN-based system (AESPNet) to identify DFU (**2023**). To distinguish between DFU and normal skin, they arranged convolution layers in parallel and used the attention module. The AESPNet consists of two segments, with convolution layers of heterogeneous kernel attached in parallel to form the final structure. The attachment of a bottleneck attention unit (BAU) followed every concatenation action in the scheme. They tested AESPNet on four classic CNN-based schemes (i.e., InceptionV3, DenseNet121, VGG16, and AlexNet), where AESPNet exhibited outstanding results with F1 scores (98 %) and sensitivity/recall (98.44 %). Though BAU improved the architecture's performance, it introduced challenges in explaining the architecture.

Biswas et al. (**2024**) [40] proposed a novel transfer learning (TL)-based system, named DFU_XAINet, for distinguishing ulcer cases from normal cases. They trained and tested five pre-trained CNN models to predict the DFU and also evaluated three XAI algorithms (i.e., LIME, SHAP, and GradCAM) to explain the predicted results. Among these models, ResNet50 exhibited high accuracy with 98.75 %. However, using a combined model in place of a single model may improve the classification results.

However, though the above papers have addressed similar problems, these papers have faced some crucial challenges. Some authors identified DFU using a single CNN model with different or same-sized kernel convolution layers. A single model is capable of capturing inadequate high-frequency features, which may result in misidentifying DFU from healthy skin. Biswas et al. [37] used multiple CNN models to identify DFU, which were able to capture high-frequency features effectively. However, their architecture was

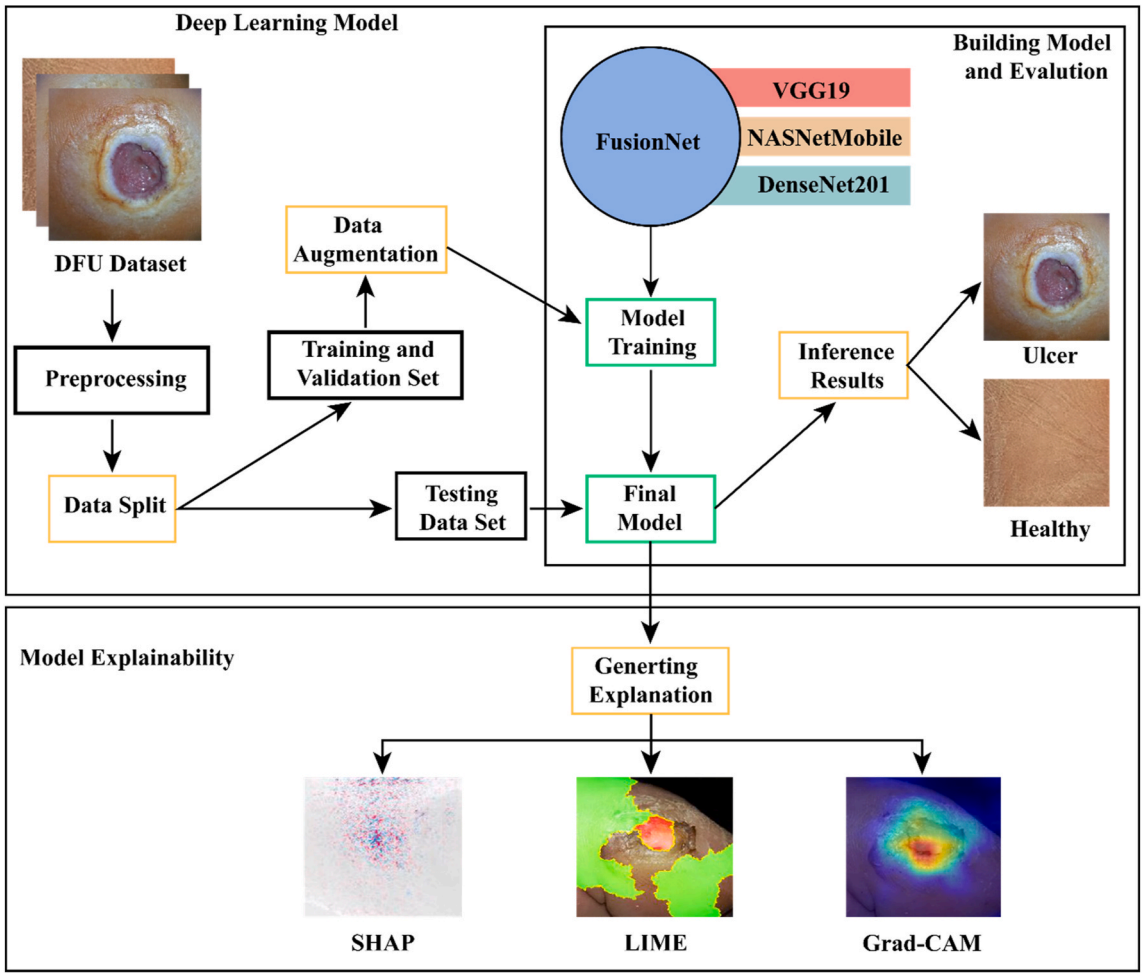


Fig. 1. Workflow of the proposed method.

very complex, with a very high number of trainable parameters due to the combination of multiple CNN networks. Das et al. [35] utilized a parallel CNN scheme with bottleneck attention modules to identify DFU cases. Due to the incorporation of bottleneck attention modules, their architecture was very complex. Apart from that, the major limitation of the above papers is the lack of explainability of their proposed system.

In summery, the novelty of the current article is as follows: 1) used a combined model to retrieve high-frequency and multi-scale features effectively, whereas the other literature used only a single model as a feature extractor; 2) reduced the computational complexity of the combined architecture using a meta-tuner module; 3) used several XAI algorithms to provide the transparency and explainability of the predicted results of the proposed model, where other literature fails to explain the model prediction system. This section succinctly outlines various works in the domain of DFU listed in Table 1, shedding light on the diverse approaches employed by researchers.

3. Proposed methodology

This section covered the suggested method. Fig. 1 shows the workflow of the proposed FusionNet for DFU detection. The FusionNet consists of the following stages: data pre-processing, feature extraction, optimized feature selection, final prediction, and predicted result explanation. Here, several pre-processing methods, including the Gaussian filter, the median filter, and the motion blur kernel method, were used to remove noisy, blurry, and irrelevant data. Then the pre-processed data was fed into three CNN models that had already been trained on the ImageNet database to pull out high-frequency features. The extracted features from each model were provided into a meta-tuner module to select the optimum features. A softmax (SM) activation function is employed to build the final predicted model using the selected optimum features. Finally, three XAI algorithms (LIME, SHAP, and GradCAM) were attached to the final model to provide transparency and explainability for the prediction results in this work. The root factors for choosing feature extraction and selection methods in identifying DFU are: (1) the multi-scale feature extraction method has the benefit of extracting more high-frequency features than the single CNN model; and (2) the fine-tuning-based feature selection method reduces the

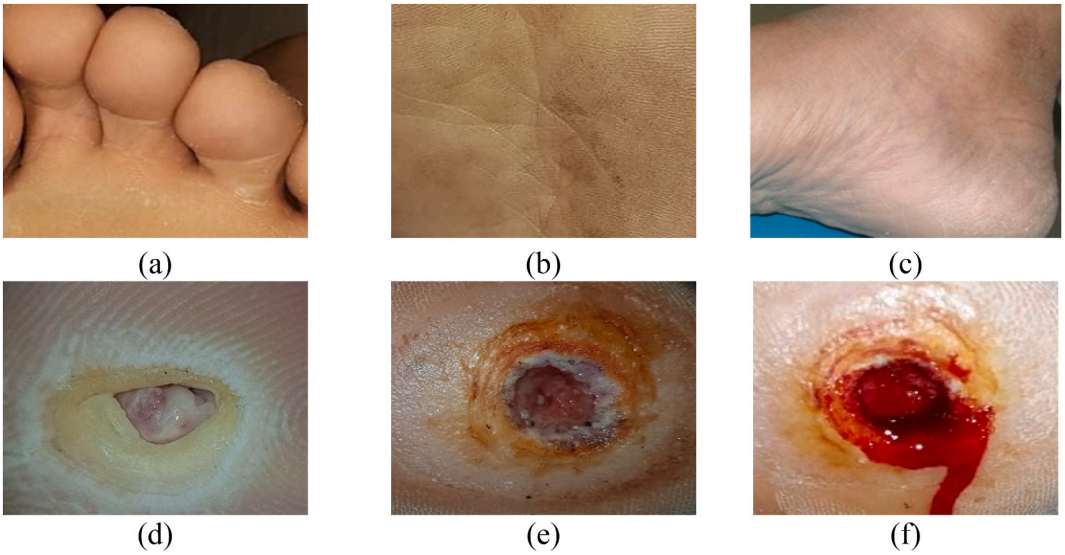


Fig. 2. Some samples of the DFU dataset.

computational complexity of the proposed architecture by selecting the optimal features. Additionally, we conducted two statistical test approaches, namely Friedman and analysis of variance (ANOVA), to find out the root statistical differences between the proposed network and other sub-networks. Fig. 4 displays the overall architecture of the suggested system. The FusionNet and all pre-trained networks are trained and validated on the same DFU dataset and meta-tuner module. Subsequently, the final FusionNet is employed to predict whether unseen images depict ulcers or healthy skin. The ensuing section provided an elaborate description of each stage of the suggested methodology. **Algorithm 1** shows the step-by-step classification procedure of the suggested network.

Algorithm 1. FusionNet for the screening of DFU images.

η = learning rate; β = batch size; δ = mini-batch size; ε = optimizer; λ = epoch;
Input: DFU Training data D^{train} (70 %), Validation data D^{valid} (10 %), and Test data D^{test} (20 %);
Output: ω = weight of the base-CNN networks;
Start:
  Resize each sample into a dimension of 224 × 224;
  Apply the data augmentation technique to enhance the volume of data;
  Retrieve the DL features from the D^{train} using selected pre-trained CNN networks;
  Combine the retrieved DL features through the concatenation layer;
  Set four fine-tuning layers: $CNN^{dropout}$, $CNN^{batch\ normalization}$, CNN^{dense}, $CNN^{softmax}$;
  Initialize the training factors: η, λ, ε, δ, and β;
  Calculate the initial weight ω through training the FusionNet;
    for $\lambda = 1$ to λ **do**
      Select a mini-batch size δ;
      Forward propagation and calculate the loss function;
      Backpropagation and updating the weight ω;
    end for
  Finally, apply XAI algorithms to explain the interpretability and transparency of the predictions of the proposed system;
End

3.1. Dataset

In this experimental setup, the publicly accessible Kaggle "DFU-dataset" [41], originating from the diabetes hospital at Nasiriyah Hospital in southern Iraq [30], played a pivotal role in evaluating the novel FusionNet. It is essential to highlight that the acquisition of informed consent and ethical statements was meticulously undertaken by all pertinent patients and individuals involved in the data selection procedure. The images in this dataset were captured based on the variety of light and viewing conditions by the image labeling experts with the help of a Samsung Galaxy Note 8 and an iPad device. This dataset was composed of four folders. Among these folders, the "patches" folder images were utilized in this experiment because these images were cropped with dimensions of 224 × 224 pixels from the images of the "original images" folder. This "patches" folder contains a total of 1055 images, with 543 of healthy (normal) and 512 of abnormal (ulcer) classes. Fig. 2 (a)-(c) shows some samples of healthy patches and Fig. 2 (d)–(f) shows some samples of ulcer patches from the DFU dataset. The "train_test_split" function is used to partition the dataset into a test set (20 %) and a

Table 2
Dataset distribution before augmentation.

Dataset	Label	Training	Validation	Testing
DFU	Healthy	390	43	110
	Ulcer	370	41	101
	Total	760	84	211

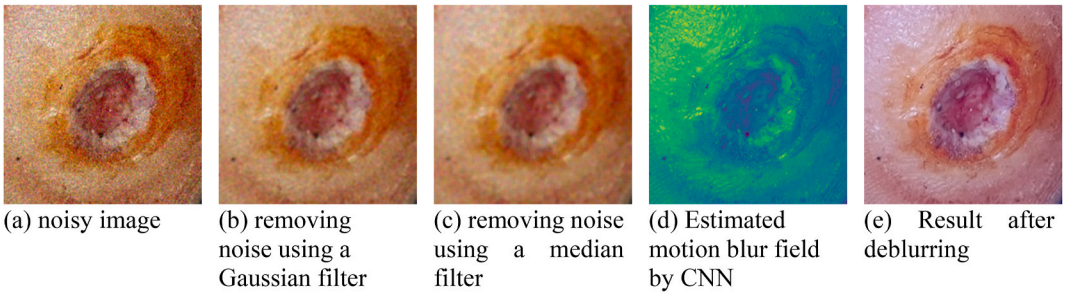


Fig. 3. Noise-removal and deblurring of a sample image.

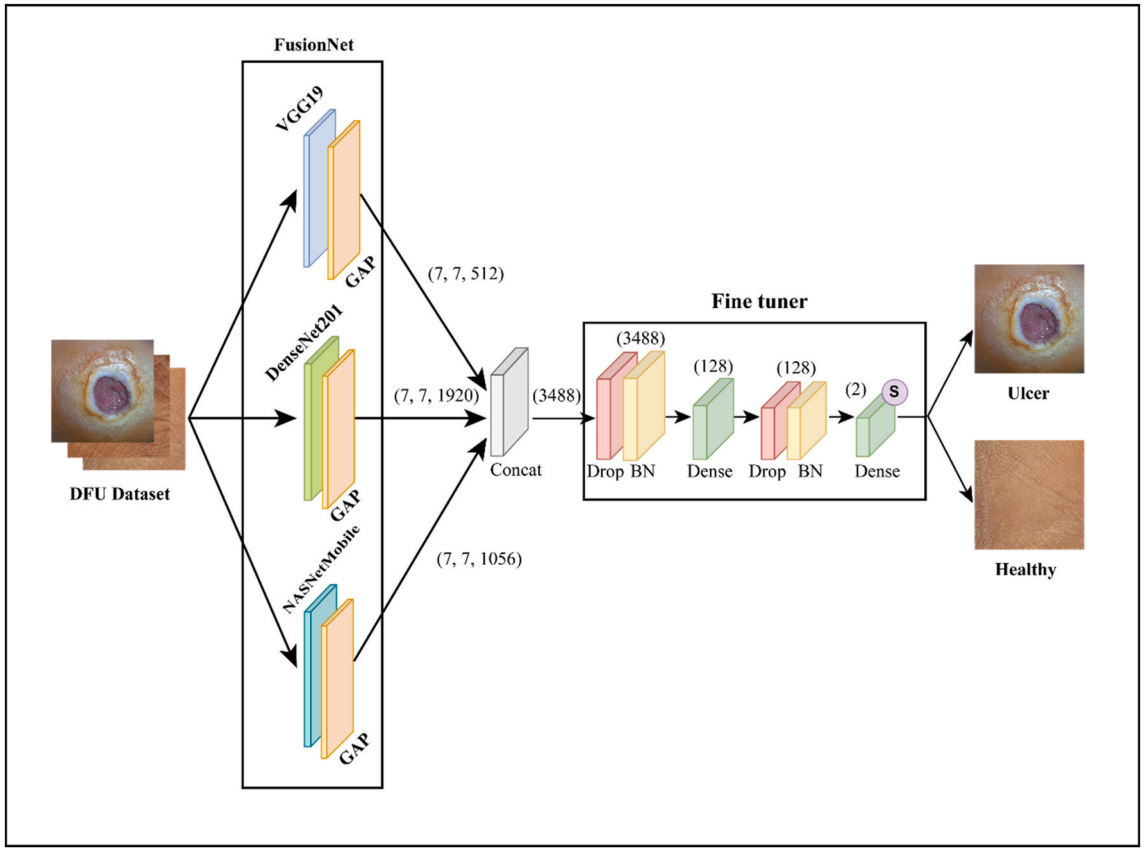


Fig. 4. Architecture of the proposed FusionNet. Here, 'GAP' stands for the global average pooling 2D layer, 'Drop' stands for the dropout layer, 'Concat' stands for the concatenation layer, 'BN' stands for the batch_normalization layer, and 'S' indicates softmax.

train set (80 %). The "train_test_split" is a special type of function in the Python programming language that was imported from the "sklearn.model_selection" package. After that, 10 % of the data is again partitioned for the validation set from the training set using the same splitting function. Lastly, the whole dataset (i.e., 1055 patches) is separated into training (i.e., 760 patches), test (i.e., 211 patches), and validation (i.e., 84 patches) phases at a ratio of 70:20:10 for each phase. Table 2 shows the number of samples in each

Table 3
Data augmentation techniques and parameters.

Strategies	Parameter values
Rotation range	90
Zooming range	2
Width shift range	0.2
Shearing range	0.4
Height shift range	0.2
Vertical flip	True
Horizontal flip	True

Table 4
Dataset distribution after augmentation.

Dataset	Label	Training	Validation	Testing
DFU	Healthy	3120	344	110
	Ulcer	2960	328	101
	Total	6080	672	211

category in the entire DFU dataset based on count.

3.2. Data pre-processing stage

Before giving the dataset to the proposed system, various pre-processing steps are performed. In the DFU dataset, the format of each sample was. jpg with a dimension of 224 × 224 and a configuration of RGB. These samples are transformed into Numpy arrays using the NumPy library in Python to facilitate quicker training and use less memory. Furthermore, we have applied a shuffling operation to train unordered images. However, while acquiring the image from the Kaggle online repository, sometimes the dataset contains noisy images. For this reason, we applied two popular image filtering operations, the Gaussian filter and the median filter, to filter out irrelevant or noisy data from the DFU dataset. Sometimes the quality of the data may be poor due to blurring, which may be caused by motion artifacts. This crucial problem is tackled in the proposed scheme by again employing the motion blur estimation method based on DL methods for motion blur kernel estimation [42]. In the motion blur kernel estimation technique, firstly, this technique estimates the probabilities of motion kernels at the patch level, then fuses the patch-based estimations into a dense field of motion kernels by CNN, and finally, deconvolve the blurry image to estimate the sharp image. Fig. 3 depicts examples of the noiseless and deblurring images from the DFU dataset. Fig. 3 (a) presents the noisy image; Fig. 3 (b)–(c) shows the samples of noiseless images; Fig. 3 (d) shows the intermediate sample after applying the estimated motion blur algorithm; and Fig. 3 (e) shows the final sample after applying the motion blur algorithm.

To provide accurate predictions, CNN networks require a large number of labeled training sets. However, gathering a huge amount of medical imagery is difficult and costly. To tackle this problem, we implemented image augmentation techniques that enhance the volume of samples during training and boost training efficacy by preventing overfitting issues. In image augmentation, we applied five image processing approaches—rotation, flipping (horizontal and vertical), zooming, shearing, and shifting—to generate the required training samples. The values of the non-binary (false or true) operations—shearing, rotation angle, zooming, and shifting—are randomly chosen from pre-defined distributions or ranges. For example, in image zooming, the images are scaled from the range [1 - zoom_range] to [1 + zoom_range]. When rotating an image, the rotation_range parameter permits random rotation from degree 0 to 360. The shearing operation changes the height and width of a sample by picking a floating-point value within a range from 0 to 1. Conversely, the flipping operation flips an image horizontally or vertically with the binary parameter False or True. With these approaches, we increase the training size to 6080 and the validation size to 672. Finally, the total amount of data increased from 1055 to 6963, with the test set containing 211 images. We applied the augmentation technique only to the train set. Augmenting the test set can generate information leakage, biases, and erroneous evaluation scenarios, resulting in incorrect classification results. That's why we use an unaltered and clean test set to develop an unbiased and reliable DFU detection system. Table 3 shows the augmentation parameters. Table 4 shows the number of total samples after applying image augmentation techniques.

3.3. Building FusionNet

This section presents the overall design of the suggested FusionNet. The widespread use of pre-trained neural networks for diverse medical image classification tasks has gained significant traction [43]. In the medical area, where obtaining a large number of categorized images for training deep CNN networks can be difficult, researchers have in recent times used powerful pre-trained CNN networks trained on ImageNet [16] to build a hybrid framework that improved medical image classification tasks. ImageNet is a huge database comprising over 14 M (million) subjects distributed across 20,000 labels, but it conducted 1.2 M subjects representing more than 1000 labels for benchmarking. These labels encompass a range of abstract objects, concepts, animals, scenes, etc.

Inspired by the above advantages, in this paper, we used five CNN-based sub-models—NASNetMobile, VGG16, MobileNet,

Table 5
Shows various concrete parameters of FusionNet.

Parameter	Value
Extracted features	3488
Activation functions	ReLU, SoftMax
Fusion operation	concatenation
Number of fine tuning layers	six

DenseNet201, and VGG19—trained on this database and simulated on the DFU dataset based on the idea of transfer learning. The weights of these sub-models were determined from the ImageNet dataset. From these sub-models, the three most successful ones—DenseNet201, VGG19, and NASNetMobile—are picked for constructing the combined network, named FusionNet. The FusionNet can contribute from both methodological and empirical perspectives. Each sub-model has distinct characteristics and architectural designs. VGG19 is a deeper network than VGG16, which captures fine-grained features, while NASNetMobile provides efficiency and scalability, and DenseNet201 facilitates feature reuse and propagation through dense connections. By combining these diverse architectures in a novel combination, the network can capture a wider range of features and representations from the data. On the other hand, if one sub-model fails to retrieve sensitive and high-level features from the data, the others can retrieve high-level features from the same data through their unique CNN layers, resulting in a reliable prediction system that could be designed. Thus, based on these facilities, we selected these sub-models based on their extracted features and proposed a novel combination using these sub-models. However, from a feature extraction perspective, Densenet201 is able to extract a large amount of features compared to other DL models, whereas VGG16 and VGG19 retrieve a smaller amount of features. On the other hand, NASNetMobile and MobileNet are able to extract a moderate number of DL features. The retrieved characteristics differ from one DL models to another:

1) NASNetMobile (1056) features.
2) VGG16 (512) features.
3) MobileNet (1024) features.
4) Densenet201 (1920) features.
5) VGG19 (512) features.

The primary goal of this network is to acquire and extract DL features through the selected networks. As depicted in Fig. 4, the fusion technique processes the input samples through three functional layers concurrently. Here, each layer indicates a pre-trained network based on DenseNet201, VGG19, and NASNetMobile, respectively. To achieve dimensionality reduction, the output from each functional layer undergoes a global average pooling (GAP) layer. The output from each GAP layer is flattened and merged using a concatenation layer to form a singular feature vector for each input sample. Following flattening, these networks produce an output feature vector with sizes of 1920, 1056, and 512, respectively (see Fig. 4). Thus, the merged feature vector is of size 3488, which is fed into a fine-tuner for tuning the network. The classification task is done through a dense (fully connected) layer. Merging all DL features, our FusionNet is comprised of 42,804,372 trainable parameters that are approximately 3, 3, and 9 times greater than the individual VGG19 (trainable parameters 14,781,890), DenseNet201 (trainable parameters 18,343,170), and NASNetMobile (trainable parameters 4,370,900) networks, respectively. The next subsections detail the fundamental architecture of each pre-trained CNN network that has been used, as well as the fine-tuning procedure.

The Fusion network is designed in a parallel fashion rather than a sequential fashion. A parallel network operates the same DFU dataset independently by using different pre-trained CNN networks. Each CNN network can retrieve discriminant multi-scale information from the input image through its unique CNN architecture. By combining retrieved information from multiple CNN networks, the fusion network can collect both low-level and high-level information, resulting in improved overall classification performance. Thus, the Fusion network can work effectively to detect DFU. Table 5 shows various concrete parameters of FusionNet.

3.3.1. DenseNet

Huang et al. [44] pioneered the development of DenseNet, an exceptional pre-trained image classifier recognized for achieving superior accuracy on the ImageNet dataset. This classifier was built based on a feed-forward architecture akin to ResNet. This classifier incorporates dense connections that facilitate the efficient exchange of crucial information across the network. In this work, we utilized DenseNet201 as our first DL feature selector classifier. This is a complex image classifier because it contains 201 neural layers, each of which is specifically designed to tackle the overfitting challenge. Following training, this classifier encompassed a total of 18,343,170 trainable parameters, which is more than others.

3.3.2. VGGNet

Simonyan et al. [45] pioneered the development of VGGNet, a basic pre-trained image classifier that was trained on the ImageNet database and achieved the top position at the ILSVRC competition. This classifier proved to be a far superior classifier to the AlexNet classifier by demonstrating a remarkable error rate (8.1 %). That's why, in our experiment, we used the VGG19 classifier as the second DL feature detector classifier. This is a simple image classifier because it contains 19 neural layers, including sixteen convolution (CNV) layers and three fully connected (FC) layers. Each CNV layer supports filter sizes from 64 to 512 and a fixed window size of 3×3. The classifier comprised five units, of which the first two contained four CNV layers and the subsequent three were allocated to the

Table 6
Summary of the proposed framework.

Layer (type)	Output Shape	Param #	Connected to
input_1	(224, 224, 3)	0	
densenet201	(7, 7, 1920)	18321984	input_1[0][0]
NASNet	(7, 7, 1056)	4269716	input_1[0][0]
vgg19	(7, 7, 512)	20024384	input_1[0][0]
GlobalAveragePooling2D	(1920)	0	densenet201[0][0]
GlobalAveragePooling2D	(1056)	0	NASNet[0][0]
GlobalAveragePooling2D	(512)	0	vgg19[0][0]
concatenate_4	(3488)	0	GlobalAveragePooling2D[0][0] GlobalAveragePooling2D_1[0][0] GlobalAveragePooling2D_2[0][0] concatenate_4[0][0]
dropout	(3488)	0	
batch-normalization	(3488)	13952	dropout[0][0]
dense	(128)	446592	batch-normalization[0][0]
dropout1	(128)	0	dense[0][0]
batch-normalization1	(128)	512	dropout1[0][0]
dense1	(2)	258	batch-normalization1[0][0]

Total params: 43,077,398.
Non-trainable params: 273,026.
Trainable params: 42,804,372.

remaining twelve. Following each block, a max-pooling (MP) layer supports a window size of 2×2 employed to uniquely identify key DL features from the adjusted activation maps. Each CNV layer operated using a rectified linear unit (ReLU) activation function. Following training, this classifier encompassed a total of 14,781,890 trainable parameters, which is more than the third classifier named NASNetMobile.

3.3.3. NASNetMobile

Zoph et al. [46] pioneered the development of NASNetMobile, another exceptional pre-trained image classifier recognized for achieving a minimum error rate (2.4) on the CIFAR-10 dataset through the ScheduledDropPath regression method. Saxena et al. [47] (**2019**) provided an optimal classifier composed of refined fundamental units that have undergone refinement through reinforcement learning. These units collectively improved the classifier's overall robustness by merging diverse functions like separable convolution, convolution, and pooling. In our study, NASNetMobile is our last DL feature detector classifier. It is a very complex image classifier because it has 769 neural layers. It was built for edge and mobile devices. Following training, this classifier encompassed a total of 4, 370,900 trainable parameters, which is fewer than others.

In summary, the principle of the combined network is to train multiple DL models (i.e., VGG19, DenseNet201, and NASNetMobile) in a parallel fashion at the same time to retrieve optimal features from the training dataset. Then each DL model converts the retrieved features into a one-dimensional feature set using a GlobalAveragePooling (GAP) layer. After that, these feature vectors are fused into a unified feature file using a concatenation operation. Thus, the combined network produces an effective and optimal feature set from the training dataset, which contributes to accurately distinguishing DFU from healthy skin.

Table 7
Summary comparison of the characteristics of the three XAI methods.

XAI	Method Characteristics	Description of DFU Image	Spatial Resolution
LIME	●An algorithm for approximating the prediction output of a black box classifier with another classifier.	●Multiple skin wounds in the sample can be successfully visualized.	The spatial resolution of a sample is altered by the number of extracted superpixels, permitting higher flexibility in the DL feature experiment.
	●Approximate models are not always accurate.	●Focuses on which superpixel area of the sample is most crucial for prediction.	
SHAP	●An algorithm for retrieving sample partially obstructive and quantifying the impact of an area leveraging a black box classifier.	●Multiple skin wounds in the sample can be identified (visualized).	The spatial resolution of a sample is altered by the stride size and kernel size, permitting higher flexibility in the DL feature experiment.
	●The impact of a combination DL activation map cannot be exhibited.	●Focuses on which area of the sample is crucial for prediction.	
Grad-CAM	●An algorithm to build up the black box classifier itself has proof for judgment.	●Targets on an important portion of large-scale wounds are why cannot detect (visualize) multiple skin wounds in the sample.	Limited amount of spatial resolution in the final convolution layer which is 7×7 (in the case of VGG16).
	●Explain the areas that impact the final probability score.	●Focuses on the important pixels that influence updating the final classification decision.	

3.4. Fine tuner module

In this section, the significance of the meta-tuner module is elucidated. The meta-tuner module is needed to adjust the weights of the model for the final classification task. The concatenated features from the concatenation layer are fed into the meta-tuner module. The meta-tuner module consists of six CNN layers with a softmax (SM) activation function. These layers are two dropout layers, two batch normalization (BN), and two dense. Each layer plays an effective role in enhancing the overall performance of the FusionNet. The key tasks of each layer are detailed below.

The incorporation of the BN (batch normalization) [48] layer is highly important to improve the performance of FusionNet. The core contribution of this BN layer is to normalize and resize the images, which represents our framework as a powerful tool for DFU diagnosis. In DL, overfitting is a major problem, which arises when the DL algorithm is excessively trained on the training data and adversely affects the test set [49]. To tackle this condition, we implement 2 dropout layers where, during model training, the first will reject a 40 percent sample and the second will reject a 20 percent sample. In addition, this kind of operation helps to greatly reduce the training time.

The dense layer establishes connections between every neuron in one layer and every neuron in the subsequent layer, creating a fully connected (FC) neural network. That's why it is also called the FC layer, which converts input samples into output predictions. The core contributions of the FC layer are to handle input images, predict class probabilities, and determine the outcomes. In this experiment, we implemented 2 FC layers with two activation functions: one is ReLU [50] and the other is softmax (SM). The SM discerns the most pertinent DL information to classify the abnormal or normal class, produces result values ranging from 0 to 1, and activates the neurons accordingly. The formula for this SM function is defined in equation (1):

$$\text{Softmax}\left(w_p\right) = \frac{e^{w_p}}{\sum_{m=1}^{n} e^{w_m}} \tag{1}$$

Where w_p indicates the p^{th} element of the input vector, e^{w_p} represents the standard exponential function for input vector w_p, n is the number of classes, e^{w_m} represents the standard exponential function for output vector w_m.

Table 6 presents the summary of the proposed framework. This table is retrieved while simulating the FusionNet for classification tasks. So this architecture has two neurons inside the last FC layer.

3.5. Explainable artificial intelligence

In this research, the incorporation of eXplainable Artificial Intelligence (XAI) played a crucial role in elucidating the decision-making system of the FusionNet. The experiment employed three prominent XAI methods, namely SHAP, Grad-CAM, and LIME, each briefly outlined in the subsequent sub-sections, to facilitate visual analysis.

3.5.1. SHaply additive exPlanations (SHAP)

We have also employed another XAI technique, namely SHaply Additive exPlanations (SHAP) [**22**] (**2017**), to elucidate the decisions made by the black box FusionNet. As a post-hoc XAI technique, SHAP offers an explanation based on DL feature relevance. This paper has employed gradient-based interpretations to describe the influence of the intermediate layers of the VGG16 classifier on the outcomes. The gradient interpreter in SHAP leverages the anticipated gradient approach to determine the overall gradients along one or more channels between two suitable dataset. SHAP produces the Shapley value of DL characteristics by setting the marginal contribution of DL characteristics ϕ_i. The formula for marginal contribution is given in equation (2):

$$\text{Marginal Contribution}, \varphi_i^k = \widehat{f}\left(z_{+i}^k\right) - \widehat{f}\left(z_{-i}^k\right) \tag{2}$$

Here, ϕ_i represents the marginal contribution of i^{th} feature, $\widehat{f}\ (z_{+i}^k)$ is the contribution features with i and $\widehat{f}\ (z_{-i}^k)$ is the contribution features without i.

The formula for Shapley value which is the average total combinations $(\phi_i\ (z))$ for a given sample z is provided in equation (3):

$$\phi_i(z) = \frac{1}{M}\sum\nolimits_{i=1}^{M} \varphi_i^k \tag{3}$$

Here $\phi_i\ (z)$ indicates average marginal contribution for sample z, ϕ_i represents the marginal contribution of i^{th} feature and M is the number of input features.

This experiment utilizes the SHAP library of [22] for elucidating the outcomes of the deep CNN classifier through a DL gradient explainer. This library serves as a powerful tool for interpreting in the context of DL models. Specifically, the gradient explainer within the SHAP was utilized to provide insights into the workings of the DL models under examination.

3.5.2. Gradient-Weighted Class Activation Mapping (Grad-CAM)

In DL, each CNN model includes a feature extraction branch and a classification branch. The classification branch holds an FC layer, and the retrieved DL information is transformed into a probability value for each label in the SM (softmax) layer. The final classified outcome of the system is the label with the maximum probability value. Grad-CAM [23], (**2017**) the last XAI algorithm employed in

Table 8
Experimental settings of the FusionNet.

Resources	Details
RAM	64 GB
CPU	Intel Core i5-12600K @ 3700 MHz
GPU	Tesla K80
Platform	Google Colab

Table 9
Training parameters with value for the proposed system.

Parameter	Value
Optimizer	adam
Metrics	accuracy
Loss Function	binary_crossentropy
Learning Rate	0.0001
Epochs	50
Batch Size	32

this study, serves as a class-discriminative localization algorithm that can produce visual interpretations without necessitating structural retraining or modifications. It achieves this by localizing relevant sample regions and utilizing the gradient information of the DL activation map from the final convolutional layer to highlight portions of the sample with the most significant impact on the probability value for the outcome of the prediction. Regions with a larger gradient are indicative of areas exerting a substantial influence on the prediction outcomes. The Grad-CAM output manifests as a heatmap visualization corresponding to a specific class label. This heatmap is instrumental in visually confirming the areas of interest within the image, as elaborated in the experimental analysis section.

3.5.3. Local Interpretable Model-agnostic explanations (LIME)

Ribeiro et al. [24] (**2016**) presented the LIME XAI algorithm, which aims to provide complete justifications for predictions produced by a black box system. The fundamental concept of LIME involves locally estimating the functionality of the black box system employing a clear, transparent glass box system, enhancing interpretability. The LIME algorithm produces perturbations by carefully deactivating and activating specific superpixels across a sample. This approach seeks to explain the results in a human-readable manner and to ascertain the relevance of the persistent superpixels in the predicted samples. LIME enhances model interpretability, promotes transparency, and instills confidence in DL systems by elucidating how the input characteristics of a black-box system influence its predictions. The initial phase of employing the LIME algorithm on an original sample is to break it down into superpixels. These superpixels determine the granularity of the region segmentation. A connected group of pixels that share the same location and color is called a superpixel. The segmentation produced by this process is more thorough and finer, making it possible to identify the regions that are crucial for accurately predicting the outcome.

The XAI is employed to offer local justifications. The local justification is a tactic provided for every prediction separately. This includes two key points: **1**) evaluating the effect of how each input affects the output, and **2**) articulating this effect in a human-understandable manner. While all three XAI algorithms mentioned encompass these fundamental points, their approaches and ensuing results differ. For instance, because the SHAP algorithm measures shapely values across the whole region of the input image, the significance of the combination of DL characteristics between areas is unclear. In LIME, the approximation results of a basic classifier are not always accurate. Grad-CAM does not always offer proof of ulcer normality; it just indicates whether areas of the sample had an impact on the final prediction's likelihood score. As a result, we believe it would be beneficial to provide prediction explanations that combine all three XAI algorithms. A brief comparison of these algorithms is presented in Table 7.

4. Performance evaluation

To show the success of FusionNet, we performed a comprehensive evaluation comparing the performance of all networks. The system configuration, evaluation metrics, outcomes analysis from qualitative and quantitative perspectives, and a conclusion will be covered next.

4.1. System configuration

In this study, instances in the same DFU data with the same train, test, and validation ratios serve as the basis for evaluating the proposed system and individual pre-trained classifiers. The proposed system was facilitated through the utilization of Keras [51], establishing the connection between Python [52] and the NN (neural network). The computational framework is shown in Table 8.

We utilize the Adam [53] optimizer with a learning rate (0.0001) for simulating our proposed system. After we categorized ulcer skin and normal skin, we utilized the binary cross-entropy as a loss function, which made the categorization task quicker. Again, we

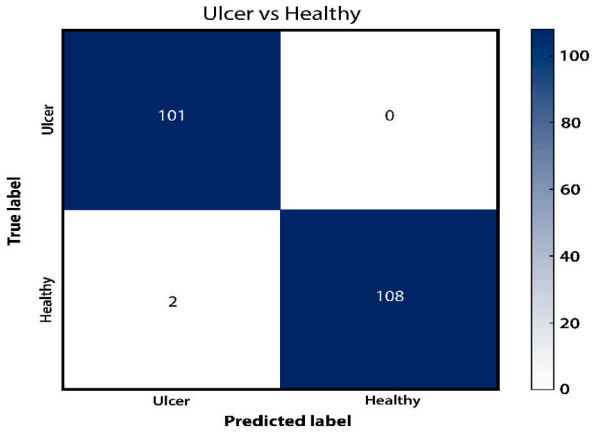


Fig. 5. Confusion matrix of the FusionNet.

Table 10
Evaluation of explainable deep learning models.

Model	Accuracy	Recall	Precision	F1 score	Error Rate	AUC
VGG16	0.981	0.964	**1.00**	0.981	0.019	0.982
VGG19	0.867	0.745	**1.00**	0.854	0.133	0.873
NASNetMobile	0.773	0.564	**1.00**	0.721	0.227	0.782
DenseNet201	0.976	0.955	**1.00**	0.977	0.024	0.977
MobileNet	0.867	0.845	0.894	0.869	0.133	0.868
FusionNet	**0.991**	**0.982**	**1.00**	**0.991**	**0.009**	**0.991**

leverage ReduceLROnPlateau (Reduce on Loss Plateau Decay) and ModelCheckpoint callbacks from the Keras library. ModelCheckpoint observes the evaluation metrics and consistently updates the network to observe the criteria such as validation loss, validation accuracy, training loss, and training accuracy. ReduceLROnPlateau reduces the learning rate if there is no progress in validation loss over a certain period of epochs. A reduced learning rate in deep learning leads to a slower training pace for the CNN model, resulting in minimal adjustments to the classifier weights. To tackle these challenges, we set batch size = 32 and epochs = 50 for exhibiting a successful model. Finally, we have evaluated the evaluation metrics—recall, AUC, accuracy, F1 score, error rate, and precision—on the test set. Table 9 shows the training parameters with values utilized for simulating the CNN classifier.

4.2. Evaluation metrics

A crucial step in building a strong DL system is system assessment. In this study, diverse evaluation parameters, like the AUC-ROC curve, and the CM (confusion matrix), are used to judge the quality of the system. Various performance measurement parameters—F1 score, recall, precision, and accuracy—can be computed using the CM. These parameters were derived from four values: false-positive (FP), false-negative (FN), true-positive (TP), and true-negative (TN). In CM, the positive label indicates ulcer instances, while the negative label indicates healthy instances. The true label indicates the proper identification, while the false label indicates the wrong identification. TP signifies the accurate recognition of ulcer instances, whereas TN represents the precise identification of healthy instances. Conversely, FP denotes the erroneous classification of ulcer instances, while FN refers to the misclassification of healthy instances. These evaluation metrics contribute to a thorough evaluation of how effectively our framework processed the input data. The formula for these evaluation matrices is given in equations (4)–(8)).

Accuracy (ACC) is the percentage of successfully identified instances. It is provided in equation (4):

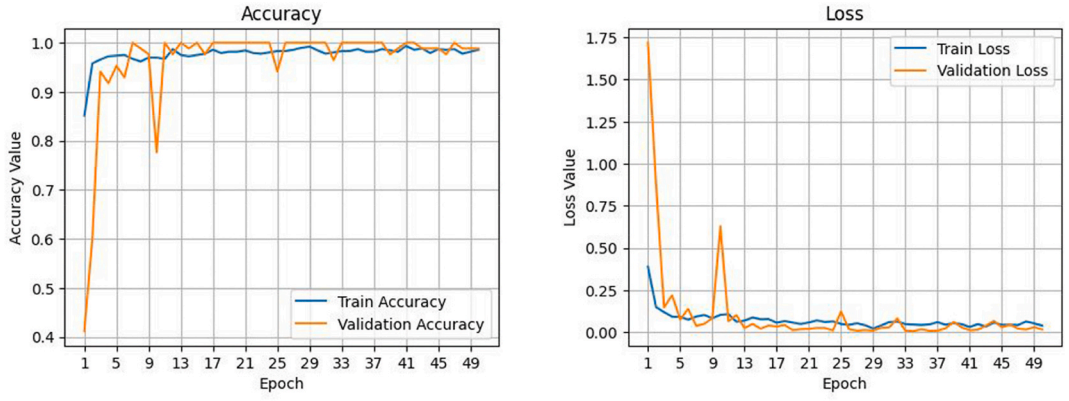
$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

Recall (REC) refers to successfully identifying TP instances by evaluating the proportion of total positive instances. Recall is given by equation (5):

$$Recall = \frac{TP}{TP + FN} \tag{5}$$

Precision (PRE) means how successfully our FusionNet predicts positive instances. Precision is computed with equation (6):

$$Precision = \frac{TP}{TP + FP} \tag{6}$$



(a) Accuracy curve **(b)** Loss curve

Fig. 6. Accuracy (left side) and loss (right side) curves of the FusionNet.

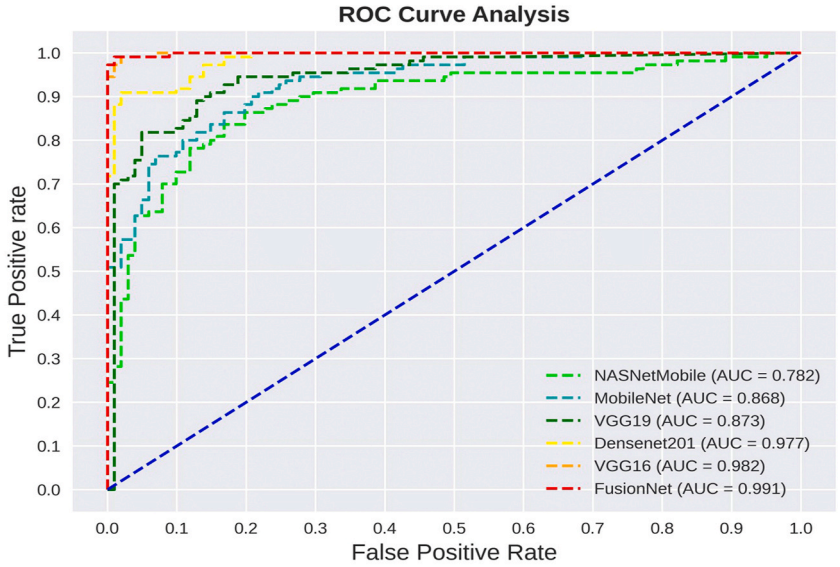


Fig. 7. ROC curve of the experimental models.

F1 score (FS) is the harmonic mean of PRE and REC scores. It is defined in equation (7):

$$F1\ score = 2 * \frac{PRE * REC}{PRE + REC} \tag{7}$$

Error rate refers to the proportion of misclassified instances about the total number of samples. Error rate is calculated by equation (8):

$$Error\ rate = \frac{FP + FN}{TP + TN + FP + FN} \tag{8}$$

Fig. 5 shows the CM (confusion matrix) of our proposed FusionNet. From CM, we show that the FusionNet correctly identifies 108 healthy samples and 101 ulcer samples. A closer look at Fig. 5 shows that the network incorrectly identifies only two healthy samples. One noteworthy benefit of the network is that it performs flawlessly in misclassifying no ulcer cases in the dataset. FusionNet's robustness is further improved by independent assessments of each model on the same DFU dataset.

4.3. Results analysis

Table 10 presents the simulation results of all experimental models. It is observed that through attaining superior accuracy (99.05 %), AUC (99.09 %), F1 score (99.08 %), recall (98.18 %), and precision (100 %), the FusionNet consistently outperforms other

Table 11
Comparison table with existing works.

References	Models	Complexity	Dataset size	Accuracy (%)	Recall (%)
Thotad et al. [38]	EfficientNet	237-layer CNN	844	98.97	98
K. Das et al. [34]	DFU_SPNet	42-layer CNN	1679	96.4	**98.4**
Goyal et al. [28]	DFUNet	14-layer CNN	1679	92.5	–
Wang et al. [29]	SVM	SVM	100	–	73.3
Biswas et al. [37]	DFU_MultiNet	VGG19 + Densenet201 + NasNetMobile + 6-layer CNN	1055	**99.1**	98.2
Alzubaidi et al. [30]	DFU_QUTNet	SVM, 58-layer CNN	1609	–	93.6
Juan et al. [33]	DFU_VIRnet	Xception (71 layers) + 29-layer CNN	2400	97.8	98.2
Das et al. [39]	AESPNet	35-layer CNN + 2 Bottleneck Attention Modules	1679	97.02	**98.44**
Alzubaidi et al. [32]	Hybrid CNN	100-layer CNN	1609	–	94.5
Biswas et al. [40]	DFU_XAINet	ResNet50 (50 layers) + 6-layer CNN	3200	98.75	97.6
Proposed Work	**FusionNet**	DenseNet201 + VGG19 + NASNetMobile + 6-layer CNN	6963	**99.1**	98.2

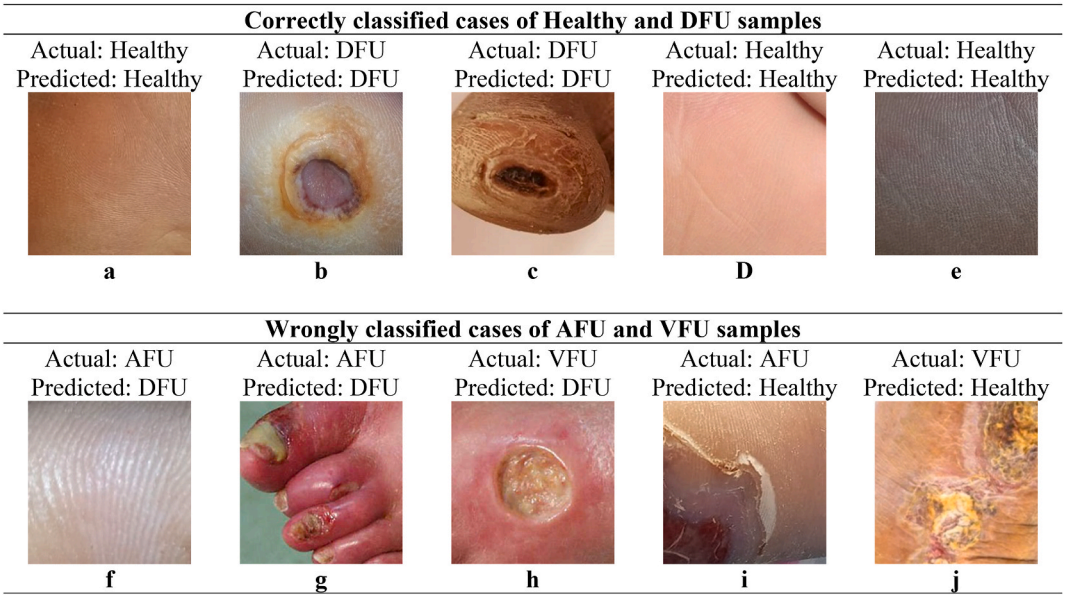


Fig. 8. Some predicted samples using the proposed method.

traditional CNN classifiers. It clearly shows how the FusionNet architecture is superior to other traditional CNN classifiers for categorizing DFU samples. The accuracy outcome suggests that, out of all the cases with any kind of diabetic symptom, the FusionNet can reliably classify ulcer cases with an accuracy of 99.05 %. Upon careful inspection of the comparison table, it is evident that among the five CNN classifiers, DenseNet201 and VGG16 exhibited high outcomes with accuracy scores of 0.976 and 0.981, respectively.

We also track our proposed system's learning curves. Fig. 6 demonstrates how our approach exhibits a moderate learning rate during training and a somewhat consistent decline in validation losses. The simulation results of our approach, extracted from the training phase, are shown in Fig. 6. A closer look at Fig. 6 shows that our suggested approach obtained 97.65 % validation accuracy, 98.68 % training accuracy, 10.03 % validation loss, and 6.17 % training loss, respectively, after the 12th epoch. Furthermore, Fig. 6(a) also assures that the overfitting issue was not noticed throughout the training procedure. But Fig. 6(b) assures that the accuracy-loss curve demonstrated a quick reduction in the loss score. Conversely, minor oscillations were observed when employing a minimal batch size.

To comprehend the class reparability of the suggested approach, we utilize the ROC (receiver operating characteristic) curve as illustrated in Fig. 7. In this curve, TPR (true positive rate) is juxtaposed against FPR (false positive rate) across diverse threshold values derived from the probability scores of DL algorithms. TPR signifies the likelihood of correctly identifying ulcer samples as ulcers; on the other hand, FPR signifies the probability of incorrectly identifying normal samples as normal. The ROC graph effectively demonstrates the strength of the FusionNet, with an AUC (as indicated by the red line in Fig. 7) of 0.991. The AUC scores of five pre-trained classifiers are 0.982, 0.873, 0.868, 0.782, and 0.977 for VGG16, VGG19, MobileNet, NASNetMobile, and, DenseNet201, respectively. Among these pre-trained classifiers, VGG16 exhibits the best discriminating power for diabetic complications. This proved that the ROC curve for VGG16 is somewhat higher compared to the other pre-trained classifiers.

The performance of the suggested FusionNet is benchmarked against existing work in Table 11. We evaluate the proposed method on a large dataset (6963 sample images) to make the evaluation stronger. The outcomes from Table 11 reveal that the suggested system

Table 12
ANOVA test reports for proposed method.

Types	Sum of Squares	Degrees of Freedom	F_statistic	p_value
treatments	74.317519	1	2.641986	0.104389
Residual	28073.153481	998	-	-

Table 13
Report of Friedman test results.

Test	Value
Friedman Test Statistic	9
p_value	0.02929
Null Hypothesis (NH)	Reject

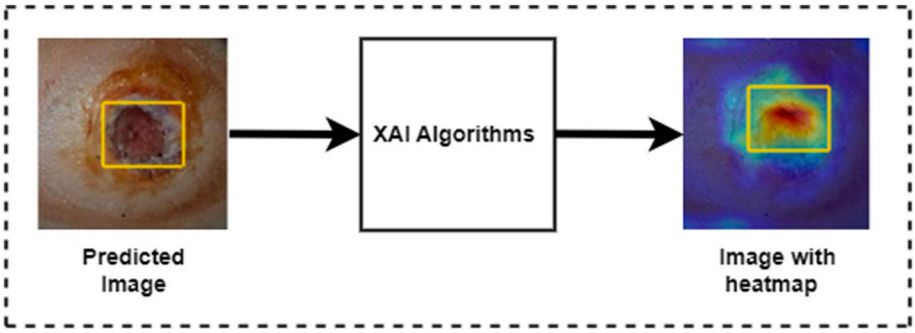


Fig. 9. An example shown how XAI algorithms explain the results of classification.

proved superior regarding all assessment metrics, except recall score. The reason behind this lower score is the somewhat higher false negative (FN) value. Upon careful inspection of the comparison table, it can be seen that FusionNet has extremely high accuracy (99.1 %) and precision (100 %) values, which makes it a substantially more potent prediction classifier for DFU ulcer cases. A classifier that obtains a high degree of precision and accuracy is typically regarded as a powerful DFU classifier.

However, the suggested system has two pitfalls: first, using this system, it is possible to assess if the sample is healthy skin or ulcer skin only. However, the system cannot allow real-time observation of pain intensity and complexity levels. Second, even if the suggested system performs well on this experimental dataset, it may yield reliable outcomes on a larger dataset through learning unique DL features from diverse samples.

Though the proposed method can distinguish the DFU from the healthy foot, this system cannot distinguish other ulcers (i.e., venous foot ulcer (VFU), arterial foot ulcer (AFU), and neurotic foot ulcer (NFU)) from the healthy foot. Fig. 8 shows some predicted samples using the proposed method. From Fig. 8, we can see that the proposed method accurately classifies healthy and DFU cases but fails to distinguish other ulcers. So the proposed method is applicable only for binary classification. Fig. 8 (a)–(e) shows correctly classified cases of healthy and DFU samples. Fig. 8 (f)–(j) shows wrongly classified cases of AFU and VFU samples.

4.4. Computational complexity analysis

In [37], a CNN-based MultiNet model was used. This model consists of three pre-trained CNN models (VGG19, Densenet201, and NASNetMobile) and a 6-layer CNN at the end of the MultiNet. In the study conducted in Ref. [38], a pre-trained CNN model named EfficientNet was used, consisting of 237 CNN layers. No additional layers were used at the end of the EfficientNet. In the study conducted in Ref. [33], a pre-trained CNN model named Xception (71 layers) and more than 29 CNN layers were used. An SVM classifier was used after the CNN model. In Ref. [34], a total of 42-layer CNN model was created, consisting of 1 input layer, 4 convolution layers (each with 1 ReLU activation layer), 3 stacked parallel units (each consisting of convolution and a ReLU activation layer), 4 transition layers (each consisting of a BatchNormalization and a LeakyReLU activation layer), 4 concatenation layers, 4 MaxPooling layers, 1 dropout layer, and 3 fully connected layers. In Ref. [32], a total of 100-layer CNN model was created, consisting of 1 input layer, 3 feature extraction parts, and 1 output classifier. In feature extraction parts, part 1 contained 3 convolution layers, each followed by one BatchNormalization layer and one ReLU layer; part 2 contained six parallel convolution blocks; and part 3 contained 1 average-pooling layer, 1 dropout layer, and 2 FC layers. In Ref. [30], a total of 58-layer CNN model was created, consisting of 1 input layer, 17 convolution layers, each followed by one batch normalization layer and one ReLU layer, 4 concatenation layers, 1 average-pooling layer, 1 dropout layer, and 2 fully connected layers. In Ref. [39], a CNN-based AESPNet architecture was designed, consisting of 35 CNN layers and 2 Bottleneck Attention Modules. In Ref. [28], a 14-layer CNN model was created, consisting of 1 input

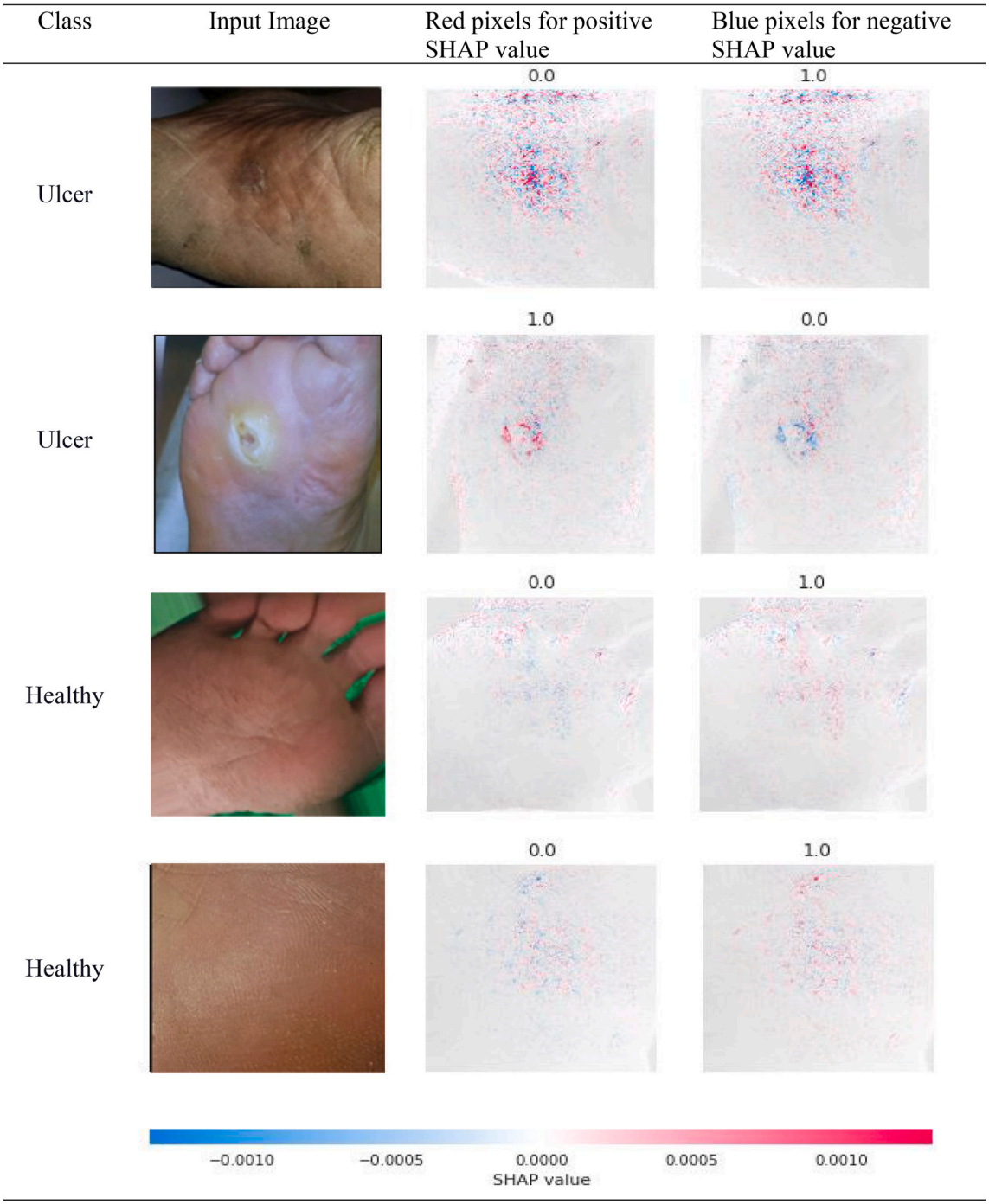


Fig. 10. Some examples of SHAP analysis for both ulcer and healthy classes.

layer, 3 convolution layers, 4 parallel convolutions with varying-sized filters, 5 max-pooling layers, and 2 fully connected layers. In Ref. [29], an SVM-based ML classifier was used for the detection and segmentation of DFU.

However, the time complexity of DL models depends on several factors, such as the number of samples, the number of extracted features, and the number of layers. In this work, assume that the number of samples is n, the number of layers of the proposed architecture is L, and the number of features extracted by the FusionNet is f. Therefore, the least time complexity of FusionNet to perform DFU detection is **O**(nfL). Table 11 provides a summary of the computational complexity of the above articles. This article analyzes the computational complexity of the CNN architecture by considering architectural components like convolution, LeakyReLU, ReLU, average pooling, max pooling, dropout, and dense layers in DL networks as one-layer unit. However, from the above discussion, we can

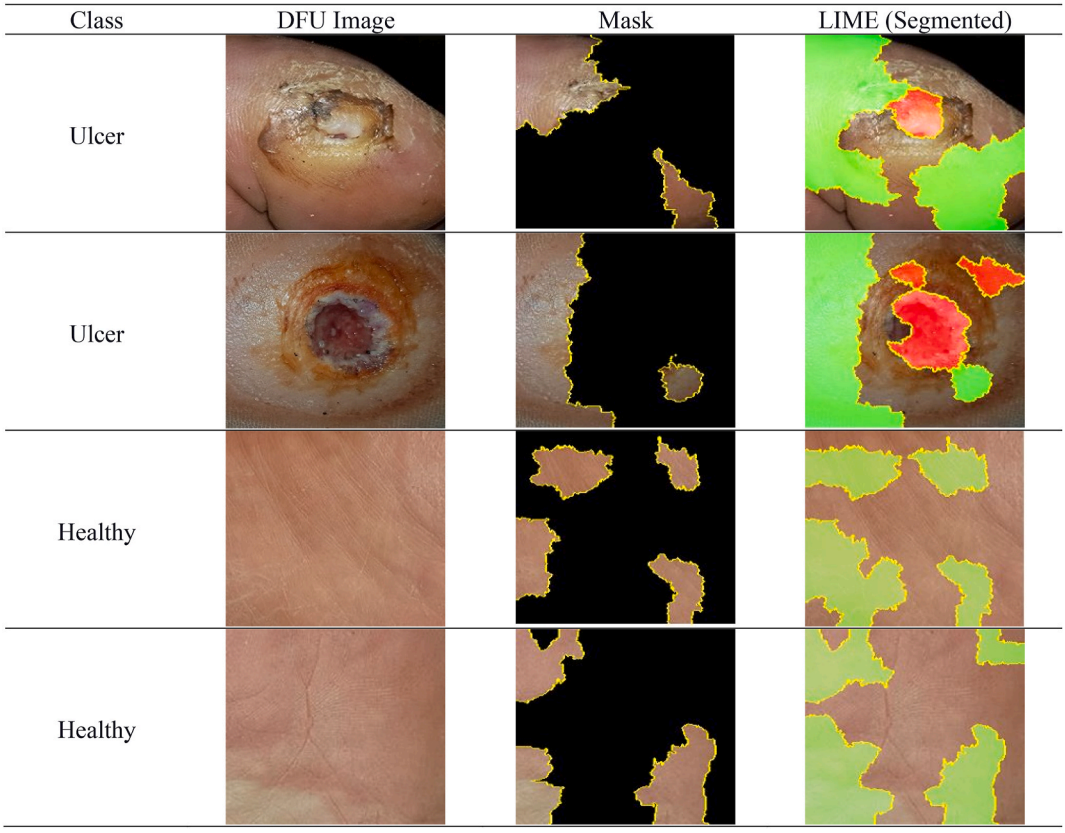


Fig. 11. Some examples of LIME analysis.

see that the proposed architecture is more complex than the other articles except [37]. Although the proposed architecture is a bit complex, the performance metrics of the proposed architecture are much higher than those of other existing works.

4.5. Statistical analysis

We conducted two statistical test approaches - Friedman and ANOVA (analysis of variance), in this article to find out the remarkable statistical differences between the proposed method and other sub-models [54]. These tests were simulated by importing the SciPy stats packages [55] developed in Python. The ANOVA test results, which highlight the remarkable statistical differences between the simulated models according to their performance metrics, are shown in Table 12. A F_statistic of 2.642 and a p_value of 0.1044 are obtained by the test reports, ensuring that there is no remarkable difference between the simulated models. From Table 13, we noted that the Friedman test is more appropriate for statistically analyzing the performance of the simulated models, although the ANOVA test showed statistically remarkable differences. The Friedman test showed statistically remarkable performance differences among the proposed methods, VGG19, NASNetMobile, and DenseNet201, with $p < 0.05$ for accuracy. The p_value of 0.02929 was obtained from the Friedman test, ensuring that the data reject the NH (null hypothesis). This result indicates that there are remarkable differences in the performance of the models used in this research.

4.6. XAI algorithms result analysis

To understand clearly which regions of the DFU image was focused by the FusionNet to explain the predicted result, we exhibit an example of Grad-CAM XAI method in Fig. 9. Here, we show the predicted DFU image utilized to input the XAI algorithm, and the heatmaps for explaining the results of classification produced by the this algorithm.

Fig. 10 displays some samples with SHAP values for the FusionNet classifier. In Fig. 10, red points indicate the positive SHAP values that contribute to increasing the output score for ulcer samples; similarly, blue points indicate the negative SHAP values that contribute to decreasing the output score. For a specific label, the prominence of the area of interest (AOI) is indicated by the overall strength of SHAP values. Thus, the SHAP approach interprets the flexibility of our suggested approach for the automatic diagnosis of DFU.

Fig. 11 exhibits the interpretability system applying LIME to the DFU dataset to prove the prediction of the FusionNet classifier. The key DL characteristics were retrieved by the XAI algorithm from the classifier's predictions; these retrieved characteristics might help

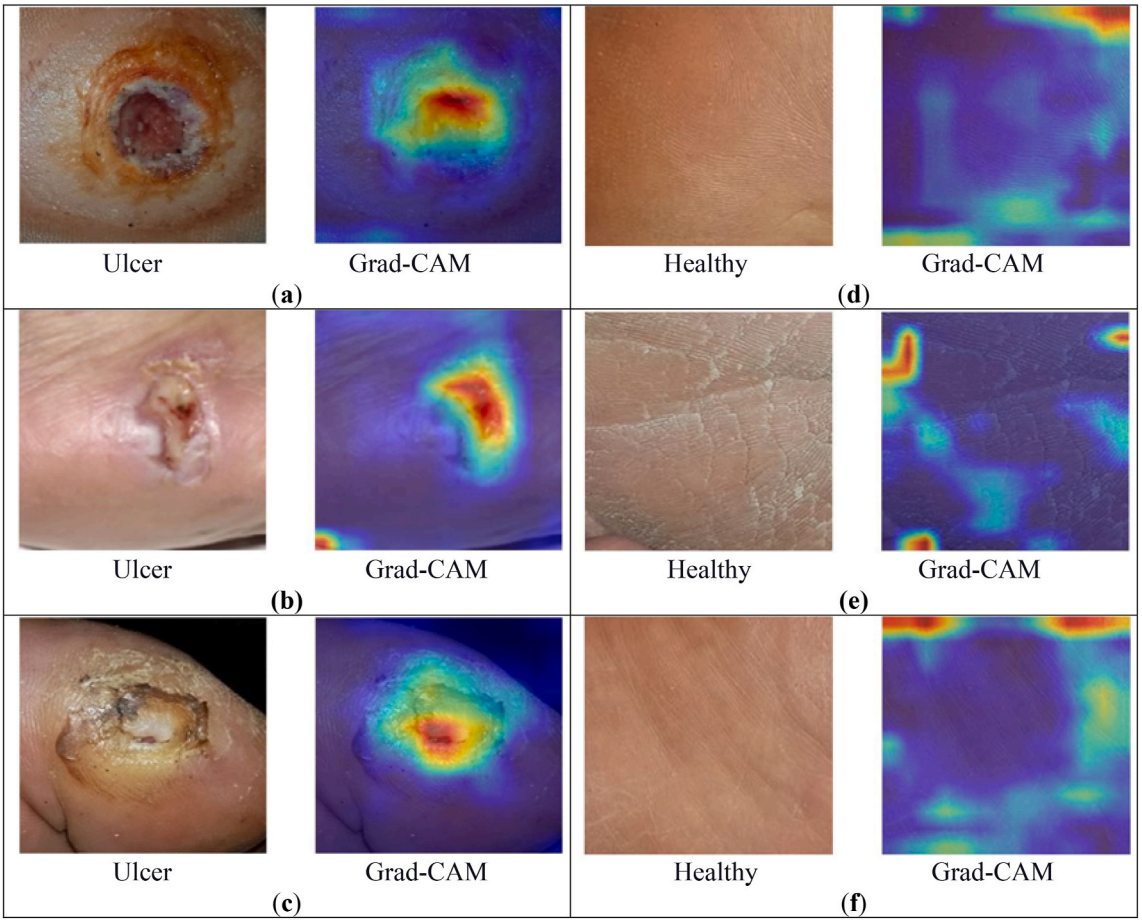


Fig. 12. Some examples of Grad-CAM analysis.

general practitioners distinguish between normal skin patches and ulcers. Upon closer inspection, Fig. 11 demonstrates how the input sample is segmented into smaller, linked areas that have a common color and position. The skin region with an ulcer is indicated by a red color, whereas the skin area with normal skin is indicated by a green color. Therefore, the resilience of our suggested approach for automated DFU diagnosis may be explained by the LIME XAI algorithm.

To comprehend which regions of the sample of the DFU dataset were highlighted using the FusionNet classifier for ulcer detection, we exhibit some samples produced by the Grad-CAM algorithm in Fig. 12. We also show the sample of the DFU dataset utilized to input the classifier, the attention maps produced by the Grad-CAM, and the overlapping of the DFU dataset and the attention maps. Fig. 12 (a)–(c) shows the ulcer samples and their Grad-CAM explanantions; Fig. 12 (d)–(f) shows the healthy samples and their Grad-CAM explanantions. From this experiment, we believe that this final visual representation can be valuable for pathologists and radiologists to pinpoint the ulcer regions to investigate. The regions scanned by the classifier for predicting ulcer presence are highlighted in yellow, with a more intense yellow indicating a higher likelihood of the projected label.

5. Conclusion

This study suggested an XAI-based multi-scale feature fusion framework (FusionNet) utilizing the TL (transfer learning) idea for automatically aiding the explanation of DL networks through visualization activation maps. To construct the FusionNet, we combined three CNN classifiers—DenseNet201, NASNetMobile, and VGG19—among five based on their performance. A global average pooling (GAP) layer is strategically attached after each classifier to preserve task-relevant information in the sample. Then, the preserved information from each GAP layer is concatenated using a concatenation CNN layer and then fed into a meta-tuner module for refining the entire network. Finally, a fully connected layer is used to classify the DFU samples. Our proposed system achieved high accuracy (99.05 %), AUC (99.09 %), precision (100.00 %), F1 score (99.08 %), and recall (98.18 %) on the test set. While the proposed classifier demonstrated a high level of performance, it is essential to note that the FusionNet is built up as a locally explained classifier with a classifier-agnostic interpretation, tailored to be shapely interpreted for a more qualitative understanding by the general public. It is essential to illustrate how a medical system functions within. To further enhance the transparency and explainability of this network, we incorporated three well-established XAI algorithms—Grad-CAM, SHAP, and LIME—to elucidate the uncertainties associated with

the classification outcomes obtained from the trained classifiers. Leveraging SHAP, with its ability to meet diverse explainability criteria such as messiness, locality, and consistency, reinforces its popularity as a robust option for model explainability. Furthermore, LIME is employed to scrutinize the top features that distinguish between ulcer and normal skin samples. In conclusion, Grad-CAM enriches interpretability by providing a detailed localization of crucial areas within a sample, enabling users to discern which aspects of the input substantially influenced the classifier's prediction for a particular level. The proposed XAI-based method could be applied in various areas related to DFU management and treatment. These application areas may include the following: 1) This method can be applied to computer-aided diagnosis (CAD) systems to reduce the diagnosis time and rapidly execute complex medical treatments; 2) trainee doctors can use this XAI-based framework to improve their diagnostic skills and treatment decision-making abilities; and 3) the proposed XAI-based method can be applied in the healthcare sector to comprehend the interpretability and transparency of the model's predictions that help DFU specialists make decisions about patient conditions.

In future research, this FusionNet paradigm may be extended to identify and categorize DFU as Charcot arthropathy, osteomyelitis, neuropathy, or ischemia. However, this approach may also be used to accurately quantify many metrics seen in DFU samples, including ulcer growth rate, volume, tissue density, and size. This quantifiable information will monitor the ongoing patient and improve their treatment planning. Additionally, in the future, we will improve our prediction outcomes by applying a novel segmentation approach to the segmented DFU dataset as well as a large-scale dataset.

Data availability statement

The dataset used in this work has been deposited into a publicly available online repository named Kaggle. The dataset can be accessed via the link: https://www.kaggle.com/laithjj/diabetic-foot-ulcerdfu.

Funding statement

Not Applicable.

Ethical approval statement

Not required.

Consent to participate statement

Not required.

CRediT authorship contribution statement

Shuvo Biswas: Writing – original draft, Validation, Resources, Methodology, Data curation, Conceptualization. **Rafid Mostafiz:** Supervision, Project administration, Methodology, Formal analysis, Conceptualization. **Mohammad Shorif Uddin:** Writing – review & editing, Visualization, Validation, Investigation. **Bikash Kumar Paul:** Software, Resources, Formal analysis, Conceptualization.

Declaration of competing interest

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.

References

[1] S. Wild, G. Roglic, A. Green, R. Sicree, H. King, Global prevalence of diabetes estimates for the year 2000 and projections for 2030, Diabetes Care 27 (5) (May 2004) 1047–1053, https://doi.org/10.2337/diacare.27.5.1047.
[2] G. Roglic, WHO Global report on diabetes: a summary, Int J Non-Commun Dis 1 (1) (2016) 3, https://doi.org/10.4103/2468-8827.184853.
[3] K. Bakker, J. Apelqvist, B. Lipsky, J. Van Netten, N. Schaper, The 2015 IWGDF guidance documents on prevention and management of foot problems in diabetes: development of an evidence-based global consensus, Diabetes/Metabolism Res. Rev. 32 (S1) (2016) 2–6.
[4] A.J. Boulton, L. Vileikyte, G. Ragnarson-Tennvall, J. Apelqvist, The global burden of diabetic foot disease, Lancet 366 (9498) (2005) 1719–1724.
[5] I.D. Federation, I.D. Atlas, IDF Diabetes Atlas, sixth ed., International Diabetes Federation, Brussels, Belgium, 2013.
[6] D.G. Armstrong, L.A. Lavery, L.B. Harkless, Validation of a diabetic wound classification system: the contribution of depth, infection, and ischemia to risk of amputation, Diabetes Care 21 (5) (1998) 855–859.
[7] P. Cavanagh, C. Attinger, Z. Abbas, A. Bal, N. Rojas, Z.-R. Xu, Cost of treating diabetic foot ulcers in five different countries, Diabetes/Metabolism Res. Rev. 28 (S1) (2012) 107–111.
[8] E. Showkatian, M. Salehi, H. Ghaffari, R. Reiazi, N. Sadighi, Deep learning based automatic detection of tuberculosis disease in chest X-ray images, Pol. J. Radiol. 87 (1) (2022) 118–124.
[9] S.I. Nafisah, G. Muhammad, Tuberculosis detection in chest radiograph using convolutional neural network architecture and explainable artificial intelligence, Neural Comput. Appl. 36 (1) (2024) 111–131, https://doi.org/10.1007/s00521-022-07258-6.
[10] R. Mostafiz, M.S. Uddin, I. Jabin, M.M. Hossain, M.M. Rahman, Automatic brain tumor detection from MRI using curvelet transform and neural features, Int. J. Ambient Comput. Intell. (IJACI) 13 (1) (Mar 2022) 1–18, https://doi.org/10.4018/IJACI.293163.
[11] J. Amin, M. Sharif, M. Raza, T. Saba, M.A. Anjum, Brain tumor detection using statistical and machine learning method, Comput. Methods Progr. Biomed. 177 (Aug 2019) 69–79, https://doi.org/10.1016/j.cmpb.2019.05.015.

[12] Md M. Hasan, Md Asaduzzaman, M.M. Rahman, M.S. Hossain, K. Andersson, D3mciAD: data-driven diagnosis of mild cognitive impairment utilizing syntactic images generation and neural nets, Brain Informatics 12960 (2021) 366–377, https://doi.org/10.1007/978-3-030-86993-9_33.
[13] A. Dahou, A. Mabrouk, M. Abd Elaziz, M. Kayed, I.M. El-Henawy, S. Alshathri, A.A. Ali, Improving crisis events detection using DistilBERT with hunger games search algorithm, Mathematics 10 (3) (Jan. 2022) 447, https://doi.org/10.3390/math10030447.
[14] S. Niu, M. Liu, Y. Liu, J. Wang, H. Song, Distant domain transfer learning for medical imaging, IEEE Journal of Biomedical and Health Informatics 25 (2020) 3784–3793, https://doi.org/10.1109/JBHI.2021.3051470.
[15] S.J. Pan, Q. Yang, A survey on transfer learning, IEEE Trans. Knowl. Data Eng. 22 (10) (Oct 2010) 1345–1359, https://doi.org/10.1109/TKDE.2009.191.
[16] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A.C. Berg, ImageNet large scale visual recognition challenge, International journal of computer vision (2015) 211–252, https://doi.org/10.1007/s11263-015-0816-y.
[17] J. Zhang, Y. Xie, G. Pang, Z. Liao, J. Verjans, W. Li, Z. Sun, J. He, Y. Li, C. Shen, Y. Xia, Viral pneumonia screening on chest X-ray images using confidence-aware anomaly detection, IEEE transactions on medical imaging 40 (3) (2020) 879–890, https://doi.org/10.1109/TMI.2020.3040950.
[18] A. Mabrouk, R.P.D. Redondo, M. Kayed, Deep learning-based sentiment classification: a comparative survey, IEEE Access 8 (2020) 85616–85638, https://doi.org/10.1109/ACCESS.2020.2992013.
[19] H.-C. Lee, A.F. Aqil, Combination of transfer learning methods for kidney glomeruli image classification, Appl. Sci. 12 (3) (Jan 2022) 1040, https://doi.org/10.3390/app12031040.
[20] A. Mabrouk, R.P.D. Redondo, M. Kayed, SEOpinion: summarization and exploration of opinion from E-commerce websites, Sensors 21 (2) (Jan 2021) 636, https://doi.org/10.3390/s21020636.
[21] G. Chandrasekaran, N. Antoanela, G. Andrei, C. Monica, J. Hemanth, Visual sentiment analysis using deep learning models with social media data, Appl. Sci. 12 (3) (Jan 2022) 1030, https://doi.org/10.3390/app12031030.
[22] S.M. Lundberg, S.-I. Lee, A unified approach to interpreting model predictions, Adv. Neural Inf. Process. Syst. 30 (2017).
[23] R.R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, Grad-cam: visual explanations from deep networks via gradient-based localization, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 618–626.
[24] M.T. Ribeiro, S. Singh, C. Guestrin, Why should I trust you?': explaining the predictions of any classifier, Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, 2016, pp. 1135–1144, https://doi.org/10.1145/2939672.2939778.
[25] W. Min, B.-K. Bao, C. Xu, M.S. Hossain, Cross-platform multi-modal topic modeling for personalized inter-platform recommendation, IEEE Trans. Multimed. 17 (10) (Oct 2015) 1787–1801, https://doi.org/10.1109/TMM.2015.2463226.
[26] X. Yang, T. Zhang, C. Xu, S. Yan, M.S. Hossain, A. Ghoneim, Deep relative attributes, IEEE Trans. Multimed. 18 (9) (Sep 2016) 1832–1842, https://doi.org/10.1109/TMM.2016.2582379.
[27] A. Mabrouk, R.P. Díaz Redondo, A. Dahou, M. Abd Elaziz, M. Kayed, Pneumonia detection on chest X-ray images using ensemble of deep convolutional neural networks, Appl. Sci. 12 (13) (Jun 2022) 6448, https://doi.org/10.3390/app12136448.
[28] M. Goyal, N.D. Reeves, A.K. Davison, S. Rajbhandari, J. Spragg, M.H. Yap, Dfunet: convolutional neural networks for diabetic foot ulcer classification, IEEE Transactions on Emerging Topics in Computational Intelligence 4 (5) (2018) 728–739, https://doi.org/10.1109/TETCI.2018.2866254.
[29] L. Wang, P.C. Pedersen, E. Agu, D.M. Strong, B. Tulu, Area determination of diabetic foot ulcer images using a cascaded two-stage SVM-based classification, IEEE (Inst. Electr. Electron. Eng.) Trans. Biomed. Eng. 64 (9) (Sep 2017) 2098–2109, https://doi.org/10.1109/TBME.2016.2632522.
[30] L. Alzubaidi, M.A. Fadhel, S.R. Oleiwi, O. Al-Shamma, J. Zhang, DFU_QUTNet: diabetic foot ulcer classification using novel deep convolutional neural network, Multimed. Tool. Appl. 79 (21–22) (2020) 15655–15677, https://doi.org/10.1007/s11042-019-07820-w.
[31] A. Doulamis, N. Doulamis, A. Angeli, A. Lazaris, S. Luthman, M. Jayapala, G. Silbernagel, A. Napp, I. Lazarou, A. Karalis, R. Hoveling, A non-invasive photonics-based device for monitoring of diabetic foot ulcers: architectural/sensorial components & technical specifications, Inventions 6 (2) (Apr 2021) 27, https://doi.org/10.3390/inventions6020027.
[32] L. Alzubaidi, A.A. Abbood, M.A. Fadhel, O. Al-Shamma, J. Zhang, Comparison of hybrid convolutional neural networks models for diabetic foot ulcer classification, J. Eng. Sci. Technol. 16 (2021) 2001–2017.
[33] J. Reyes-Luévano, J.A. Guerrero-Viramontes, J.R. Romo-Andrade, M. Funes-Gallanzi, DFU_VIRnet: a novel visible-infrared CNN to improve diabetic foot ulcer classification and early detection of ulcer risk zones, Biomed. Signal Process Control 86 (2023) 105341, https://doi.org/10.2139/ssrn.4010975.
[34] S.K. Das, P. Roy, A.K. Mishra, DFU_SPNet: a stacked parallel convolution layers based CNN to improve Diabetic Foot Ulcer classification, ICT Express 8 (2) (2022) 271–275, https://doi.org/10.1016/j.icte.2021.08.022.
[35] S.K. Das, P. Roy, A.K. Mishra, Fusion of handcrafted and deep convolutional neural network features for effective identification of diabetic foot ulcer, Concurrency Comput. Pract. Ex. 34 (5) (2022) e6690.
[36] M. Kaselimi, E. Protopapadakis, A. Doulamis, N. Doulamis, A review of non-invasive sensors and artificial intelligence models for diabetic foot monitoring, Front. Physiol. 13 (Oct. 2022) 924546, https://doi.org/10.3389/fphys.2022.924546.
[37] S. Biswas, R. Mostafiz, B.K. Paul, K.M. Mohi Uddin, M.M. Rahman, F.N.U. Shariful, DFU_MultiNet: a deep neural network approach for detecting diabetic foot ulcers through multi-scale feature fusion using the DFU dataset, Intelligence-Based Medicine 8 (2023) 100128, https://doi.org/10.1016/j.ibmed.2023.100128.
[38] P.N. Thotad, G.R. Bharamagoudar, B.S. Anami, Diabetic foot ulcer detection using deep learning approaches, Sensors International 4 (2023) 100210, https://doi.org/10.1016/j.sintl.2022.100210.
[39] S.K. Das, S. Namasudra, A. Kumar, N.R. Moparthi, AESPNet: attention enhanced stacked parallel network to improve automatic diabetic foot ulcer identification, Image Vis Comput. 138 (Oct 2023) 104809, https://doi.org/10.1016/j.imavis.2023.104809.
[40] S. Biswas, R. Mostafiz, B.K. Paul, K.M.M. Uddin, Md A. Hadi, F. Khanom, DFU_XAI: a deep learning-based approach to diabetic foot ulcer detection using feature explainability, Biomedical Materials & Devices (Mar 2024) 1–21, https://doi.org/10.1007/s44174-024-00165-5.
[41] Dataset: diabetic foot ulcer (DFU). Available: https://www.kaggle.com/datasets/laithjj/diabetic-foot-ulcer-dfu.
[42] J. Sun, Wenfei Cao, Zongben Xu, J. Ponce, Learning a convolutional neural network for non-uniform motion blur removal, in: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Boston, MA, USA, Jun 2015, pp. 769–777, https://doi.org/10.1109/CVPR.2015.7298677.
[43] S.J. Pan, Q. Yang, A survey on transfer learning, IEEE Trans. Knowl. Data Eng. 22 (10) (Oct 2010) 1345–1359, https://doi.org/10.1109/TKDE.2009.191.
[44] G. Huang, Z. Liu, L. Van Der Maaten, K.Q. Weinberger, Densely connected convolutional networks, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Honolulu, HI, Jul 2017, pp. 2261–2269, https://doi.org/10.1109/CVPR.2017.243.
[45] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, arXiv preprint arXiv:1409.1556 (Sep. 2014), https://doi.org/10.48550/arXiv.1409.1556.
[46] B. Zoph, V. Vasudevan, J. Shlens, Q.V. Le, Learning transferable architectures for scalable image recognition, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, IEEE, Salt Lake City, UT, Jun 2018, pp. 8697–8710, https://doi.org/10.1109/CVPR.2018.00907.
[47] F. Saxen, P. Werner, S. Handrich, E. Othman, L. Dinges, A. Al-Hamadi, Face attribute detection with MobileNetV2 and NasNet-mobile, in: 2019 11th International Symposium on Image and Signal Processing and Analysis (ISPA), IEEE, Dubrovnik, Croatia, Sep. 2019, pp. 176–180, https://doi.org/10.1109/ISPA.2019.8868585.
[48] J. Koushik, Understanding convolutional neural networks,, 2023.
[49] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, Dropout: a simple way to prevent neural networks from overfitting, J. Mach. Learn. Res. 15 (2014) 1929–1958.
[50] G.E. Dahl, T.N. Sainath, G.E. Hinton, Improving deep neural networks for LVCSR using rectified linear units and dropout, in: IEEE International Conference on Acoustics, Speech and Signal Processing, IEEE, Vancouver, BC, Canada, May 2013, pp. 8609–8613, https://doi.org/10.1109/ICASSP.2013.6639346.
[51] F. Chollet, Keras: Deep learning library for theano and tensorflow 7 (8) (2015) T1. Available:, https://keras.io/.
[52] N. Ketkar, Deep Learning with Python, Apress, Berkeley, CA, 2017, https://doi.org/10.1007/978-1-4842-2766-4.

[53] D.P. Kingma, J. Ba, Adam: a method for stochastic optimization, in: the 3rd International Conference for Learning Representations, San Diego, 2015, https://doi.org/10.48550/arXiv.1412.6980.
[54] J. Demsar, J. Demsar, Statistical comparisons of classifiers over multiple data sets, J. Mach. Learn. Res. 7 (2006) 1–30.
[55] S. Seabold, J. Perktold, Statsmodels: econometric and statistical modeling with Python, in: Presented at the Python in Science Conference, 2010, pp. 92–96, https://doi.org/10.25080/Majora-92bf1922-011. Austin, Texas.