Optimized Classification with CBGRU-CapNet on Thermographic Analysis to Detect Diabetic Foot Ulcers with Precise Diagnosis

Ashok Babu Ane[1*], Gireesh Babu C N[2], Karthik Govindan Manoharan[3], Rafeeq Md.[4], Bhupathi Rajarao[5], and Kadiyala Vijaya Kumar[6]

[1] Department of Electronics & Communication Engineering, P.V.P Siddhartha Institute of Technology, Vijayawada 520007, Andhra Pradesh, India.

[2] Department of Computer Science and Engineering, BMS Institute of Technology & Management, Bengaluru 560064, Karnataka, India.

[3] Department of Information Security, School of Computer Science and Engineering (SCOPE), Vellore Institute of Technology, Vellore 632014, Tamil Nadu, India

[4] Department of Computer Science and Engineering, CMR Engineering College, Hyderabad 501401, Telangana, India

[5] Department of Electronics and Communication Engineering, Eluru College of Engineering and Technology, Eluru 534004, Andhra Pradesh, India

[6] Department of Computer Science and Engineering − Artificial Intelligence, G. Pullaiah College of Engineering and Technology, Kurnool 518002, Andhra Pradesh, India

e-mail: aneashokbabu@gmail.com

Abstract. Diabetic foot ulcers (DFUs) are a severe complication of diabetes mellitus, often leading to lower limb or foot amputation due to infections and a high recurrence rate. Recent research focuses on early detection of underlying conditions that cause skin and tissue damage before visible lesions appear by analysing infrared thermograms of the plantar aspect of both feet. Most existing studies rely on the publicly available INAOE dataset, which includes thermographic images of both healthy and diabetic subjects. However, advancements in the field have been supported by the newly released STANDUP dataset, which offers a more extensive and diverse collection of thermograms. This study utilizes a broader dataset to improve DFU classification. Features are extracted from preprocessed input data using both statistical methods and deep feature extraction techniques. A novel deep learning model, CBGRU-CapNet − combining convolutional layers, bidirectional gated recurrent units (BiGRU), and capsule networks − is employed for classification. To enhance accuracy and reduce computational complexity, the study introduces a fine-tuning approach using the Modified Directional Transport Metaheuristic Algorithm (mDTMA). Experimental results demonstrate that the proposed model achieves superior predictive performance and robustness, making it highly suitable for accurate and consistent DFU detection across multiple evaluation metrics. The results demonstrate that the proposed model provides superior predictive power and robustness, indicating its suitability for tasks requiring high accuracy besides consistency across multiple estimation metrics.

Keywords: Diabetic foot ulcer; thermography; directional transport metaheuristic algorithm; bidirectional gated recurrent units; deep features; Diabetes mellitus.

Paper #9198 received 12 Dec 2024; revised manuscript received 18 Feb 2025; accepted for publication 25 Feb 2025; published online 30 Mar 2025. doi: 10.18287/JBPE25.11.010306.

1 Introduction

Diabetes mellitus (DM) is another name for diabetes. Low blood sugar, or hyperglycemia, is the root cause of this chronic condition that can cause amputation of lower limbs, heart disease, blindness, and renal failure, among other devastating complications [1]. Diabetes foot ulcers (DFUs), commonly known as sores, are open sores that can develop on the foot of a diabetic. More than 422 million people had diabetes in 2014, up from 108 million in 1980, according to statistics from around the world. Among adults, for example, the total frequency increased from 4.7% in 1980 to 8.5% in 2014 [2]. In addition, 600 million individuals will have diabetes globally by the end of 2035, according to predictions [3]. The fact that most of these people are from impoverished countries with inadequate healthcare systems and generalized ignorance is also important to consider. Amputation of the lower limbs may be necessary in 15–25% of diabetic individuals who develop DFU later in the course of the disease if adequate care is not accessible [4].

Due to insufficient diagnosis and treatment, almost one million diabetics with "high-risk feet" lose a limb every year [5]. Extra care is required of those with diabetes, who must maintain good personal cleanliness, take their prescriptions as prescribed, and see their doctors frequently for exams [6]. Consequently, the financial burden will be particularly heavy for families dealing with diabetes, particularly in underdeveloped nations where the cost of treatment is approximately 5.7 periods the yearly income [7]. The prevalence of diabetes is expected to increase, putting an increasing number of entities at risk for diabetic foot ulcers [8]. The already-scarce resources and expertise will be even more stretched thin as a result of this. Inadequate treatment causes almost one million diabetics to lose a vital body part every year, such as a foot [9].

To get to the correct diagnosis, doctors have to go through mountains of medical records. Manual diagnostic procedures are more prone to error and can be costly, while computer-aided diagnostic techniques have the potential to be more efficient and cost-effective [10]. Technological advancements in the field of wearable health and mobile health have alleviate complications associated with diabetes by tracking damaging foot stress and inflammation, enhancing patients' well-being, and extending the time it takes for remission to set in Ref. [11]. There is an immediate need for a cost-effective, non-invasive, remotely deployable, and highly reliable automated DFU diagnostic method. Deep learning (DL), intelligent systems, and the proliferation of applications have all contributed to the emergence of DFU disease detection as a vibrant and rapidly evolving area of research [12]. The DFU test now includes numerous important mechanisms for early diagnosis, thanks to current clinical practice. A diabetic foot specialist or evaluates the patient's medical history and does an evaluation of the DFU. Additional tests, including an X-ray, CT scan, or MRI, administered as needed.

Technological advancements in the last several years have allowed the biomedical field to make great strides, particularly in areas such as the explanation of protein structures, the classification of biomedical images, and the alignment of genetic sequences [13]. In order to comprehend, evaluate, and preserve the growing body of biomedical data, dependable and effective computer systems are essential. DL-based methods are capable of addressing these intricate problems. It is possible to efficiently acquire hierarchical DL features from raw data using DL techniques, and neural networks in particular [14]. Because of this, these algorithms are able to decipher complex biomedical data, including its representations and linkages. Due to the increased focus and investment in DL recently, it has grown and is now widely used in many industries. The biomedical industry has started to prioritize the use of DL techniques because of their ability to assess massive datasets, find patterns that are relevant to tasks, and generate predictions. In the modern, digital crucial for the diagnosis of numerous health issues [15].

When it comes to medical imaging, traditional ML and DL algorithms succeed by picking and extracting features that are more shape, colour, and size sensitive. Using ML and DL approaches, DFUs have been successfully discovered and detected in previous studies. Numerous scholars have looked into these challenging difficulties, but thus yet, no practical answers have been identified [16]. Unfortunately, a lot of ML and DL algorithms are complicated and opaque, so it is hard to figure out how the model is making judgments or predictions and to spot possible biases or mistakes. Medicine, economics, and law enforcement are just a few fields where model errors can have far-reaching consequences. The difficulty in tuning and generalizing black box models to fresh data is further compounded by their sensitivity to hyperparameter choice [17]. Another problem that can arise with black box models is overfitting, which can cause them to perform poorly when tested with new data. Despite black box models' success in many domains, their opaqueness and lack of interpretability raise serious concerns.

If possible, ML and DL models should be able to estimate the severity of the lesion and categorize people at risk of emerging ulcers using a single thermogram that includes the plantar feet. Thorough data interpretation is of the utmost importance in the healthcare setting. While numerous indicators have been suggested for use in thermography-based foot disease identification, it is difficult to ascertain which ones are most indicative of DFUs [18]. It could be difficult to understand data when there are a lot of features. Experts' differing interpretations of the data can cause more variation in clinical decision-making due to discrepancies in illness diagnoses. Consequently, in order to improve the overall cost-performance for classification and decrease decision variability and data misinterpretation, it is necessary to develop a subset of significant variables for thermography-based foot disorder detection [19]. Because lowering the amount of features can memory

resources, classifiers with improved cost-performance ratios are attained by using a subset of features containing relevant information [20]. This appropriate subset of characteristics cannot be established due to the imbalanced datasets cases and the absence of consistency among thermograms.

To find out what properties are important for early DFU detection, ML and DL models have been investigated. The research primarily used the INAOE dataset, which is comprised of thermograms showing the plantar aspect of both feet, and was the only publicly available dataset, with a few exceptions. Just recently, a comparable dataset called STANDUP was made public. It offers ways to improve upon the existing state of the art by just adding more samples to train the ML and DL representations. Additionally, the supplementary dataset allows for the assessment of the set of state-of-the-art features' generalizability, which was previously determined using classical and DL methods.

How the residue of the paper is structured is as shadows: In Section 2, the relevant literature is reviewed; in Section 3, the suggested classical is fleetingly described; in Section 4, results are analysed; and in Section 5, conclusion is drawn.

2 Related Works

The research was carried out by Özgür et al. [21] from November 2023 to February 2024. The information was gathered from three Turkish hospitals' endocrinology and metabolic diseases departments. Data was collected using an identification form besides the Foot Care Scale. The significance of the variables was predicted using gradient boosting methods. The research included XGBoost, LightGBM, and Random Forest as its machine learning methods. Patients whose scores were either higher than the mean FCS-OD score or lower than it was predicted using the procedures. The best result was achieved by XGBoost, with an area under the curve (AUC) of 0.7469. In all of the models, gender (0.0038), perceived health condition (0.0218), and treatment regimen (0.0027) were the most common components. The best area under the curve (AUC) model, XGBoost, also found that income level (0.0055) and A1c (0.0020) were predictors of score.

The goal of the study by Chen et al. [22] was to find potential treatment targets by detecting genes associated with macrophages in DFU. This work utilised datasets from the gene RNA-seq datasets linked to DFU. Macrophage subpopulation distribution in the DFU was shown by analysis of the given single-cell data. For additional bioinformatics analysis, four separate RNA-seq datasets were combined into a single DFU cohort. Diverse machine learning techniques for biomarker identification, enrichment analysis, and differential expression (DEG) analysis were all a part of it. Western transcription-quantitative polymerase chain reaction (RT-qPCR) were used to authorize the important findings. Lastly, western blotting and RT-qPCR were used to confirm the results. Using single-cell analysis, to were able to recover 802 genes associated with macrophages. There were 743 differentially expressed genes (DEGs) found following the study of expression. Through the process of cross-analysis, 37 differentially expressed genes (DEGs) related with macrophages were discovered. Four machine learning techniques were used to screen and cross-analyse 37 junctions. Lastly, HMOX1 was recognised as a biomarker that could be very useful. Several metabolic pathways, including the insulin signalling system, were substantially connected with HMOX1. In DFU samples, HMOX1 was shown to be considerably overexpressed. Finally, this study's analytical results showed that HMOX1 is a macrophage-associated biomarker in DFU that could be useful. Our work sheds light on how macrophages contribute to this condition and could lead to more effective targeted treatments for diffuse focal ulceration (DFU).

Using machine learning, Hong et al. [23] assessed the likelihood of diabetic foot ulcers (DFUs) returning in older diabetic patients with the goal of improving DFU prevention and care. The objective is to develop reliable models that can predict the likelihood of DFU recurrence in older diabetic patients by taking into account risk factors such age, blood sugar control, alcohol intake, and smoking. Machine learning models were built using data obtained from 138 older diabetic patients following data cleansing, screening for outliers, and feature integration. A 93% success rate was achieved using Support Vector Machine (SVM). The experimental consequences demonstration that SVM is a good tool for forecasting the probability of DFU recurrence in older diabetic patients, giving doctors a better gauge to work with. This research shows that machine learning is very useful for forecasting the likelihood of recurrence of foot ulcers in older diabetes patients, which is crucial for their management. This strategy utilises computer-assisted medical procedures to control diabetes in the elderly in a way that allows for timely intervention, which in turn reduces the likelihood of patient recurrence.

In their study on semantic segmentation of photographs of diabetic foot ulcers (DFUs), Saeedi et al. [24] looked at how the U-Net model was applied inside a federated learning framework. Training in both a centralised and a federated learning environment is the goal, with a set of randomly generated U-Net model parameters serving as the basis. We will use an open-source dataset of diabetic foot ulcers supplied by Medetec because medical photos are quite sensitive. There is no longer any need to centralise raw data thanks to federated learning, which allows us to decentralise our approach to machine learning. One of the methods utilised is to compare and contrast centralised and federated machine learning systems through simulations. According to the findings, federated learning in conjunction with the U-Net architecture can obtain a noteworthy dice score of 0.9, which is comparable to the performance of centralised models, while also eliminating the need for centralised data collection. Attaining a balance between privacy issues and analytical accuracy, this conclusion highlights the potential of federated learning to improve detection approaches for DFUs. The federated learning codebases provided by this study allow interested researchers to

replicate and build upon the findings, making a significant contribution to the biomedical imaging field. On GitHub, you can find the source code.

Verma [25] has used a deep learning model to analyse thermal photographs of the foot in diabetic patients in order to diagnose foot ulcers. The model's efficacy was estimated by comparing it to previous research. The study utilised thermal pictures that are available for free online. There are two categories of diabetic foot photos in the dataset: normal and pathological. Out of a total of 1055 photographs in the collection, 543 depict the patient's normal foot and the remaining images show the patient's aberrant foot. After the study's dataset was pre-processed using watershed segmentation and clever edge detection, it was transformed into a new dataset. The pre-processed dataset was subsequently enhanced and balanced by data augmentation. Following this, learning model was utilised for prediction purposes in order to diagnose foot ulcers. The pre-processed dataset, improved by clever edge detection and segmentation, can lower computing cost and improve model performance for accurate predictions. Both the original datasets were used to evaluate our proposed model, which utilised EfficientNetB0. The pre- dataset had edge detection and segmentation applied to it. Impressively, ResNet50 achieved 89% and 89.1% accuracy for the two datasets, correspondingly, whereas EfficientNetB0 outperformed with 96.1% and 99.4% accuracy, respectively. The results were incredibly encouraging.

Wu et al. [26] proposed a developed YOLOv5-based model for vascular stenosis small target detection. Different types were identified by using GhostC3 and tested on the publicly available dataset for detection of diabetic feet.

Using hybrid deep learning models, Chee et al. [27] have studied the use of gait diabetes detection. Hybrid data, like ECG data and the Pima Database, are used in current research. This study uses a hybrid deep layers make up the suggested CNN-LSTM model. CNN-LSTM can learn patterns for classification and extract significant features by merging the two models that are used to compare the CNN-LSTM model's performance to that of the CNN and LSTM models. The accuracy of CNN-LSTM is 91.25%, which is higher than that of current techniques. Therefore, this study shows that non-invasive diabetes detection methods have the potential to displace more conventional invasive approaches. Future research could examine muscle activation and forces in conjunction with gait acceleration to enhance the model's ability to detect diabetes.

The objective of Castillo-Morquecho et al. [28] is to use thermography to examine the foot sole investigate any relationships with clinical variables. A machine learning method was also created for classification. There were twenty-seven age- and sex-matched controls and twenty-three diabetic patients. The plantar foot surface's thermograms were obtained and divided into areas of interest. Using predetermined regions of interest, the mean foot temperature and temperature change index were computed. Glycated haemoglobin, body mass index, and temperature measurements were all correlation analysis. Principal component investigation and support vector machines were used in a two-layered cross-validation model for classification. Mean foot temperature was found to have significant positive correlations with both body mass index ($\rho = 0.35$, $p = 0.013$) and glycated haemoglobin ($\rho = 0.44$, $p = 0.0015$). There were no discernible relationships among the temperature change index and clinical variables. Based on thermal data, the machine learning model classified the diabetic and control groups with a moderate sensitivity (78.3%) and high overall accuracy (90%) and specificity (100%).

The main goal of Agrawal et al. [29] is to use insoles with wireless plantar pressure sensors to identify and categorise diabetic feet from healthy ones using dynamic plantar pressure. Comparing machine learning (ML) and statistical-based classification techniques is the secondary goal. In line with earlier studies, data from 150 participants who wore the insoles while walking showed that diabetic feet have higher plantar pressure than healthy feet. With an accuracy of 0.85, the Adaptive Boosting (AdaBoost) machine learning model outperformed the statistical approach, which had an accuracy of 0.67. These results imply that plantar pressure features can be used to efficiently classify diabetic and healthy feet when paired with machine learning models and insoles with built-in pressure sensors. Future studies will concentrate on classifying different stages of diabetic neuropathy using these insoles and machine learning, with the goal of predicting foot ulcers early in the home environment.

Using bioinformatics and machine learning analysis, Shi et al. [30] seek to find and validate putative biomarkers of Gln metabolism linked to DFU. Two microarray datasets were obtained from the Gene Expression Omnibus (GEO) database to identify the related genes. To examine the relationship between immune cell infiltration status and deGlnMRGs, the connections were investigated. Notably, WGCNA can be used to pinpoint genes that are differentially expressed (DEGs) within particular clusters. The best machine learning model was then created and vetted. Additionally, two clusters were found in DFU. Immune heterogeneity was found in these two clusters, according to immunological infiltration analysis.

In improve the accuracy of diagnosing Diabetic Foot Ulcers, El-Kady et al. [31] has concentrated heavily on medical image analysis. The hybrid models were evaluated by using the publicly available dataset. Notable results were obtained; the ResNet50 scored 0.75 on the F_1-score and averaged 0.76 on accuracy and precision. With an F_1-score of 0.84, a regular accuracy of 0.84, besides precision of 0.85, the hybrid model outperformed these measures. More accurate and efficient DFU identification in clinical settings may be possible thanks to this study, which adds to the growing in medical image processing. The diagnostic accuracy was gradually increased due to the implementation of hybrid proposed model.

Building a convolutional neural network to categorise thermal imageries of the feet as “normal” or “diabetic” has been the primary focus of Sulayman et al. [32]. The authors have performed multiple experiments on various parameters, including optimisers, activation functions, and transfer learning base models, in an effort to increase the model’s accuracy. This work uses the TensorFlow library, and the Keras package for scripting the code. Matplotlib and Excel are used to simulate the experiments and plot the results as training and validation accuracy and loss graphs. For the sake of comparison, only one parameter is varied in each experiment; all other parameters stay the same. So that the findings are solely impacted by the optimiser change, other parameters are kept constant for all trials in an experiment to find an effective optimiser, and the same process is used for all the other experiments. The built convolutional neural network makes advantage of transfer learning to enhance training efficiency and speed while compensating for the short dataset. A comparison is provided between ResNet-50, EffecientNetB3, and VGG16, three of the top transfer learning models. Adam and SGD were evaluated in terms of optimisers. The sigmoid function and the ReLU activation function were compared. The prevailing belief was that the optimal outcome could be achieved by integrating the most effective parameters from each experiment. With a validation accuracy of 90.5%, it was clearly an improvement above previous results. Afterwards, an accuracy of 86.7% was achieved when the prototypical was pictures that were not included in the training dataset.

Illuminate®, a point-of-care platform developed by Sridhar et al. [33], combines autofluorescence intensity imaging with deep learning in less than 2 min. The DenseNet with UNet model was used to detect the gram-type DFU model.

The purpose of the study by Cakir et al. [34] is to assess how well a DL algorithm can distinguish between bone marrow signals from osteomyelitis. The study’s local ethics committee gave their assent to the retrospective design. Partitioning the 148 patients into their respective T1- and T2-weighted images yielded 679 labelled regions 714 labelled regions (160 CNO, 272 OM, and 282 TR), respectively. For T1-weighted images, the ResNet-50 accuracy values were calculated to be 96.2% and 97.1%, correspondingly. Also, for T2-weighted images, the ResNet-50 achieved an accuracy of 95.6% and the EfficientNet-b0 96.8%. In addition, ResNet-50 has a sensitivity of 91.1% for CNO, 92.4% for OM, and 96.6% for TR, as reported by BCC. EfficientNet-b0, on the other hand, has a sensitivity of 93.2% for T1, 97.6% for OM, and 98.1% for TR. For TR, ResNet-50 achieves specificity scores of 98.1%, 97.9%, and 94.7% in T1-weighted pictures, and 98.6%, 97.5%, weighted images, respectively. In a similar vein, EfficientNet-specificity b0’s for T1-weighted pictures are 99.1%, 98.5%, and 98.7%, while for T2-weighted images it is 98.7%, 98.7%, and 98.4%. Deep learning techniques accurately distinguish CNO, OM, and TR in diabetic feet without invasive procedures.

By combining state-of-the-art technology with traditional algorithms, smart sensors linked to an Internet of Things infrastructure and a Raspberry Pi, Murugan and Mohankumar [35] suggest a method for early diagnosis of diabetic complications. It greatly enhances the treatment that is given to people with diabetes. An example of real-time data is the utilisation of a Convolutional Neural Network (CNN) model to interpret things like heel pressure, temperatures, and mobility habits. By distinguishing between normal foot health and the first signs of chronic difficulties, the CNN allows for remote monitoring and rapid treatments to lessen the likelihood of diabetes’s complications. Individuals are able to take charge of their health thanks to a user-friendly interface and the rapid notifications provided by a cloud-based infrastructure. The application of this novel tactic signifies an important advancement in the management of foot complications associated with diabetes. It provides a preventative measure to guarantee patients’ health and welfare and enhance systemic management. The system’s potential to revolutionise diabetes treatment methods is demonstrated by how easily it can be integrated into daily life. This paves the way for a thorough and effective method of reducing diabetic foot complications.

3 Materials Besides Methods

A state-of-the-art set of features may be extracted thermograms using the deep learning methodology presented in this study. The feature extraction process took into account three input datasets that were created by combining separate datasets. Using both traditional and deep learning methods, the study extract a subset of features from the input datasets. The CBGRU-CapNet classifier was fed a subset of features that were common to all the methods used. The study work employs the mDTMA to fine-tune the CBGRU-CapNet settings. Each set of retrieved characteristics from the STANDUP besides expanded databases was evaluated and compared using the suggested classifier as a reference. Furthermore, features that were more relevant and robust were retrieved in this work and compared to features that were extracted using only the INAOE dataset. In this section, brief clarification of projected model for DFU is mentioned in Fig. 1 and each of its block is explained in detailed.

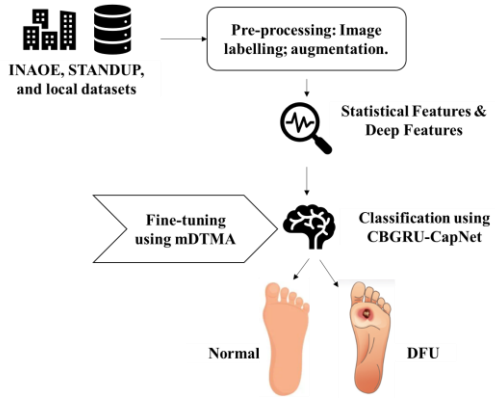

Fig. 1 Workflow of the research model.

3.1 Datasets Description

Throughout this work, infrared (IR) photos captured at various locations and with various sensors are used as part of the datasets. For every dataset, you can also find RGB versions of the same scene. The sole of the foot can be more easily separated from its surroundings with the help of these photos. Also, a different demographic sample was used to launch the matching acquisition campaigns at different times. You may find more information on the infrared acquisition procedure used for the INAOE, STANDUP, and local datasets elsewhere.

3.1.1 STANDUP Dataset

For DFU finding using infrared thermography, the STANDUP dataset was created and released in June 2023 [36]. Among the 227 participants in the study, 145 had diabetes and 82 were considered to be in good health. After 10 min of resting in a position to achieve thermodynamic equilibrium, photographs were taken at T0 for each participant. Then, after a cold stress test, photographs were taken at T10. R0 was for people without neuropathy or ischaemia; R1 was for people with neuropathy but no ischaemia; and R2 was for people with ischaemia in the diabetic group. This study did not differentiate between diabetic and healthy patients when determining pathology grades; all subjects were included for classification. In addition, T0 photos are the main focus of the analysis, which allows for useful comparisons with other INAOE and local. In other words, thermograms from people who were cold-stress tested were not taken into account. As a consequence, 145 individuals with diabetes and 38 healthy controls were removed from STANDUP dataset. Take note that this dataset is severely skewed towards cases of diabetes, just like the INAOE dataset.

3.1.2 INAOE Dataset

Among the 167 volunteers included in the INAOE thermogram dataset, which was made public in December 2019, 122 had diabetes and 45 did not [37]. The INAOE dataset has an imbalance in the number of cases with diabetes. It was initially planned to use this dataset to compare the distribution of plantar area temperatures in participants with and without diabetes. Nonetheless, thermograms have found extensive application in the early diagnosis of DFU.

3.1.3 Local Dataset

This effort is associated with the public release of the local dataset, which was acquired in 2021 [38–39]. The dataset includes samples taken from 22 healthy unpaid helpers at four separate time points; however, metaphors from the 15-min resting site before acquisition were used. This dataset was combined INAOE dataset to create a larger dataset that aims to make up for the INAOE dataset's bias towards diabetes cases.

3.2 Pre-Processing

In order to recover the accuracy of the classification, it is necessary to clean the input data before processing it.

3.2.1 Image Labeling Approach

On the other hand, a large quantity of labelled data is necessary for a deep learning model to perform optimally, particularly one that is CNN-based. However, it can be time-consuming and expensive to acquire a mountain of medical data. Consequently, methods such as regularisation, data picture labelling can be utilised to improve the performance of DL models and address overfitting. This problem is solved by expanding the dataset using a patch labelling strategy. This method of sample labelling involves selecting large portions of a sample, cropping them, and then assigning the appropriate class to the cropped samples. Initially, to extract the Region of Interest (ROI) from the samples, the research work uses the 224-pixel-wide and 224-pixel-high windows value. Classification of the patches into ulcer or healthy categories is based on the areas that encompass both ulcer skin and normal skin. Instead of focussing on the full sample, the DFU_ XAI framework labelled areas that were relevant (i.e., areas connected to ulcers) within bigger samples. This allowed the DL models to be more efficient. We used this method to extract relevant DL data for classification. Overfitting can be mitigated in part by training DL models with relevant DL data, which decreases computational load and memory allocations.

3.2.2 Data Augmentation

Data augmentation approaches are utilised to augment data in order to make it more varied and larger. These methods include random scaling besides rotation, besides contrast enhancement using new colour spaces. We used the rotation method in our research, which involves spinning each of the input skin patches through two different angles of 90° and 180°. The "train_test_split" function makes short work of dividing the DFU datasets into a "train" set consisting of 80% of the data besides a "test" set containing of 20%. This method is imported using the "sklearn.model_selection" package in Python. Ten percent of the training set's data is separated once more for validation using the same splitting procedure.

3.3 Features Extraction

This subsection outlines the two distinct feature extraction methodologies that to compare: the statistical method and the deep features-based method.

3.3.1 Statistical Features Extraction

Extracting features from raw data using various statistical procedures is what statistical features extraction is all about. Following a method that is commonly used in literature, we calculated 9 statistical indicators for each variable and time window: median, standard deviation,

skewness, kurtosis, quartile. In order to facilitate the ML algorithm's eventual discovery of patterns in the data, the acquired values are then normalised so that all of the values are on the same scale. Ultimately, we obtained a matrix in $\mathbb{R}^{N\times 9\times V}$, where V is the sum of variables and N is the number of time windows.

3.3.2 Convolutional Neural Network

One category of neural network developed to detect input spatial patterns is the CNN [40]. It is based on the principle of weight sharing, wherein a single weight is utilised by multiple neurones in a layer. Hence, a layer's output, also known as the feature map, can be determined by taking the layer's convolutional product with a specific matrix, also known as a convolutional filter. For this reason, the feature m in the formula for a 1D-CNN with a single filter is mentioned as $M \in \mathbb{R}^{\mathrm{K}}$ can be written as linear elements in the input layer $I = \{I_{j,h}\} \in \mathbb{R}^{N\times H}$ weighted by the filters $W = \{W\} \in \mathbb{R}^{L}$:

$$M_k = f\left(\sum_{l=1}^{L}\sum_{h=1}^{H} W_l l_{(k-1)*s+l,h}\right), \quad (1)$$

where s is the stride of the convolution and the dimension of the feature map, K is equaled to $\left[\frac{N+1-L}{s}\right]$.

The transposed convolutional layer, also called the deconvolutional layer, is another variety of convolutional layer. For the convolutional layer, it is a common way to go back to the input's original dimensions while still keeping the spatial relationships between data intact. The feature map is thus derived from the aforementioned formulation. $M \in \mathbb{R}^{K}$ has the dimension equal to $K = N + (L-1) * s$.

3.3.3 AutoEncoder

Reducing the feature space's dimensionality is the goal of an AutoEncoder (AE) neural network. Projecting the raw data into a subset called latent space allows us to complete this task. By filtering out irrelevant data and background noise, the projection is engineered to retain key details.

The encoder and decoder are the two main parts of an AE. The first one maps the inputs onto the latent space, and the second one uses the encoded input to rebuild the features. The decoder is required to train the complete AE, even if our aim only requires the encoder. By using the l2 norm of reconstruction provided by the AE as a loss function, the network is trained to extract a meaningful latent space.

The AE is built using the CNN in our application. Specifically, there are layers in the encoder and two deconvolutional layers in the decoder. To get a latent space with dimensions comparable to the statistical method, the kernel size and stride of the encoder layers are selected. On the flip side, deconvolutional layer hyperparameters are selected with the goal of achieving Decoder symmetry with respect to the Encoder.

The AE is built using the CNN in our application. Specifically, there are two convolutional layers in the encoder and two deconvolutional layers in the Decoder. The goal of selecting the encoder layers' kernel sizes and strides is twofold: first, to achieve a latent space with dimensions comparable to those of the statistical method; and second, to maintain a smooth dimensionality drop between them. Deconvolutional layered feature maps, on the other hand, are created with an effort to maintain encoder-decoder symmetry. For activation function, the study uses the ReLU. Other options, like sigmoid and tanh, have also been investigated. But they do not seem to boost ReLU's performance much at first look, so to stick with this activation function.

3.4 Classification Using CBGRU-CapNet

In this work, the CBGRU-CapNet is used to detect the DFU from the extracted statistical and deep features from CNN-AE. The multi-layered architecture of the model is examined in the following subsections.

3.4.1 Convolutional Layer

Three convolutional layers − conv1, conv2, and conv3 − in a convolutional neural network (CNN) are used by the model to extract regional patterns input data. These layers use kernel sizes of 2, 3, and 4, sliding filters over token embeddings to create feature maps. These maps capture interactions among neighboring tokens, extracting spatial, and temporal highlights associated with DFU. With 128 channels of varying sizes, the convolutional layers perform operations and derive a feature sequence *fs* from the input data. This sequence undergoes maximum pooling and concatenation, ultimately contributing to the subsequent layer's final feature vector.

$$fn = f(wtxt + b). \quad (2)$$

3.4.2 BIGRU Layer

A bidirectional Gated Recurrent Unit (BiGRU), a specific kind of recurrent patterns and relationships within the data, is used in this layer. The GRU's bidirectional capability allows it to analyse input sequences both forward and backward, at the same time collecting data from the past and the future. In the BiGRU setup, a forward GRU processes succeeding sequences (f1 to f128), while a backward GRU handles preceding sequences (f128 to f1).

The combination of CNN and BiGRU in a proposed architecture is a key to boosting accuracy. The integration of CNN and BiGRU not only ensures the extraction of features but also finds unique trends and patterns in the contextual features of the various ROI of three datasets. Hence, the above-discussed combination provides extraordinary mapping between the input images and extracted features in the DFU detection process.

$$ForwardHiddenState(h_f) = GUR(L_{fs}),\ n \in [1{,}128], \quad (3)$$

$$BackwardHiddenState(h_b) = GUR(L_{fs}),\ n \in [128,1], \quad (4)$$

$$TotalHiddenState(h_t) = [h_f, h_b]. \quad (5)$$

3.4.3 GRU Cell

The cell is equipped with gate (i_t), the forget gate (f_t), and the output gate (o_t). These gates serve the purpose of preserving and modifying information related to the data preceding time *t*. The cell states and updating the parameters is governed by Eq. (6) to Eq. (7).

The Gated Recurrent Unit (GRU) is a particular kind of cell created to detect long-range relationships in sequential input while minimising vanishing gradient issues. It is made up of a candidate activation ($\tilde{h}_t$) that computes the intermediary hidden state, an update gate (z_t), and a reset gate (r_t), which manage the movement of information, and the last hidden state to be computed (h_t), which acts as the cell's output. The equations for the GRU cell are as shadows:

$$(updategate) z_t = \sigma(W_z \cdot [h_{t-1}, x_t]), \quad (6)$$

$$(Resetgate) r_t = \sigma(W_r \cdot [h_{t-1}, x_t]), \quad (7)$$

$$\tilde{h}_t = tanh(W_h \cdot [r_t \cdot h_{t-1}, x_t]), \quad (8)$$

$$(newhiddenstate) h_t = (1 - z_t) \cdot h_{t-1} + z_t \cdot \tilde{h}_t. \quad (9)$$

Here, σ stands for the sigmoid activation function, $tanh$ for the hyperbolic tangent activation function, W_z, W_r, and W_h respectively, for weight matrices, h_{t-1} for the prior hidden state, and x_t for the input at time step *t*. The update gate z_t regulates how much of the preceding hidden state and candidate activation are combined to create the new hidden state h_t, while the reset gate r_t regulates which portions of the previous hidden state are disregarded when calculating the candidate activation.

To capture context-rich sequences in both directions (forward as well as backward), the model applies a BiGRU layer to the result of the convolutional layer. The outputs of the BiGRU for forward and backward directions are represented by Eqs. (3) and (4), respectively. The contextual representation of the input text is combined with information from both directions in the BiGRU output. This representation based on BiGRU combines the forward (h_f) and backward (h_b) hidden states, aligning them to reconstruct comprehensive, contextually-informed sequences. Eq. (5) ultimately expresses the combined contextual sequence as the final hidden state h_t, which is then passed to the capsule network layer.

3.4.4 Capsule Layer

Furthermore, the precision of DFU detection is ameliorated by using the capsule layer. Traditional CNNs lack distinct feature extraction and lose essential information due to pooling methods. Capsule networks address this by capturing linguistically supplemented features that consider order and local context. They excel in tasks like classification and retrieval, outperforming CNNs. Capsules comprise smaller capsules and use vectors to convey classification likelihood and aspect orientation, resulting in more efficient representation. Unlike CNNs, capsules generate vectors instead of scalars, and the dynamic routing algorithm adjusts weights, enhancing feature extraction.

The Capsule Network (CapsNet) is handling hierarchical relationships and spatial hierarchies in data. Capsules are a fundamental building block in CapsNets, and they aim to represent specific patterns or entities in a more robust manner. The key equations for Capsules and routing by-agreement are given. Capsule C_i in layer l is defined as a vector s_i representing the instantiation parameters of an entity, such as the pose (position, orientation, etc.). In order to guarantee that the length of s_i is from 0 to 1, it also includes a squash function connected to it.

$$s_i = W_i . x_i, \quad (10)$$

$$v_i = squash(s_i). \quad (11)$$

Routing-by-Agreement (Dynamic Routing). Information is sent from capsules to higher-level capsules via dynamic routing to guarantee agreement among each capsule on the presence of entities.

$$c_{ij} = \frac{\exp(b_{ij})}{\sum_k \exp(b_{ik})} (Softmax\ to\ compute\ routing\ weights), \quad (12)$$

$$s_j = \sum_i c_{ij} . u_{j|i}\ (Weighted\ sum\ of\ predictions), \quad (13)$$

$$v_j = squash(s_j)\ (Squash\ function), \quad (14)$$

where C_i represents a capsule in the lower layer, v_i is the output of that capsule, W_i represents the weight matrix, x_i is the input to the capsule, c_{ij} are the routing weights, b_{ij} are the log prior probabilities, $u_{j|i}$ is the prediction vector from capsule *i* to capsule *j*, v_j is the output of the higher-level capsule.

CapsNets are designed to hierarchical dealings by routing information between capsules and have shown promise in various tasks, especially in vision.

Our model utilizes capsules to improve DFU classification. As a result, the final hidden state, denoting the output from the BiGRU layer, is transmitted to the capsule network layer. Dynamic routing is used for building coupling coefficients c_{ij}, which enables the model to ignore trivial terms in the input data. The importance of DFU-related characteristics is correlated with the coupling coefficient c_{ij}; larger values denote more significant features, which capture crucial semantic representations of ROI in several directions. The coupling coefficient c_{ij} is computed using the Softmax function, and the output of the capsule, s_j, is obtained by adding up all of the prediction vectors.

3.4.5 Dense and Output Layers

The connected layer receives the output vector, v_j, from capsule network. Finally, using the sigmoid function on the dense layer output, the output layer classifies DFU. The proposed CBGRU-CapNet model utilizes a binary cross-entropy loss function.

3.5 Fine-Tuning of CBGRU-CapNet Using mDTMA

Here, a metaheuristic manifold algorithm [41] is used to fine-tune the CBGRU-CapNet parameters. Its architecture is comparable to the classic metaheuristic approach on Euclidean space, but its implementation of operations is different. Each operation must be performed in this implementation to maintain the individual's position on the manifolds. To be more precise, mDTMA, the anticipated algorithm, begins with a randomly generated population, with each person drawn from the manifold's points. The next step is to optimise the population towards better locations using a unique manifold version of the update and selection technique. This can be done by selecting individuals with superior fitness values for a goal, which is a real-valued function. Throughout the process, the population stays on the manifold, ensuring that workable answers will be found. In an effort to make the proposed algorithm more accessible, we have tried to name it after well-known operations from classic metaheuristic algorithms, like selection; however, these operations have been modified or improved to meet the needs of manifolds.

Optimisation issues involving real-valued functions on matrix manifolds are taken into account in this study. Each element of the solution set can be naturally characterized in matrix arrays, since the set is located on a matrix manifold. This allows for the progress of differential geometry through the formulation and control of matrix algebra. Positive definite simplex manifolds, Grassman manifolds, oblique spaces, product manifolds (rotation with Euclidean space), sphere manifolds, and matrices manifolds are among the manifolds to which our approaches are applied.

3.5.1 Main Framework of mDTMA

Initialising the population, selecting the tournament, crossing, mutation of polynomials, directional retraction are the fundamental steps of mDTMA. Directional are valuable because they facilitate mutation and crossover processes.

Let us consider a real valued function $f : \mathcal{M} \rightarrow R$, in which a submanifold $R^{n\times p}$ is embedded, which is explicit limit set in $R^{n\times p}$. For instance, Stiefel manifold $St(p,n)$ is a collection of submanifold of $R^{n\times p}$. In particular, when $p = 1$, it sphere; when $p = n$, it is orthogonal matrices $O(n)$.

Let us represent $\mathcal{P}^t \in \mathcal{M}$ as size of N population for t group, $\mathcal{M}$ as manifold, and $\mathcal{P}^0$ as the initial population, where $\mathcal{P}^0$is defined as $\mathcal{P}^0 = \{x_1^0, x_2^0, \dots, x_N^0\}$, each agent (chromosome) x_n^0 appears as $x_n^0 \rightarrow \{R^{a\times b}, \dots, R^{c\times d}\}_k$ and k denotes the element and the issue definition determines the sum of matrices (or solution components) in each agent.

The mDTMA procedure starts populace $\mathcal{P}^0$, where each agent is located on the manifold *M*. Initially, we used tournament selection to return the indices of N solutions by 2-tournament preference (random list of candidates) based on their fitness values. The contender with the lowest fitness score $f(x)$ will be designated selection. When compared to the conventional crossover, in which the offspring are not given any parental figures, this crossover is unique. The two parents only partially operate in this stage; the remaining work is done by the retraction step, which to call pre-crossover. In order to increase the variety of targets and their exploitability, mutation is a necessary step. Two goals from the previous two steps will typically depart from the manifold in the case that the provided *M* is not linear, according to directional transport. The directed transport operator takes the objective and breaks it down into its component parts, keeping the direction and length. In contrast to parallel transport, which uses a geodesic path while maintaining length but not angle, this operator for directional transport takes both into account. Finally, two possibilities are generated by this stage. Using the retraction are retrieved to the manifold as a new generation.

3.5.2 Implementation of Operators

In agents, pre-cross over as $x_i^t, x_j^t \rightarrow \{R^{a\times b}, \dots, R^{c\times d}\}_k$, k iterations are executed, one element at a time. Let us assume that for every element in both agents, parent1 and parent2, *beta* are randomly generated where *beta* belongs to $R^{c\times d}$, $beta \in R^{c\times d}$

To obtain the goal, let us compute the crossover using beta generated d-length vector:

$$v_1(target1) \leftarrow \left[beta \odot (Q_1^t - Q_2^t) \odot \frac{1}{2}\right], \qquad (15)$$

$$v_2(target2) \leftarrow \left[-beta \odot (Q_1^t - Q_2^t) \odot \frac{1}{2}\right], \qquad (16)$$

where $\odot$ achieves element-wise multiplication.

If you want to describe an isometric movement along a curve from tangent vectors to tangent space, you can use parallel transports. Assuming the existence and uniqueness of a minimising geodesic, the curve is typically left out when to implicitly use it as the curve. On the other hand, solving equation is often expensive, and it is necessary for algorithmic parallel transport. Consequently, the pragmatic option pertains to computationally efficient methodologies, transporters, and a stand-in for equivalent transport that is not constrained by connection limitations.

Orthogonal projector: Orthogonal map $Proj_x : \varepsilon \rightarrow \varepsilon$ categorized by three properties: (1) $im(Proj_x) = T_x\mathcal{M}$; (2) $Proj_x \circ Proj_x = Proj_x$; (3) $\langle u - Proj_x(u), v\rangle = 0$ for all $v \in T_x\mathcal{M}$ and $u \in \varepsilon$.

Denote $Proj_x$ the orthogonal from ε to $T_x\mathcal{M}$ and deduce $Proj_x^{\perp} = \boldsymbol{Id} - Proj_x$ which is the orthogonal projector to $N_x\mathcal{M}$. Operation $T_x\mathcal{M}$ and $N_x\mathcal{M}$ are complements in ε, where $\boldsymbol{Id}$ represents the individuality operator.

To calculate the orthogonal projection, we need to choose a Riemannian manifold, M. $T_x\mathcal{M} = \{v \in R^d : x^T v = 0\}$ is well defined by orthogonal accompaniment of x in Euclidean space ε. Therefore, any component aligned with x is removed by embedded manifold (the underlying Euclidean metric) to a tangent space of x. Evaluate the orthogonal projectors' features, for all $u, v \in \varepsilon$, there is $\langle Proj_x(u), v - Proj_x(v) = 0\rangle$. This movement $z \in \varepsilon$ from a point $x \in \mathcal{M}$ how tangential and normal motion are composed. To suggest an operator, a directed transport operator, move the candidate isometrically informally.

Let us equip $\mathcal{M}$ with metric. Then, there is a map $\eta_x(z) \cdot \eta_x(z) : \mathcal{E} \to T_x\mathcal{M} \oplus N_x\mathcal{M} \to (Proj_x(z) \oplus Proj_x^{\perp}(z))$, where $\eta_x(z)$ describes the procedure to $z \in \mathcal{E}$ of the point x on the manifold $\mathcal{M}$.

Given $Proj_x : \mathcal{E} \to T_x\mathcal{M} : u \mapsto Proj_x(u) = u - \langle x, u\rangle x$, which describes a map of $T\,\mathcal{M} \to \mathcal{M}$. The projection $Proj_x^{\perp}$ constituent satisfies $R_x(Proj_x^{\perp}) = x$, only when manifold $\mathcal{M}^{d-1}$ with embedded space R^d, for the purpose of defining retraction. There are some areas where to could theoretically avoid doing the same searches twice.

4 Results and Discussion

Our work is executed on Dell computers powered by Intel Core i7 processors. The utilisation of a robust GPU, and more especially the P100 prototypical, was critical in expediting the training process. The computational load was efficiently handled by our system with 32 GB of system RAM, thanks to the Kaggle kernel environment. Python 3.9.7 and TensorFlow 2.12.0 were used for the implementation. Table 1 discusses the training parameters.

Table 1 Training hyperparameters.

Setting	Hyperparameter
32	Batch size
90%	Training besides validation split
10%	Test split
mDTMA	Optimizer
150 × 150 px	Input size
Categorical cross-entropy	Loss function
0.001	Learning rate
10	Epochs

4.1 Validation Study of Proposed Model on First Dataset

Fig. 2 provides the graphical description of projected model on different training and testing data ratio on first dataset in terms of different metrics. The comparative analysis of various algorithms in terms of different metrics is shown in Fig. 3.

The proposed model demonstrates notable improvements in performance as the training data percentage increases. With a 60−40 training-testing split, the model accomplishes an accuracy of 69.04%, sensitivity of 90%, specificity of 43%, and an F_1-score of 73%, indicating a moderate detection.

The comparative analysis across various models (Table 2) highlights that the projected model achieves the highest presentation in all metrics, with an accuracy of 97.38%, F_1-score of 97.05%, precision of 96.98%, besides recall of 97.12%. ResNet50 performs notably well with an accuracy of 95.02% and F_1-score of 95.61%, proving its capability in handling complex classification tasks. CNN-LSTM and CNN also maintain competitive performance, achieving accuracies of 94.36% and 94.12%, respectively. Meanwhile, SVM, and XGBoost yield slightly lower metrics but still demonstrate strong classification accuracy and robustness. Overall, the proposed model's superior metrics underscore its effectiveness and reliability for the intended application, outperforming other approaches in the study.

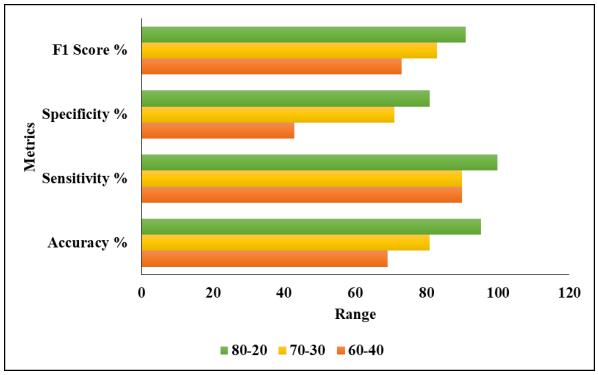

Fig. 2 Visual study of proposed prototypical on first dataset.

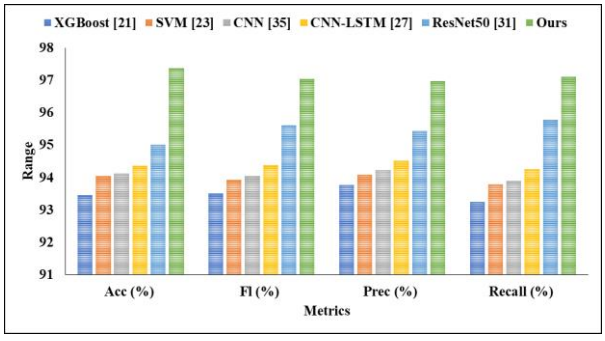

Fig. 3 Graphical description of projected model.

Table 2 Validation analysis of the desired model using existing techniques.

Methods	Accuracy (%)	F_1-score (%)	Precision (%)	Recall (%)
XGBoost [21]	93.46	93.51	93.78	93.25
SVM [23]	94.06	93.94	94.09	93.79
CNN [35]	94.12	94.05	94.22	93.89
CNN-LSTM [27]	94.36	94.39	94.53	94.26
ResNet50 [31]	95.02	95.61	95.44	95.79
Ours	97.38	97.05	96.98	97.12

4.2 Validation Analysis of Proposed Classical on Second Dataset

Fig. 4 provides the graphical description of predictable model on second dataset by using different training ratio.

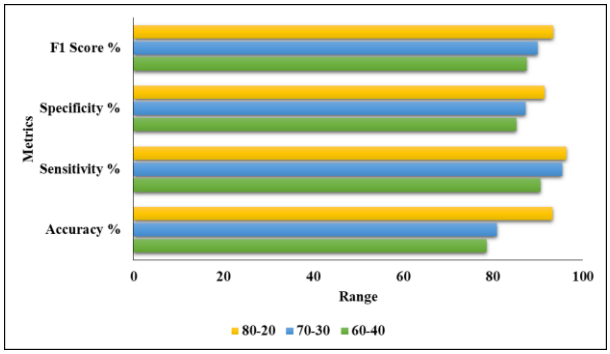

Fig. 4 Visual analysis of proposed model by modifying the second data training and testing ratio.

The proposed model exhibits a clear performance improvement with an increase in the training data percentage, reflecting the model's robustness with more training data. For the 60−40 training-testing split, the model accomplishes an accuracy of 78.57%, sensitivity of 90.65%, specificity of 85.23%, besides an F_1-score of 87.58%, showing solid performance. With a 70−30 split, the model's accuracy rises to 80.95%, sensitivity to 95.52%, specificity to 87.25%, and F_1-score to 89.95%, indicating enhanced detection capabilities. The highest metrics are observed with an 80−20 split, where the model achieves 93.26% accuracy, 96.36% sensitivity, 91.52% specificity, and an F_1-score of 93.36%. This upward trend underscores the model's improved reliability and effectiveness in correctly identifying relevant patterns as more data becomes available for training.

The comparative analysis of various methods highlights the superior performance of the proposed model and it is visually shown in Fig. 5. XGBoost achieves an accuracy of 82.99% score of 83.74%, precision of 81.03%, and recall of 85.45%. SVM performs slightly better with an accuracy of 84.14%, F_1-score of 85.43%, precision of 83.48%, and recall of 86.39%. CNN achieves 86.18% accuracy, F_1-score of 86.05%, precision of 86.31%, and recall of 89.79%. The CNN-LSTM model further improves, with an accuracy of 90.18%, F_1-score of 89.91%, precision of 89.69%, and recall of 90.13%. ResNet50 outperforms the previous methods, achieving an accuracy of 91.33%, F_1-score of 91.03%, precision of 90.48%, and recall of 91.59%. The proposed model surpasses all others, with an accuracy of 93.26%, F_1-score of 92.36%, precision of 91.29%, and recall of 93.46%, demonstrating its effectiveness and robustness in classification tasks.

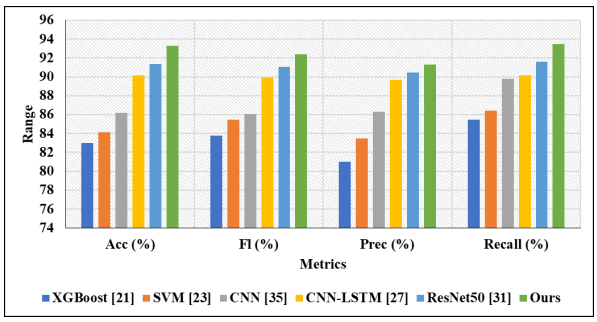

Fig. 5 Graphical comparison of projected model with existing works on second dataset.

4.3 Validation Analysis of Projected Model on Third Dataset

Figs. 6 and 7 presents the visual analysis of various models with proposed model on third dataset in terms of various metrics.

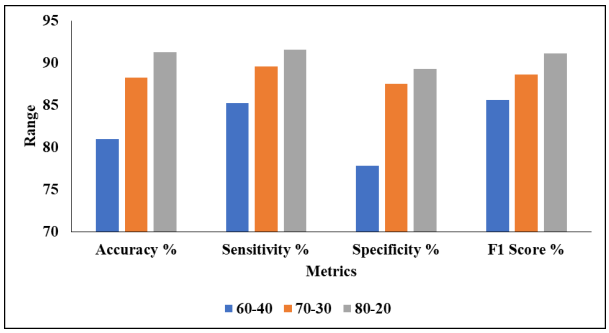

Fig. 6 Training and testing ratio of proposed model.

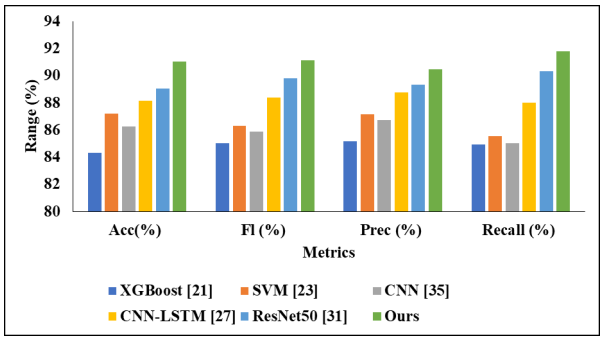

Fig. 7 Visual comparison of proposed model with existing techniques.

In this performance analysis of the proposed model across different training/testing splits, the accuracy, sensitivity, specificity, and F_1-score show a positive trend as the training percentage increases. With a 60−40 split, the model accomplishes an accuracy of 80.95%, sensitivity of 85.25%, specificity of 77.85%, besides an F_1-score of 85.64%. Moving to a 70−30 split, the accuracy rises to 88.25%, sensitivity to 89.58%, specificity to 87.52%, and F_1-score to 88.64%, indicating improved performance. With an 80−20 split, the model achieves the highest metrics, accuracy of 91.23%, sensitivity of 91.56%, specificity of 89.25%, and F_1-score of 91.12%. This analysis recommends that increasing the training data proportion enhances the model's predictive capabilities, particularly in achieving higher accuracy and F_1-scores.

This comparative analysis of classification methods highlights the improved performance of the proposed model ("Ours") over established techniques such as XGBoost, SVM, CNN, CNN-LSTM, and ResNet50. The proposed model attained the highest accuracy of 91.02%, F_1-score of 91.12%, precision of 90.46%, and recall of 91.79%. Among existing models, ResNet50 performed best, with 89.02% accuracy and a 90.31% recall, while CNN-LSTM achieved a strong F_1-score of 88.36%. The results designate that the proposed model consistently outperforms other methods across altogether metrics, demonstrating its efficacy and robustness for classification tasks, particularly in precision and recall metrics, which are crucial for balanced model performance.

4.4 Analysis of Our Model on Merged Data

Tables 3 and 4 delivers the comparative study of various models on merged data in terms of loss and accuracy.

In the analysis of model performance based on parameters, training accuracy, and loss, the proposed model exhibits exceptional results with a training accuracy of 100% and a minimal loss of 0.0021, indicating highly effective training. CNN-LSTM follows closely with an accuracy of 99.97% and a loss of 0.0027, and CNN achieves 99.91% accuracy with a slightly higher loss of 0.0042. SVM also demonstrates strong performance with 99.86% accuracy and 0.0057 loss, despite having significantly fewer parameters at 8.1M. In contrast, ResNet50, with the highest parameter count (86M), shows relatively lower training accuracy (74.84%) and a higher loss of 0.5115, suggesting overfitting or inefficiency with the given dataset. This analysis highlights the proposed model's balance between parameter efficiency and training effectiveness, outperforming in accuracy and loss across the evaluated models.

Table 3 Various Models' investigation on merged dataset.

Model name	Parameters	Training Accuracy	Loss
XGBoost [21]	61.9 M	0.8763	0.3233
SVM [23]	8.1 M	0.9986	0.0057
CNN [35]	5.3 M	0.9991	0.0042
CNN-LSTM [27]	11.2 M	0.9997	0.0027
ResNet50 [31]	86 M	0.7484	0.5115
Ours	22.9 M	1.0000	0.0021

Table 4 Comparative analysis of different models on merged data.

Model name	Accuracy %	Precision %	Recall %	F_1-score %
XGBoost [21]	87	87	89	87
SVM [23]	89	88	90	89
CNN [35]	91	90	92	92
CNN-LSTM [27]	93	92	93	93
ResNet50 [31]	95	94	94	94
Ours	96	96	96	96

In the comparative analysis of different models on merged data, the projected model outperforms other approaches with score of 96% across all metrics. ResNet50 follows closely, achieving 95% accuracy, and F_1-score, while CNN-LSTM achieves 93% across these metrics, indicating strong performance but slightly lower than ResNet50 and the proposed model. CNN achieves a balanced 91% in all categories, with SVM and XGBoost at 89% and 87%, respectively. The results demonstrate that the proposed model provides superior predictive power and robustness, indicating its suitability for tasks requiring high accuracy besides consistency across multiple estimation metrics.

5 Conclusion

To enhance classification accuracy and minimize decision variability and data misinterpretation, it is crucial to identify a subset of significant variables for early thermographic detection of foot illnesses. However, due to the imbalance in datasets concerning diabetic cases and inconsistencies among thermograms, defining an optimal subset of features remains challenging. In this study, lightweight deep neural network models for DFU recognition were developed, integrating convolutional networks, BiGRU, and capsule networks. The proposed approach effectively captures features from multiple orientations through the capsule network, while the CNN layer identifies semantic and contextual details. Additionally, the BiGRU model leverages its hidden layers to extract valuable patterns from contextual information. To further refine classification performance, the mDTMA was employed for optimization. The results indicate that the proposed model improves classification accuracy and efficiently localizes ulcers, thereby increasing diagnostic confidence among medical professionals. Beyond enhancing patient care and outcomes, this approach can aid in accurately diagnosing foot ulcers caused by conditions beyond diabetes.

Future research could focus on developing a multiscale transfer learning framework that integrates multiple pre-trained CNN models to enhance efficiency and classification accuracy. This approach aims to reduce training time while improving specificity in skin lesion categorization. Furthermore, expanding the proposed framework to detect a broader range of dermatological diseases and various skin lesions can be explored by leveraging multiple pre-trained CNN networks instead of a single model. Additionally, incorporating explainable AI techniques could provide deeper interpretability, aiding clinicians in decision-making and ensuring reliable diagnosis.

6 Data Availability Statement

The INAOE and STANDUP datasets employed in this study are publicly obtainable at https://ieee-dataport.org/open-access/plantar-thermogram-databasestudy-diabetic-foot-complications and https://www.standupproject.eu/manager/?conf=default&route=/STANDUP_Database (accessed 2 Feb 2024), respectively. The local dataset has been made available at https://www.iac.es/en/science-and-technology/technology-transfer-iactec/forms (accessed 17 Feb 2024). Notice that, due to privacy restrictions, only 19 of the 22 images sets originally acquired were released.

Disclosures

Author declares there is no conflict of interest.

References

1. J. Reyes-Luévano, J. A. Guerrero-Viramontes, J. Rubén Romo-Andrade, and M. Funes-Gallanzi, "DFU_VIRNet: A novel Visible-InfraRed CNN to improve diabetic foot ulcer classification and early detection of ulcer risk zones," Biomedical Signal Processing and Control 86, 105341 (2023).
2. S. K. Das, P. Roy, P. Singh, M. Diwakar, V. Singh, A. Maurya, S. Kumar, S. Kadry, and J. Kim, "Diabetic foot ulcer identification: A review," Diagnostics 13(12), 1998 (2023).
3. A. Alqahtani, S. Alsubai, M. P. Rahamathulla, A. Gumaei, M. Sha, Y.-D. Zhang, and M. A. Khan, "Empowering foot health: Harnessing the adaptive weighted sub-gradient convolutional neural network for diabetic foot ulcer classification," Diagnostics 13(17), 2831 (2023).
4. D. G. Armstrong, T.-W. Tan, A. J. M. Boulton, and S. A. Bus, "Diabetic foot ulcers: A review," JAMA 330(1), 62 (2023).
5. S. Nagaraju, K. V. Kumar, B. P. Rani, E. L. Lydia, M. K. Ishak, I. Filali, F. K. Karim, and S. M. Mostafa, "Automated diabetic foot ulcer detection and classification using deep learning," IEEE Access 11, 127578–127588 (2023)..
6. C. Cao, Y. Qiu, Z. Wang, J. Ou, J. Wang, A. H. Hounye, M. Hou, Q. Zhou, and J. Zhang, "Nested segmentation and multi-level classification of diabetic foot ulcer based on mask R-CNN," Multimedia Tools and Applications 82(12), 18887–18906 (2023).
7. O. Niță, L. I. Arhire, L. Mihalache, A. D. Popa, G. Niță, A. Gherasim, and M. Graur, "Evaluating classification systems of diabetic foot ulcer severity: A 12-Year retrospective study on factors impacting survival," Healthcare 11(14), 2077 (2023).
8. E. Tamir, O. Rabau, Y. Beer, Y. Smorgick, H. Kaufman, and A. S. Finestone, "A novel classification for diabetic foot ulcers of the first ray," Advances in Skin & Wound Care 36(1), 30–34 (2023).

9. M. S. A. Toofanee, M. Hamroun, S. Dowlut, K. Tamine, V. Petit, A. K. Duong, and D. Sauveron, "Federated learning: Centralized and P2P for a siamese deep learning model for diabetes foot ulcer classification," Applied Sciences 13(23), 12776 (2023).
10. P. N. Thotad, G. R. Bharamagoudar, and B. S. Anami, "Diabetic foot ulcer detection using deep learning approaches," Sensors International 4, 100210 (2023).
11. S. K. Das, S. Namasudra, A. Kumar, and N. R. Moparthi, "AESPNet: Attention enhanced stacked parallel network to improve automatic diabetic foot ulcer identification," Image and Vision Computing 138, 104809 (2023).
12. B. Gunapriya, T. Rajesh, A. Thirumalraj, and B. Manjunatha, "LW-CNN-based extraction with optimized encoder-decoder model for detection of diabetic retinopathy," Journal of Autonomous Intelligence 7(3), (2023).
13. M. H. Alshayeji, S. ChandraBhasi Sindhu, and S. Abed, "Early detection of diabetic foot ulcers from thermal images using the bag of features technique," Biomedical Signal Processing and Control 79, 104143 (2023).
14. K. McDermott, M. Fang, A. J. M. Boulton, E. Selvin, and C. W. Hicks, "Etiology, epidemiology, and disparities in the burden of diabetic foot ulcers," Diabetes Care 46(1), 209–221 (2023).
15. V. Sathya Preiya, V. D. A. Kumar, "Deep learning-based classification and feature extraction for predicting pathogenesis of foot ulcers in patients with diabetes," Diagnostics 13(12), 1983 (2023).
16. A. Hernandez-Guedes, N. Arteaga-Marrero, E. Villa, G. M. Callico, and J. Ruiz-Alzola, "Feature ranking by variational dropout for classification using thermograms from diabetic foot ulcers," Sensors 23(2), 757 (2023).
17. I. Khosa, A. Raza, M. Anjum, W. Ahmad, and S. Shahab, "Automatic diabetic foot ulcer recognition using multi-level thermographic image data," Diagnostics 13(16), 2637 (2023).
18. S. Biswas, R. Mostafiz, B. K. Paul, K. M. Mohi Uddin, M. M. Rahman, and 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, 100128 (2023).
19. M. Bundó, B. Vlacho, J. Llussà, I. Bobé, M. Aivar, C. Ciria, A. Martínez-Sánchez, J. Real, M. Mata-Cases, X. Cos, M. Dòria, J. Viade, J. Franch-Nadal, and D. Mauricio, "Prediction of outcomes in subjects with type 2 diabetes and diabetic foot ulcers in Catalonian primary care centers: a multicenter observational study," Journal of Foot and Ankle Research 16(1), 8 (2023).
20. T. Lan, Z. Li, and J. Chen, "FusionSegNet: Fusing global foot features and local wound features to diagnose diabetic foot," Computers in Biology and Medicine 152, 106456 (2023).
21. S. Özgür, S. Mum, H. Benzer, M. K. Toran, and İ. Toygar, "A machine learning approach to predict foot care self-management in older adults with diabetes," Diabetology & Metabolic Syndrome 16(1), 244 (2024).
22. Y. Chen, Y. Zhang, M. Jiang, H. Ma, and Y. Cai, "HMOX1 as a therapeutic target associated with diabetic foot ulcers based on single-cell analysis and machine learning," International Wound Journal 21(3), e14815 (2024).
23. S. Hong, Y. Chen, Y. Lin, X. Xie, G. Chen, H. Xie, and W. Lu, "Personalized prediction of diabetic foot ulcer recurrence in elderly individuals using machine learning paradigms," Technology and Health Care 32, 265–276 (2024).
24. M. Saeedi, H. T. Gorji, F. Vasefi, and K. Tavakolian, "Federated versus central machine learning on diabetic foot ulcer images: Comparative simulations," IEEE Access 12, 58960–58971 (2024).
25. G. Verma, "Leveraging smart image processing techniques for early detection of foot ulcers using a deep learning network," Polish Journal of Radiology 89, 368–377 (2024).
26. C. Wu, C. Xu, S. Ou, X. Wu, J. Guo, Y. Qi, and S. Cai, "A novel approach for diabetic foot diagnosis: Deep learning-based detection of lower extremity arterial stenosis," Diabetes Research and Clinical Practice 207, 111032 (2024).
27. L. Z. Chee, S. Sivakumar, K. H. Lim, and A. A. Gopalai, "Gait acceleration-based diabetes detection using hybrid deep learning," Biomedical Signal Processing and Control 92, 105998 (2024)..
28. R. Castillo-Morquecho, E. Guevara, J. L. Ramirez-GarciaLuna, M. A. Martínez-Jiménez, M. G. Medina-Rangel, and E. S. Kolosovas-Machuca, "Digital infrared thermography and machine learning for diabetic foot assessment: thermal patterns and classification," Journal of Diabetes & Metabolic Disorders 23(2), 1967–1976 (2024).
29. D. K. Agrawal, W. Jongpinit, S. Pojprapai, W. Usaha, P. Wattanapan, P. Tangkanjanavelukul, and T. Vitoonpong, "Smart insole-based plantar pressure analysis for healthy and diabetic feet classification: statistical vs. machine learning approaches," Technologies 12(11), 231 (2024).
30. H. Shi, X. Yuan, X. Yang, R. Huang, W. Fan, and G. Liu, "A novel diabetic foot ulcer diagnostic model: identification and analysis of genes related to glutamine metabolism and immune infiltration," BMC Genomics 25(1), 125 (2024).
31. A. M. El-Kady, M. M. Abbassy, H. Hamdy, A. Farid, A. Moussa, "Advancing diabetic foot ulcer detection based on resnet and gan integration," Journal of Theoretical and Applied Information Technology 102(6), 2258–2268 (2024).
32. S. A. Sulayman, A. R. Qishqish, and G. F. Dayhoom, "Early detection of diabetic foot using thermal imaging and deep learning techniques," African Journal of Advanced Pure and Applied Sciences, 564–567 (2024).
33. A. Sridhar, R. Balasubramanian, J. A. Krishnaswamy, G. Radhakrishnan, and B. Pesala, "Deep-learning aided multispectral autofluorescence imaging for rapid bioburden detection and gram-type classification on diabetic foot ulcers," Optics and Biophotonics in Low-Resource Settings X 14 (2024).

34. M. Cakir, G. Tulum, F. Cuce, K. B. Yilmaz, A. Aralasmak, M. İ. Isik, and H. Canbolat, "Differential diagnosis of diabetic foot osteomyelitis and charcot neuropathic osteoarthropathy with deep learning methods," Journal of Imaging Informatics in Medicine 37(5), 2454–2465 (2024).
35. S. Murugan, N. Mohankumar, "Smart foot monitoring: A cutting-edge solution for early detection of diabetic complications using CNN model," International Journal of Advances in Signal and Image Sciences 10(1), 45–55.
36. D. Bouallal, A. Bougrine, R. Harba, R. Canals, H. Douzi, L. Vilcahuaman, and H. Arbanil, "STANDUP database of plantar foot thermal and RGB images for early ulcer detection," Open Research Europe 2, 77 (2022).
37. D. A. Hernandez-Contreras, H. Peregrina-Barreto, J. de J. Rangel-Magdaleno, and F. J. Renero-Carrillo, "Plantar thermogram database for the study of diabetic foot complications," IEEE Access 7, 161296–161307 (2019).
38. N. Arteaga-Marrero, L. C. Bodson, A. Hernández, E. Villa, and J. Ruiz-Alzola, "Morphological foot model for temperature pattern analysis proposed for diabetic foot disorders," Applied Sciences 11(16), 7396 (2021).
39. E. Villa, N. Arteaga-Marrero, and J. Ruiz-Alzola, "Performance assessment of low-cost thermal cameras for medical applications," Sensors 20(5), 1321 (2020).
40. R. Aluvalu, T. Sharma, U. M. Viswanadhula, A. D. Thirumalraj, M. V. V. P. Kantipudi, and S. Mudrakola, "Komodo dragon mlipir algorithm-based CNN model for detection of illegal tree cutting in smart IoT forest area," Recent Advances in Computer Science and Communications 17(6), 1–12 (2024).
41. V. Snášel, L. Kong, and S. Das, "From constraints fusion to manifold optimization: A new directional transport manifold metaheuristic algorithm," Information Fusion 113, 102596 (2025).