

Applications of Machine Learning in Diabetic Foot Ulcer Diagnosis using Multimodal Images: A Review

Veena Mayya, Venkat Tummala, Chaduvula Upendra Reddy, Pranjal Mishra, Rajasekhar Boddu, Diana Olivia*, Sowmya Kamath S

Abstract—Diabetes related complications such as Diabetic Foot Ulcers (DFU) may necessitate recurrent hospitalisations and expensive treatments. Uncontrolled diabetes can result in severe DFUs, resulting in amputation of lower limbs or feet, prolonged debilitation and diminished quality of life. Early diagnosis and proactive management are reported to significantly enhance the prognosis and reduce the onset of further complications. In this study, research works on developing clinical decision support systems (CDSS) for the identification and segmentation of DFU are systematically reviewed. The techniques employed range from traditional image processing techniques to approaches based on deep learning (DL). A taxonomy of DFU CDSSs is presented, categorised into two groups: RGB-based techniques and thermal imaging-based approaches. To the best of our knowledge, this is the first attempt at a comprehensive study of CDSSs for DFU related investigative tasks, based on different imaging modalities. We also delve into the difficulties experienced in the process of creating efficient, reliable, and accurate models for the early detection of DFU, and highlight the vast potential for further research in this emerging domain.

Index Terms—Diabetes related complications, Thermography, Image processing, Clinical Decision Support Systems, Artificial intelligence

I. INTRODUCTION

AS per recent International Diabetes Federation reports [1], over 463 million people globally have been affected by diabetes, which is predicted to reach 700 million, by 2045

Manuscript received September 23, 2022; revised March 13, 2023.

Veena Mayya is a Senior Assistant Professor at the Department of Information & Communication Technology, at Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, Karnataka, India. (Email: veena.mayya@manipal.edu).

Venkat Tummala is an undergraduate student in the Department of Computer Science and Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, Karnataka, India. (Email: venkat.tummala@learner.manipal.edu).

Chaduvula Upendra Reddy is an undergraduate student in the Department of Information & Communication Technology, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, Karnataka, India. (Email: chaduvula.reddy@learner.manipal.edu).

Pranjal Mishra is an undergraduate student in the Department of Computer Science and Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, Karnataka, India. (Email: pranjal.mishra2@learner.manipal.edu).

Rajasekhar Boddu is a lecturer at the Department of Software Engineering, College of Computing and Informatics, Haramaya University, Dire Dawa, Ethiopia. (Email: Rajsekhar.Boddu@haramaya.edu.et).

Diana Olivia is an Associate Professor at the Department of Information & Communication Technology, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, Karnataka, India. (Corresponding Author, Email: diana.olivia@manipal.edu)

Sowmya Kamath S is a Grade-1 Assistant Professor at the Department of Information Technology, National Institute of Technology Karnataka, Surathkal, Mangalore 575025, India. (Email: sowmyakamath@nitk.edu.in).

[1]. One in three diabetic patients experiences diabetic foot ulcers (DFU), which are 34% more likely to occur in people with diabetes over their lifetime [2]. The massive increase in DFU prevalence over the past few decades has posed a critical challenge to global healthcare systems. In more serious cases, DFU can result in frequent hospitalization, reduced quality of life, expensive rehabilitation therapy, disfigurement of the feet or lower limbs, or, in more extreme cases, even death. Thus, to enable effective and timely treatment and prevent foot or lower limb amputation, it is essential to recognise these complications as early as possible.

The condition of a patient suffering from DFU is exacerbated by certain associated clinical symptoms caused by an inadequate oxygen-rich blood supply due to a condition called ischaemia. The prevalent approaches adopted by medical practitioners in contemporary clinical practise for detecting ischaemia and infection are physical examination, blood testing, and Doppler examinations of leg blood arteries. Traditional DFU assessment techniques are therefore cost-prohibitive, expert-dependent, and time-consuming. Typically, DFUs are characterised by erratic structures and irregularly shaped exterior limits. Blisters, redness, callus formation, and substantial tissue types like bleeding, scaly skin, granulation, and slough are some of the visible characteristics of DFU and the skin around it that vary depending on the stage. Therefore, correct evaluation of these visual indicators like colour descriptors, temperature features, and texture features would form the basis of ulcer evaluation using computer vision algorithms [3, 4]. The emergence of artificial intelligence in medicine with technologies like machine learning, deep learning, and computer vision has helped make tremendous strides in the field of medical imaging in recent years. There is thus a significant potential for adopting such learning based approaches for automatic disease diagnosis [5–7].

Recently, automated diagnosis of diseases for improving care delivery has received significant research attention [8–20]. The evaluation of DFU has primarily employed different types of image datasets, namely, RGB and thermal image datasets. Some sample RGB and thermal images are shown in Fig. 1. For RGB image datasets, there are two versions: one is the wound DFU dataset, which contains normal and abnormal DFU wound images; the other is the infection and ischaemia dataset, which contains infection and non-infection images as well as ischaemia and non-ischaemia images. The thermal images provide foot thermograms that allow visualisation of the temperature variations of the foot. Complex algorithms analyse these changes and allow for

further research. Researchers have adopted a wide variety of techniques and models for the task of DFU diagnosis and management of associated conditions.

In this review, relevant research publications were selected based on two major criteria. Firstly, each of the selected articles incorporated a combination of advanced machine learning and deep learning approaches. Additionally, the research work chosen has been published within the last five to six years. This ensures that our review is up-to-date with the latest breakthroughs in DFU machine learning applications. A formal methodology was defined for the selection of research work for our comprehensive review as per the PRISMA guidelines [23], and is illustrated in Fig. 4. The inclusion criteria are determined by the PICO principle (population, intervention, comparison, and outcomes). The *study population* considered for the evaluation of DFU is characterised using RGB and thermal image datasets. The *interventions* considered for the study encompass the types of evaluations used, i.e., segmentation, detection, and classification. The *comparison* is performed as a measure of differentiation between the efficacy of thermal and RGB imaging modalities. Finally, the *outcomes* are measured by considering standard datasets and evaluation metrics. The exclusion criteria considered allowed for the pruning of articles with incomplete data on the main indicators and those without full-text access.

II. REVIEW OF CDSS FOR DFU DIAGNOSIS

Typically, identifying anomalies in DFU images can be categorised into three major tasks – (1) Segmentation (2) Detection and (3) Classification. Medical image segmentation has made extensive use of deep learning, and several articles describing its application in this field have been published. Deep learning models have been adapted to differentiate the region of interest (ROI) in images for further processing for diabetic wound segmentation [24].

Detection is a process that focuses on recognising and locating disease/lesions in images, utilising infrared thermography for ulcer detection and prevention in high-risk diabetic feet. The effectiveness of both deep learning and classic machine learning techniques is leveraged for classification and grading [25]. Although traditional machine learning techniques perform well, such approaches are particularly slow due to the many intermediate steps. Deep learning approaches have demonstrated their superiority through the use of object localization meta-architectures, which were used to train end-to-end models on DFU datasets to classify and localise ulcers on full foot images.

Image classification is a significant task in the disease diagnosis process and has been extensively explored in literature. In recent studies, techniques like Support Vector Machine (SVM), convolutional neural networks (CNNs), and K-nearest neighbours have been used, in addition to pretrained models like Alexnet, Resnet, and VGG16. Ensemble deep learning models have been adapted for binary classification of DFU images for tasks like infection vs. non-infection and ischaemia vs. non-ischaemia [3]. Custom-designed deep neural models have also been leveraged for specialized tasks [26]. Based on the data modalities, these works can be broadly categorized into two groups - thermal imaging based CDSS and RGB imaging based CDSS. A detailed discussion

on the various approaches in these two groups are discussed in this section. The graph shown in Fig. 2 details about the number of thermal and RGB-based publications for the past three years.

A. Thermal imaging based CDSS

Diagnostic thermal images are captured using infrared thermography (IRT), a quick, non-invasive, and non-contact approach for examining the temperature distribution of the foot and assessing any thermal changes [27]. Unlike other medical imaging techniques, IRT does not use harmful radiation because it only detects the thermal signatures emitted by the object under examination. Research studies in this domain aim to perform classification (binary and multi-class) and segmentation of diabetic foot ulcers. A range of machine learning models like Multilayer Perceptron (MLP), Support Vector Machine (SVM), logistic regression, etc, and deep learning algorithms like AlexNet, GoogleNet, ResNet, and MobileNet have been utilised for classification and segmentation. Figure 3 summarises the procedure utilised by the majority of the existing works. Commonly used preprocessing techniques are summarised in Figure 5.

One of the major challenges in designing a CDSS based on thermal imaging is the limited availability of publicly accessible thermal datasets. Data augmentation is employed to increase the size of the dataset in order to combat this [28]. Techniques like dataset augmentation with 5 fold cross-validation, oversampling, and various preprocessing approaches such as data annotation, alignment, and feature extraction have been used to prepare the dataset. The most extensively utilised thermal DFU dataset is the Plantar Thermogram Database [22], which contains thermograms of 122 diabetic and 45 non-diabetic subjects. The distribution of temperature in the plantar region can be studied using this dataset. Many research projects rely on proprietary databases, making it impossible to replicate their findings. Table I presents a summary of existing research works that utilized thermal images datasets for DFU diagnosis. Fig. 2 depicts the year-wise distributions of publications.

Muralidhara et al. [29] initially formulated a balanced dataset using the plantar thermogram database [22], then performed multi-class classification using the CNN network. Alshayegi et al. [30] extracted the features using Scale Invariant Feature Transform (SIFT) and Speeded Up Robust Features (SURF) from the plantar thermogram database [22]. These features are then trained using various machine learning models such as SVM, Random Forest, K-Nearest Neighbor (KNN), etc. The authors reported the best results with a combination of SURF features and a SVM classifier.

Filipe et al. [31] divided the foot into five different clusters using the K-means algorithm with different values of temperature and extracted the features from plantar thermogram images [22]. The authors then performed a binary classification and, only for diabetic cases, performed multi-class classification using SVM, weighted k-NN, logistic regression, etc. The authors achieved the best accuracy of 0.980 for binary classification and 0.966 for multi-class classification using the SVM quadratic model. Isaza et al. [32] proposed data augmentation through the change of the amplitude in the Fourier Transform and achieved 100%



Fig. 1: Sample DFU images A)RGB (Normal)[21], B)RGB (Abnormal)[21], C)Thermal (Normal)[22], D)Thermal (Abnormal)[22].

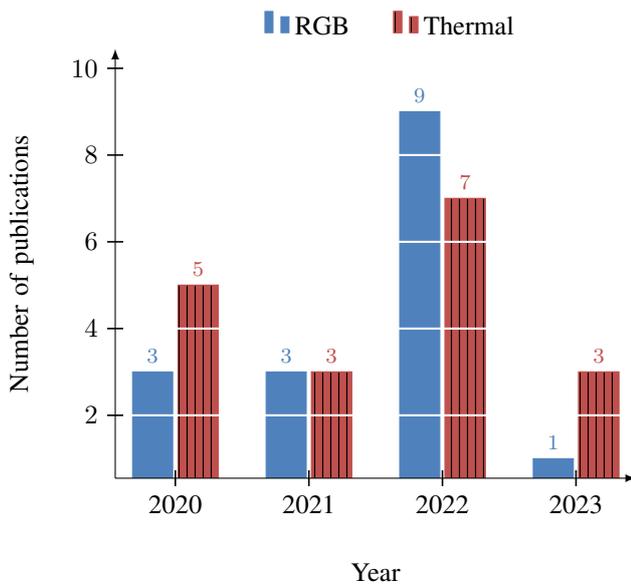


Fig. 2: Year-wise publications.

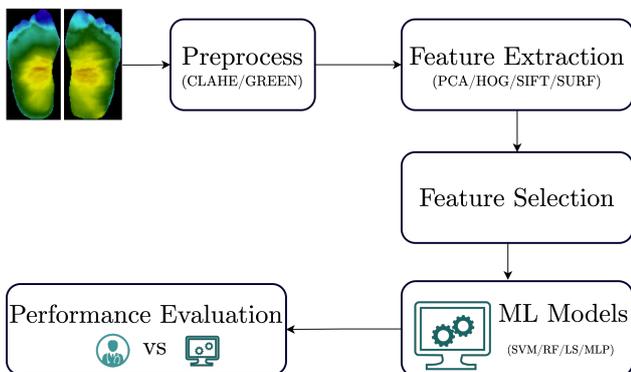


Fig. 3: ML based DFU detection systems.

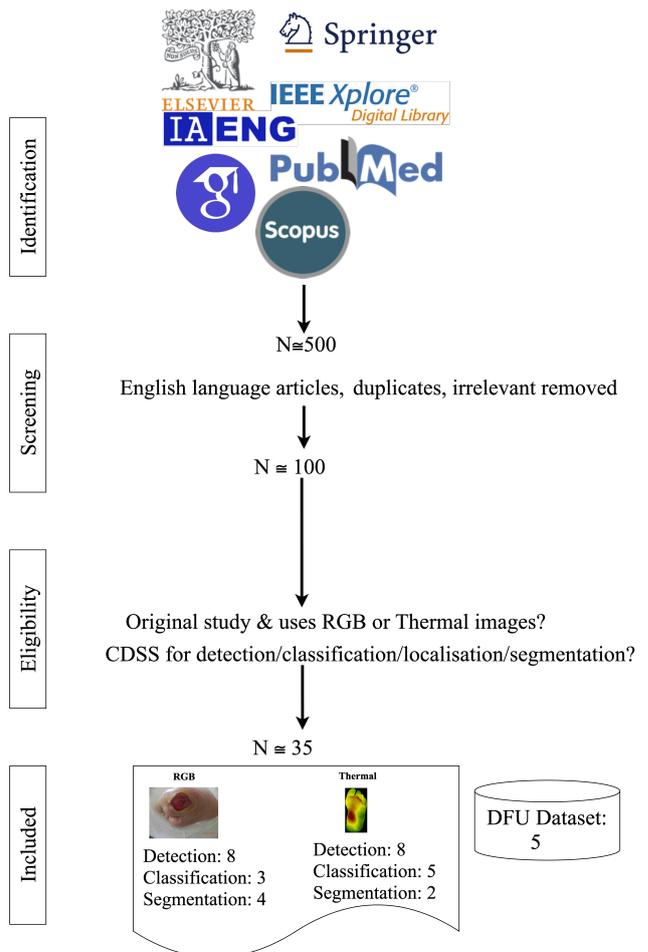


Fig. 4: Formal selection criteria (as per PRISMA Guidelines).

accuracy using ResNet50v2 for the thermogram dataset [22]. Munadi et al. [33] segmented the foot region and fused the

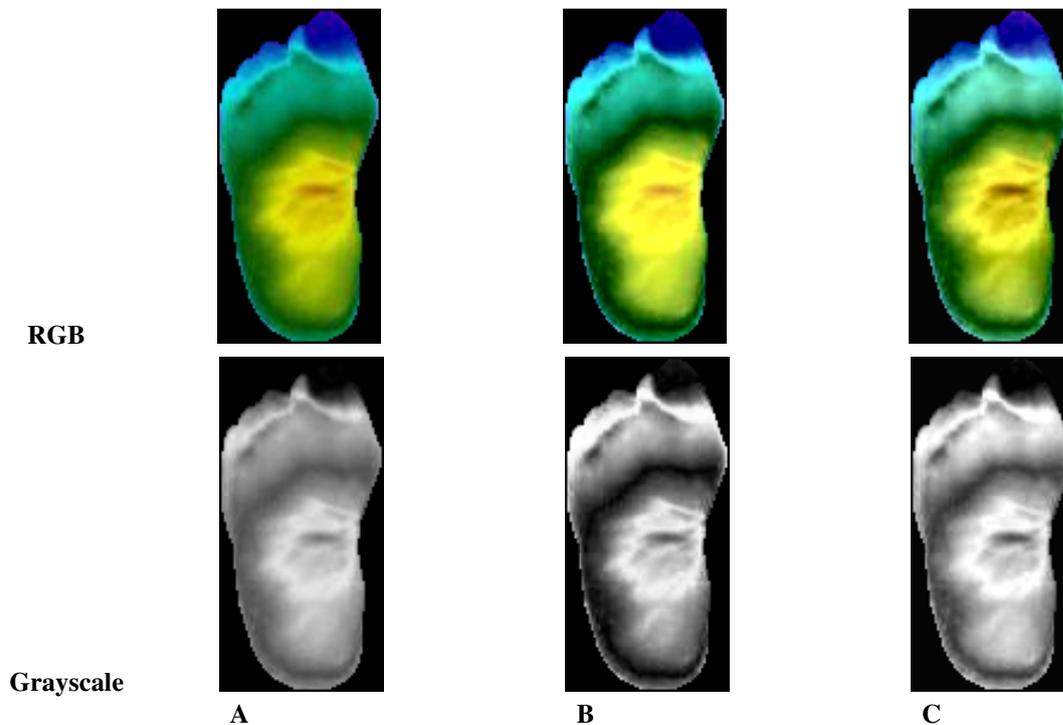


Fig. 5: Pre-processed images A)Original, B)Histogram equalization, C)Contrast limited adaptive histogram equalization.

predictions of MobileNetV2 and ShuffleNet to achieve 100% accuracy using the plantar thermogram database [22].

Hernandez-Guedes et al. [34] utilised pretrained convolutional AutoEncoder (AE) [35] without skip-connections and achieved an accuracy of 0.937 for classifying thermograms[22]. Khandakar et al. [36] explored different machine-learning (ML) approaches for classifying thermograms [22] with feature engineering and found that the classical ML classifier's exceptionally good compared to the performance of the 2D CNN models. Bouallal et al. [37] segmented the foot sole from the plantar foot images using a double encoder-ResUnet. The authors achieved an average intersection over union (IoU) of 97% due to the fusion of thermal and RGB colour information.

B. RGB imaging based CDSS

RGB images from existing DFU datasets are obtained in a variety of ways, including by direct capture using Android devices or Apple tablets at varying brightness levels and angles, while a few datasets provide images captured using professional cameras [3]. All such images are typically captured with the appropriate permission and ethical clearance from the patients.

Existing work mostly makes use of four publicly available datasets. The Diabetic Foot Ulcers Grand Challenge 2020 challenge (DFUC 2020) dataset [56] consists of a total of 4,000 images (with a 50% split for training and testing) along with extra 200 images for normality checks. The DFUC2021 dataset [53] provides pathology labels for 152 ulcers with ischaemia, 1703 ulcers with infection, and 372 ulcers with both conditions. With the usage of natural augmentation, 15,683 DFU patches were generated, with 5,955 training and 5,734 testing patches. The DFUC2022 dataset [60] is the largest image segmentation dataset in which, clinicians manually delineated ulcer areas. The FUSC (Foot ulcer

segmentation challenge) dataset [61] consists of 1,010 wound images, which are augmented to obtain 3,645 training images and 405 images.

Table II presents a summary of existing research works that utilized RGB images datasets for DFU diagnosis. These studies used a variety of preprocessing techniques, including the zero-centre procedure for producing patches followed by pixel normalisation of the datasets. Additionally, data augmentation techniques such as horizontal and vertical flips, as well as picture scaling, were used to enhance photos. Shades of Gray [62], a technique based on colour constancy, is used to compensate for noise and illumination issues caused by varied capturing equipment. All researchers pre-processed these datasets to make the images uniform in size. Several experiments, such as segmentation, classification, and detecting DFU, used these datasets. Figure 6 summarises the procedure utilised by the majority of the existing works. For DL based CDSS, the feature extraction and selection phase is unnecessary because these models can learn the features from the training data provided during the training phase.

Goyal et al. [5] proposed DFUNet for classifying the RGB foot image patches into normal and abnormal cases, and achieved an F-measure of 0.939 for 172 test patches. Venkatesan et al. [46] initially augmented the images using the Synthetic Minority Oversampling Technique (SMOTE) and classified them into normal and abnormal classes using NFU-Net. They achieved 100% accuracy for the binary classification of 150 test images. Das et al. [45] fused the ML based low level handcrafted features and CNN based high level features for DFU identification and achieved 0.9537 F1-score for 336 test images [5].

Goyal et al. [3] incorporated an ensemble CNN model for more effective recognition of ischaemia and infection in RGB DFU images. Das et al. [49] experimented with seven

TABLE I: Summary of thermal image based DFU CDSSs.

Work	Dataset/Sample size	Methodology/Remarks
Filipe et al. [38] (Binary classification with foot segmentation).	Public thermograms dataset [22] containing 122 diabetic and 45 non-diabetic patients.	K-means clustering implemented to divide the foot images into clusters and temperature is calculated for each cluster. The AUC achieved is 0.84 along with sensitivity of 0.73 and F-score of 0.81.
Saminathan et al. [39] (Binary classification).	Private dataset with 36 Diabetic and 24 Control patients.	Classification was done by implementing the SVM. Accuracy of 0.908 was obtained.
Khandakar et al. [28] (Binary classification)	Public thermograms dataset [22].	ResNet18, ResNet50, DenseNet201, InceptionV3, VGG19, and MobileNetV2 were employed as pre-trained CNNs. DenseNet201 and MobileNetV2 performed the best with 0.94 accuracy.
Munadi et al. [33] (Binary classification).	Public thermograms dataset [22].	Segmented the foot region and fused the predictions of MobileNetV2 and ShuffleNet to achieve accuracy of 1.
Isaza et al. [32] (Binary classification)	Public thermograms dataset [22].	Proposed data augmentation through the change of the amplitude in the Fourier Transform and achieved 100% accuracy using ResNet50v2.
Alshayegi et al. [30] (Binary classification)	Public thermograms dataset [22].	Features extracted using SIFT and SURF. Classified using traditional ML algorithms like SVM, RF, KNN, etc. and achieved accuracy of 0.978, precision of 0.979, sensitivity of 0.978, specificity of 0.978.
Abian Hernandez-Guedes et al. [34] (Binary classification)	Public thermograms dataset [22].	Utilised pretrained convolutional AutoEncoder (AE) [35] without skip-connections and achieved an accuracy of 0.937.
Hernandez et al. [40] (Binary classification).	Publicly available INAOE dataset. The Database was extended using SMOTE.	Features were extracted using LASSO and random forest approaches, Subsequently, the extracted features were employed to classify subjects using a support vector machine (SVM) model with an accuracy of 91%.
Puneeth et al. [41] (Binary classification).	Private dataset with 488 image patches. Augmentation is used to increase the number of images to 1688 patches.	Early detection and prognosis of diabetic foot ulcers using the EfficientNet with 98.97% accuracy is implemented.
Lan et al. [42] (Binary classification).	The dataset was gathered by Shanghai Municipal Eighth People's Hospital. Apart from DFU, all of the non-DF images had chronic wounds.	To distinguish between DF images and non-DF images with an accuracy of 95.78%, the FusionSegNet extracts global foot features and local wound features in the second stage.
Cruz-Vega et al. [26] (Multi-class classification)	Public thermograms dataset [22].	Data augmentation and models like MLP, SVM, AlexNet, and GoogleNet were implemented. A custom DL structure (DFTNet) was created, achieving a precision of 0.94 and accuracy of 0.945.
Maldonado et al. [43] (Classifier combined with segmentation to obtain a multi-class output)	Private database was built which comprised 249 images. Segmentation and annotation were done as part of the preprocessing.	Transfer Learning was utilised on a pre-trained Masked R-CNN model, with 95.61% accuracy, 96.5% sensitivity and 92.41% specificity.
Muralidhara et al. [29] (Multi-class classification)	Public thermograms dataset [22].	Formulated a balanced dataset, then performed multi-class classification using CNN network. Achieved a mean accuracy of 0.9827, mean sensitivity of 0.9684 and mean specificity of 0.9892.
Filipe et al. [31] (Multi-class classification)	Public thermograms dataset [22].	The authors performed a binary classification, and only for diabetic cases performed multi-class classification using SVM quadratic model, and achieved 0.966 for multi-class classification.
Khandakar et al. [36] (Multi-class classification)	Public thermograms dataset [22].	Explored different ML approaches with feature engineering, multilayer perceptron (MLP) showed an accuracy of 90.1% .
Prabhu et al. [44] (DFU segmentation).	Private dataset with 50 total plantar thermograms. 40 samples belong to the DM (Diabetes mellitus) group and 10 belong to the control group	Adaptive C-means algorithm is implemented and it is evaluated by using the Dice coefficient and the root mean square deviation which are 0.941 and 5.986 respectively.
Bouallal et al. [37] (DFU segmentation).	Private dataset with 54 healthy and 145 diabetic patients from the National Hospital Dos de Mayo, Peru.	Segmented the foot sole from the plantar foot images using double encoder-ResUnet. They achieved an average intersection over union (IoU) of 97% due to the fusion of thermal and RGB colour information.

TABLE II: Summary of RGB image based DFU CDSSs.

Work	Dataset/Sample size	Methodology/Remarks
Goyal et al. [5] (Binary classification of normal and abnormal)	Public dataset with 1423 training patches, 84 validation patches, and 172 test patches that were generated from the 397 original foot images.	A of important aspect of CNNs architecture - depth and parallel convolution layered DFUNet model was proposed. Achieved F-measure of 0.939 for binary classification.
Das et al. [45] (Binary classification of normal and abnormal)	Public dataset provided by Goyal et al. [5]	Fused the ML based low level handcrafted features and CNN based high level features for DFU identification and achieved 0.9537 F1-score for 336 test images.
Venkatesan et al. [46] (Binary classification)	Public dataset provided by Goyal et al. [5]	Initially augmented the images using Synthetic Minority Oversampling Technique (SMOTE) and classified into normal and abnormal classes using NFU-Net. The authors achieved 100% accuracy for the binary classification of 172 test images.
Das et al. [47] (Binary classification of normal and abnormal)	Public dataset provided by Goyal et al. [3]	DFU_SPNet, AlexNet, VGG16, DFUNet and DFU_QUTNet models were compared and DFU_SPNet outperformed most of the models with 0.954 F1 score for binary classification.
Alzubaidi et al. [48] (Binary classification)	Private dataset with 1609 skin patches. 542 are normal and 1067 are abnormal	It is structured around the concept of Directed Acyclic Graph (DAG). The proposed network exceeds several state-of-the-art models in terms of performance.
Goyal et al. [3] (Binary classification: ischaemia vs non-ischaemia and infection vs non-infection)	Public dataset of 1459 DFU images with 7136 patches for training, 1019 patches for validation, and 2038 patches (testing) from the 2611 original dataset	BayesNet (DL), Randomforest (DL), Multilayer perceptron (DL), InceptionV3 (CNN), ResNet50(CNN), InceptionResNetV2 (CNN), Ensemble (CNN) model performed the best with 0.902 for ischaemia binary classification and 0.722 for infection binary classification.
Das et al. [49] (Binary classification of ischaemia and infection in DFU)	Public dataset provided by Goyal et al. [3]	Seven variants of the proposed architecture have been experimentally evaluated, Res4Net, Res5 ... till Res10Net and Ensemble CNN. Res4Net performed the best in Ischaemia detection with .978 F1 score and Res7Net in infection detection with 0.798 F1 score.
Liu et al. [50] (Binary classification of ischaemia and infection in DFU)	Public dataset provided by Goyal et al. [3]	Experimented with EfficientNet. Achieved F1 score of 0.9939 in ischemia and 0.9792 in infection detection respectively, for the augmented test images.
Garaawi et al. [51] (Binary classification of ischaemia and infection in DFU)	Public dataset provided by Goyal et al. [3]	Fused RGB input images and Mapped LBP codes, the resultant images were trained using custom CNN model. Achieved 0.975 F-Measure for ischaemia and 0.787 for infection detection.
Yap et al. [52] (Multi-class classification)	Public DFUC2021 [53]	Analysis of a various methods with and without pretrained models like VGG16, ResNet101, InceptionV3, DenseNet121, EfficientNetB0. The overall best performance is given by EfficientNetB0 with F1 score of 0.55.
Lingmei et al. [54] (Multi-class classification)	Public DFUC2021 [53]	The authors validated the performance of the model on the DFUC2021 test set and achieved 0.593 F1-score.
Goyal et al. [25] (DFU localization)	Private dataset of 1775 images split into training 1242 images, 178 validation and 355 in test images.	Utilized Faster R-CNN with InceptionV2 model to achieve a mean average precision of 91.8% for DFU localization.
Oliveira et al. [55] (DFU localization)	public DFUC2020 [56] dataset	Different localization models such as Faster RCNN, Faster RCNN FP and Faster RCNN DFU were compared, and the overall best performed model was Faster RCNN DFU with F1-score of 0.948.
Huang et al. [57] (DFU segmentation)	Private dataset with 727 images	Faster R-CNN was utilised for DFU segmentation and achieved an accuracy of 90% for 200 test images from augmented dataset.
Hüseyin et al. [58] (Segmentation)	Medetec pressure wound image dataset	They used the segmentation architectures FCN, PSP, UNet, SegNet, and DeepLabV3 and used the backbone (base) models ResNet, VGG-16, MobileNet, EfficientNet, and Vanilla CNN. The MobileNet-UNet showed an accuracy of 99.67%.
Cao et al. [59] (DFU segmentation and multi-level classification)	Private dataset with 1000 images along with DFUC2020 [5] dataset	Built a semantic segmentation of DFU using Mask R-CNN and performed multi-level classification. The authors achieved an F1 score of 0.7696 for multi-level classification of DFU.

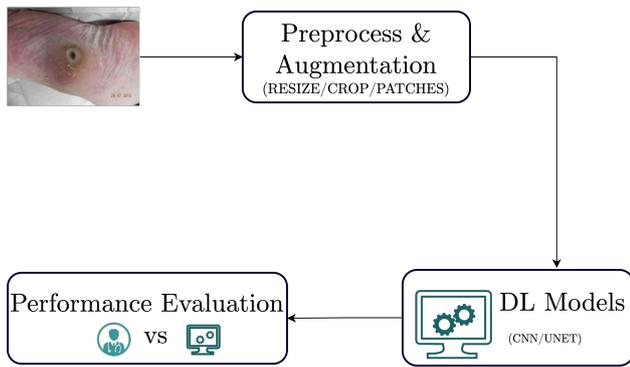


Fig. 6: DL based DFU CDSS.

variants of the proposed architecture that were designed based on Res4Net, Res5 and ensemble CNN. Res4Net performed the best in ischaemia detection with a 0.978 F1 score and Res7Net in infection detection with a 0.798 F1 score for Goyal et al.'s dataset [3]. Liu et al. [50] experimented with EfficientNets achieving an F1 score of 0.9939 in ischemia and 0.9792 in infection detection, respectively, for the augmented test images. Garaawi et al. [51] fused RGB input images and Mapped LBP codes, the resultant images were trained using a custom CNN model to perform binary classification for the recognition of ischaemia and infection. They reported a 0.975 F-Measure for ischaemia and a 0.787 for infection detection.

An asymmetric convolutional transformer network was proposed by Lingmei et al. [54] for the multi-class (4-class) classification task of DFU. The authors validated the performance of the model on the DFUC2021 test set, and were able to achieve 0.593 F1-score. Yap et al. [52] analysed various pretrained models like VGG16, ResNet101, InceptionV3, DenseNet121, EfficientNetB0. The overall best performance is given by EfficientNetB0 with F1 score of 0.55, for the multi-class classification. Goyal et al. [25] localized DFU region on full foot images using faster R-CNN with InceptionV2 model and achieved a mean average precision of 91.8% for 355 testing images. Huang et al. [57] segmented the DFU using the private dataset with 727 images. They utilised faster R-CNN for DFU segmentation and achieved an accuracy of 90% for 200 test images from an augmented dataset. Cao et al. [59] built a semantic segmentation of DFU using Mask R-CNN and performed multi-level classification using a private dataset consisting of 1000 images. The authors achieved an F1 score of 0.7696 for the multi-level classification of DFU. Cassidy et al. [63] localised DFU region using the trained model and developed a cloud-based framework for remote detection of DFU.

III. EVALUATION METRICS

To effectively evaluate and benchmark the performance of CDS systems designed for DFU diagnosis, several standard metrics can be utilized. Some such metrics are sensitivity (true positive rate (TPR) or recall), specificity (false positive rate (FPR)), F1-score, precision, the area under the receiver operating characteristic curve (AUC), overlapping error, accuracy, boundary-based evaluations, and the dice similarity coefficient have been used to evaluate the performance. Each of these metrics is primarily concerned with assessing the

performance of false positives (FP), true positives (TP), false negatives (FN), and true negatives (TN) identified by the CDSS. Here, TP are the number of DFU cases identified correctly by the CDSS system to be positive, matching expert opinion. TN represents the correctly detected non-DFU cases, and FN represents the number of incorrectly rejected DFU cases, matched with human expert opinion. The number of cases that were wrongly classified by the CDSS as DFU but not by human experts is known as FP .

The most popular metrics used for the DFU detection/classification/segmentation task are *Sensitivity* (also called Recall or True Positive Rate (TPR)), *Specificity* (False Positive Rate (FPR)), *Precision* (Positive Predictive Value), *F1-score* (Dice similarity coefficient), *Accuracy*, *Area under Receiver Operative Characteristic Curve* (AUC) and *Intersection over Union* (IOU, Jaccard similarity index). Sensitivity (Eq. 1) is the ratio of accurately identified findings, and a higher value indicates a better system. Specificity (Eq. 2) is the ratio of correctly identified non-DFU cases, and the higher the value, the better is the system. Precision (Eq. 3) illustrates how many of the DFU that were labeled positive, were correctly identified. The system performs better with higher value. F1-score (Eq. 4) is the harmonic mean of precision and recall, and the higher the value, the better the system. Accuracy (Eq. 5) is the ratio of correctly predicted MA to the total MA observations. AUC is the area under ROC curve is a probability curve (plotted as sensitivity versus fall-out), and the area under the curve represents measure of class separability. The greater the AUC value, the better the system distinguishes between DFU and non-DFU. IOU (Eq. 6) evaluates how close the predicted finding is to the ground truth finding. The intersection is 0 when there is no overlap between the ground truth and prediction bounding boxes As a result, IOU will be 0 as well.

$$TPR = \frac{TP}{(FN + TP)} \quad (1)$$

$$FPR = \frac{TN}{(FP + TN)} \quad (2)$$

$$Precision = \frac{TP}{(TP + FP)} \quad (3)$$

$$F1\ score = \frac{2 \times (TP)}{(2 \times TP + FP + FN)} \quad (4)$$

$$Accuracy = \frac{TP + TN}{(TP + FP + FN + TN)} \quad (5)$$

$$IOU = \frac{TP}{(TP + FN + FP)} \quad (6)$$

Tables III summarize the performance of various existing works using thermal DFU images for classification and detection tasks. Tables IV, V, and VI summarize the performance of various existing works using RGB DFU images for classification, segmentation, and detection tasks.

IV. DISCUSSION AND OBSERVED GAPS

From the detailed review of existing works, we were able to draw some significant conclusions, highlighting the goodness of available models and gaps in research that require further efforts from the research community. A CDSS

TABLE III: Performance of state-of-the-art models built on thermal DFU images - summary.

Task	Method	Accuracy	TPR	FPR	AUC	F1-Score	Precision
Binary classification	DFTNet [26]	0.9453	0.9534	0.9375	0.9455	0.9457	0.9401
	DenseNet201 [28]	0.9401	0.9401	0.9078	0.9401	0.9401	0.9401
	Custom CNN [29]	0.9827	-	-	-	0.9621	0.9626
	SVM+SURF [30]	0.9781	-	-	-	-	-
	SVM Quadratic[31]	0.912	-	-	-	0.825	0.826
	CNN+ML [36]	0.91	-	-	-	0.91	0.91
	K-means [38]	0.743	0.7300	0.778	0.840	0.805	0.899
	SVM [39]	0.9561	0.9650	0.9241	-	-	-
	ResNet50v2[32]	1	1	1	1	1	1
	MobileNetv2 + ShuffleNet [33]	1	1	1	1	1	1
Detection	SVM[40]	0.91	-	-	-	0.90	0.98
	Masked RCNN [43]	0.9028	0.9029	0.9028	-	-	-
	Faster RCNN [57]	0.900	-	-	-	-	-
	EfficientNet [41]	0.989	-	-	-	98	99
	FusionSegNet [42]	0.957	-	-	0.989	0.949	-

TABLE IV: Multi-class classification performance of state-of-the-art models built on RGB DFU images - summary.

Method	Accuracy	TPR	FPR	AUC	F1-Score	Precision
DFUSPnet [48]	0.964	0.984	0.951	0.974	0.954	0.926
DFUQUtNet [47]	-	-	0.936	-	0.845	0.954
DFUnet [5]	0.925	0.943	0.911	0.962	0.939	0.945
ACTNet [54]	-	-	-	0.824	0.593	-
EfficientNetB0 [52] multi-class (4-class) classification	0.620	-	-	-	0.570	0.550
DFUSC [59] multiclass classification	-	-	-	-	0.769	0.846

TABLE V: Segmentation performance of state-of-the-art models built on RGB DFU images - summary.

Method	Accuracy	TPR	FPR	AUC	F1-Score	Precision
ASURA [64]	0.84	-	-	0.91	0.90	0.92
CONVNET [65]	-	-	-	0.308	0.348	0.40
FCN-16s [66]	-	0.9	0.988	-	-	-
Patch-based CNN [67]	-	0.917	0.973	-	-	-
Object Detection Fast R-CNN method [57]	0.90	-	-	-	-	-
Faster R-CNN DFU [55]	-	-	-	-	0.948	0.914
Faster R-CNN FP [55]	-	-	-	-	0.919	0.865
MobileNet-UNet [58]	0.996	-	-	-	0.99	0.995

TABLE VI: Detection performance of state-of-the-art models built on RGB DFU images - summary.

Method	Time (in sec)	mAP	FPR	Parameters (in million)	F1-Score	Precision
SSD-MobileNet [68]	0.28	0.84	-	22.6	-	-
SSD-InceptionV2 [68]	0.37	0.87	-	53.5	-	-
Faster R-CNN with InceptionV2 [68]	0.48	0.91	-	52.2	-	-
R-FCN with Resnet 101 [68]	0.90	0.90	-	199.1	-	-
FRCNN R-FCN [69]	-	0.65	0.75	-	0.67	0.61
FRCNN ResNet101 [69]	-	0.65	0.73	-	0.66	0.59
FRCNN Inception-v2-ResNet101 [69]	-	0.64	0.75	-	0.67	0.60
EffDet [69]	-	0.62	0.69	-	0.69	0.69
EfficientDet-D5 [70]	-	0.50	-	-	-	-
EfficientDet-D6 [70]	-	0.51	-	-	-	-
NFU-Net (PReLU) (D2) [46]	1.00	-	-	-	1.00	1.00
Alex Net (D2) [46]	0.918	-	-	-	0.919	0.933
LeNet (D2) [46]	0.896	-	-	-	0.889	0.877
LRC [45]	0.937	-	-	9.5	0.954	-
Ischaemia classification [49]	0.90	-	-	-	-	-
Infection classification [49]	0.73	-	-	-	-	-
EfficientNet- Infection classification [50]	-	0.979	-	-	0.979	0.979
EfficientNet- Ischaemia classification [50]	-	0.993	-	-	0.993	0.993
DFU-RGB-TEX-NET [51]	-	-	-	-	0.981	0.952
DFU_SPNet [47]	0.974	-	-	-	-	-
DFU_QUTNet [48]	-	-	-	-	0.945	0.954

designed for automated DFU diagnosis based tasks can provide crucial alternatives for manual screening, especially in rural areas where trained experts are scarce and in semi-urban areas where skilled physicians are not always available. However, numerous challenges must be overcome when deploying a CDSS in a real-time clinical setting.

- *Data quality:* DFU images are often acquired from several sources and annotated by physicians with varied

degrees of experience; the CDSS should be robust enough to be able to accurately predict disease onset and progression while being able to manage any discrepancies or outliers caused by these differences in data quality and expert opinion.

- *Data imbalance:* Currently, publicly available DFU datasets are extremely skewed in terms of class distribution, with many infection instances but just a few is-

chaemia cases. Generative Adversarial Networks (GAN) can be utilized to synthetically augment the dataset, to offset the lack of bigger datasets.

- *Data modality:* Currently, publicly available DFU datasets mostly include RGB images, while, thermogram datasets are very scarce and less in number. This could be due to the requirement of a sophisticated setup and advanced infrared cameras for capturing thermograms. The combination of diagnosis and image features, such as, patient ethnicity, the presence of ischemia, the depth of DFU to the tendon, and neuropathy would aid in the development of a more robust DFU diagnosis system.
- *Data loss:* Many existing works focused on reducing the image size to to 128x128. This may have resulted in some information loss, which can be possibly prevented via ensemble of results from image patches.
- *Varied processing requirements:* Multimodal DFU data is often collected from a variety of sources (e.g., imaging data from multiple devices). Each data modality may need its own preprocessing and AI approaches. An ensemble of results from different preprocessing approaches could aid improve the robustness of the CDSS.
- *Overfitting issue:* For good performance, CNNs require a large volume of labelled training set. Since some of the available DFU datasets are small, the CNN parameters may be poorly tuned, resulting in overfitting. GANs could be utilized to synthetically augment the dataset, thus aid in development of a generalizable CDSS.
- *Unsupervised learning issue:* K-means clustering is typically employed before feature extraction. The quality of the output in K-means clustering is determined by the arbitrary selection of the initial centroid, allowing for different results for each cluster within the same image.
- *Dealing with multiple labels:* Designing methods to detect the co-occurrences of multiple conditions, such as ischaemia and infection, pose many significant challenges.
- *Interpretability:* The inability to provide an explainable interpretation for the projected outcomes may cause scepticism among healthcare professionals. CDSSs based on machine learning and deep learning are considered black boxes, and in domains such as healthcare, it is frequently required to provide information on how and why the system recommended a specific outcome. Thus, interpretability is already a crucial aspect in CDSS for early diagnosis of DFU. Recently introduced Explainable AI (XAI) techniques [71] such as score deviation maps and recursive division methods can be investigated.

V. CONCLUDING REMARKS

Diabetes related complications like DFUs are a treatable and preventable chronic condition that can result in repeated hospitalisation, the amputation of lower limbs or toes, and in more severe situations, even fatality. Its identification could

therefore prevent additional complications and enable better prognosis. In this research work, a thorough assessment of significant works on CDSS for DFU was undertaken, encompassing the identification and segmentation of DFU using both traditional image processing methods and DL-based approaches. A taxonomy of DFU CDSS classified into two groups, techniques based on thermal imaging and approaches based on RGB, was presented and discussed. Detailed discussions on widely used data augmentation, pre-processing, and modelling methodologies, as well as publicly accessible DFU databases, were also presented. Additionally, the most common evaluation metrics used to assess CDSS for DFU were also described. The benefits and drawbacks of various existing approaches were discussed and illustrated via a task-wise comparative evaluation of state-of-the-art approaches on standard datasets. The majority of performance enhancements were observed in binary classification. There is scope for improvement in multi-class classification, which is more advantageous for distinguishing between infective vs peripheral vascular DFU.

Additionally, the primary difficulties that must be overcome in designing DFU CDSS were highlighted and studied in detail, for throwing light on future directions for research in the field. The need for scoring-based CDSSs that are built on interactive interfaces similar to those provided by inexpensive, widely available devices like smartphones is emphasized, as they would help doctors make an early diagnosis of DFU in a way that is both user-friendly and economical.

DATA AVAILABILITY

This review is based on available public databases and suitable references could be explored for further details.

REFERENCES

- [1] P. Saeedi, I. Petersohn, P. Salpea, B. Malanda, S. Karuranga, N. Unwin, S. Colagiuri, L. Guariguata, A. A. Motala, K. Ogurtsova, J. E. Shaw, D. Bright, and R. Williams, "Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the international diabetes federation diabetes atlas, 9th edition," *Diabetes Research and Clinical Practice*, vol. 157, pp. 107–117, 2019.
- [2] D. G. Armstrong, A. J. Boulton, and S. A. Bus, "Diabetic foot ulcers and their recurrence," *New England Journal of Medicine*, vol. 376, no. 24, pp. 2367–2375, 2017.
- [3] M. Goyal, N. D. Reeves, S. Rajbhandari, N. Ahmad, C. Wang, and M. H. Yap, "Recognition of ischaemia and infection in diabetic foot ulcers: Dataset and techniques," *Computers in Biology and Medicine*, vol. 117, pp. 103–113, 2020.
- [4] V. Mayya, K. Karthik, K. S. Sowmya, K. Karadka, and J. Jeganathan, "Coviddx: Ai-based clinical decision support system for learning covid-19 disease representations from multimodal patient data." in *HEALTHINF*, 2021, pp. 659–666.
- [5] M. Goyal, N. D. Reeves, A. K. Davison, S. Rajbhandari, J. Spragg, and M. H. Yap, "Dfunet: convolutional neural networks for diabetic foot ulcer classification," *IEEE*

- Transactions on Emerging Topics in Computational Intelligence*, pp. 1–12, 2018.
- [6] Y. Dalia, A. Bharath *et al.*, “deepoa: Clinical decision support system for early detection and severity grading of knee osteoarthritis,” in *2021 International Conference on Computer, Communication and Signal Processing (ICCCSP)*. IEEE, 2021, pp. 250–255.
- [7] I. M. Nedumkunnel, L. E. George, and N. A. Rosh, “Explainable deep neural models for covid-19 prediction from chest x-rays with region of interest visualization,” in *2021 International Conference on Secure Cyber Computing and Communications (ICSCCC)*. IEEE, 2021, pp. 96–101.
- [8] A. Sopharak, B. Uyyanonvara, and S. Barman, “Automatic microaneurysm detection from non-dilated diabetic retinopathy retinal images using mathematical morphology methods,” *IAENG International Journal of Computer Science*, vol. 38, no. 3, pp. 295–301, 2011.
- [9] R. Sarno, S. Izza Sabilla, D. Rahman Wijaya, and Hariyanto, “Electronic nose for detecting multilevel diabetes using optimized deep neural network,” *Engineering Letters*, vol. 28, no. 1, pp. 31–42, 2020.
- [10] K. Karthik and S. S. Kamath, “Swarm optimisation-based bag of visual words model for content-based x-ray scan retrieval,” *International Journal of Biomedical Engineering and Technology*, vol. 40, no. 2, pp. 168–183, 2022.
- [11] V. Mayya, S. K. S, U. Kulkarni, D. K. Surya, and U. R. Acharya, “An empirical study of preprocessing techniques with convolutional neural networks for accurate detection of chronic ocular diseases using fundus images,” *Applied Intelligence*, vol. 53, no. 2, pp. 1548–1566, 2023.
- [12] B. Mukesh, T. Harish, V. Mayya, and S. Kamath, “Deep learning based detection of diabetic retinopathy from inexpensive fundus imaging techniques,” 2021, Conference paper.
- [13] V. Mayya, S. K. S, and U. Kulkarni, “Automated microaneurysms detection for early diagnosis of diabetic retinopathy: A comprehensive review,” *Computer Methods and Programs in Biomedicine Update*, vol. 1, pp. 100–013, 2021.
- [14] H.-M. Song, Y. Liu, J.-S. Wang, and B. Zhou, “Research on segmentation algorithms of retinal vessel images,” *IAENG International Journal of Computer Science*, vol. 49, no. 2, pp. 286–298, 2022.
- [15] X.-D. Li, J.-S. Wang, W.-K. Hao, and X.-R. Zhao, “Bi-directional long-term short-term memory classification method of diabetes datasets,” *Engineering Letters*, vol. 30, no. 4, pp. 1603–1611, 2022.
- [16] V. Mayya, S. Shevgoor, U. Kulkarni, M. Hazarika, P. D. Barua, and U. R. Acharya, “Multi-scale convolutional neural network for accurate corneal segmentation in early detection of fungal keratitis,” *Journal of Fungi*, vol. 7, no. 10, p. 850, 2021.
- [17] Erwin, H. K. Putra, B. Suprihatin, and F. Ramadhini, “A hybrid clahe-gamma adjustment and densely connected u-net for retinal blood vessel segmentation using augmentation data,” *Engineering Letters*, vol. 30, no. 2, pp. 485–493, 2022.
- [18] M. Bansal and N. Singh, “Retinal vessel segmentation techniques: A review,” vol. 1, 2017, Conference paper, pp. 442–447.
- [19] K. Karthik and S. Kamath, “Automated view orientation classification for x-ray images using deep neural networks,” in *Smart computational intelligence in biomedical and health informatics*. CRC Press, 2021, pp. 61–72.
- [20] K. Katara and S. Shevgoor, “Msdnet: A deep neural ensemble model for abnormality detection and classification of plain radiographs,” *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–15, 2022.
- [21] L. Alzubaidi, M. A. Fadhel, S. R. Oleiwi, O. Al-Shamma, and J. Zhang, “Dfu_qutnet: diabetic foot ulcer classification using novel deep convolutional neural network,” *Multimedia Tools and Applications*, vol. 79, pp. 15 655–15 677, 2019.
- [22] D. Hernandez-Contreras, H. Peregrina-Barreto, J. Rangel-Magdaleno, and F. Renero, “Plantar thermogram database for the study of diabetic foot complications,” *IEEE Access*, vol. PP, pp. 1–10, 11 2019.
- [23] L. Shamseer, D. Moher, M. Clarke, D. Ghersi, A. Liberati, M. Petticrew, P. Shekelle, and L. A. Stewart, “Preferred reporting items for systematic review and meta-analysis protocols (prisma-p) 2015: elaboration and explanation,” *BMJ*, vol. 349, 2015.
- [24] C. Cui, K. Thurnhofer-Hemsi, R. Soroushmehr, A. Mishra, J. Gryak, E. Domínguez, K. Najarian, and E. López-Rubio, “Diabetic wound segmentation using convolutional neural networks,” in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2019, pp. 1002–1005.
- [25] M. Goyal, N. D. Reeves, S. Rajbhandari, and M. H. Yap, “Robust methods for real-time diabetic foot ulcer detection and localization on mobile devices,” *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 4, pp. 1730–1741, 2019.
- [26] I. Cruz-Vega, D. Hernandez-Contreras, H. Peregrina-Barreto, J. Rangel-Magdaleno, and J. Ramirez-Cortes, “Deep learning classification for diabetic foot thermograms,” *Sensors (Switzerland)*, vol. 20, no. 6, 2020.
- [27] C. B. Pereira, X. Yu, S. Dahlmans, V. Blazek, S. Leonhardt, and D. Teichmann, *Infrared Thermography*. Cham: Springer International Publishing, 2018.
- [28] A. Khandakar, M. Chowdhury, M. Ibne Reaz, S. Md Ali, M. Hasan, S. Kiranyaz, T. Rahman, R. Alfkey, A. Bakar, and R. Malik, “A machine learning model for early detection of diabetic foot using thermogram images,” *Computers in Biology and Medicine*, vol. 137, 2021.
- [29] S. Muralidhara, A. Lucieri, A. Dengel, and S. Ahmed, “Holistic multi-class classification & grading of diabetic foot ulcerations from plantar thermal images using deep learning,” *Health Information Science and Systems*, vol. 10, no. 1, 2022.
- [30] M. H. Alshayegi, 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*, vol. 79, pp. 104–143, 2023.

- [31] V. Filipe, P. Teixeira, and A. Teixeira, "Automatic classification of foot thermograms using machine learning techniques," *Algorithms*, vol. 15, no. 7, 2022.
- [32] A. Anaya-Isaza and M. Zequera-Diaz, "Fourier transform-based data augmentation in deep learning for diabetic foot thermograph classification," *Biocybernetics and Biomedical Engineering*, vol. 42, no. 2, pp. 437–452, 2022.
- [33] K. Munadi, K. Saddami, M. Oktiana, R. Roslidar, K. Muchtar, M. Melinda, R. Muharar, M. Syukri, T. F. Abidin, and F. Arnia, "A deep learning method for early detection of diabetic foot using decision fusion and thermal images," *Applied Sciences (Switzerland)*, vol. 12, no. 15, 2022.
- [34] A. Hernandez-Guedes, I. Santana-Perez, N. Arteaga-Marrero, H. Fabelo, G. M. Callico, and J. Ruiz-Alzola, "Performance evaluation of deep learning models for image classification over small datasets: Diabetic foot case study," *IEEE Access*, vol. 10, pp. 124 373–124 386, 2022.
- [35] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.
- [36] A. Khandakar, M. E. H. Chowdhury, M. B. I. Reaz, S. H. M. Ali, T. O. Abbas, T. Alam, M. A. Ayari, Z. B. Mahbub, R. Habib, T. Rahman, A. M. Tahir, A. A. Bakar, and R. A. Malik, "Thermal change index-based diabetic foot thermogram image classification using machine learning techniques," *Sensors*, vol. 22, no. 5, 2022.
- [37] D. Bouallal, H. Douzi, and R. Harba, "Diabetic foot thermal image segmentation using double encoder-resunet (de-resunet)," *Journal of Medical Engineering and Technology*, vol. 46, no. 5, pp. 378–392, 2022.
- [38] V. Filipe, P. Teixeira, and A. Teixeira, "A clustering approach for prediction of diabetic foot using thermal images," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 12251 LNCS, pp. 620–631, 2020.
- [39] J. Saminathan, M. Sasikala, V. Narayanamurthy, K. Rajesh, and R. Arvind, "Computer aided detection of diabetic foot ulcer using asymmetry analysis of texture and temperature features," *Infrared Physics & Technology*, vol. 105, pp. 103–113, 2020.
- [40] 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*, vol. 23, no. 2, 2023.
- [41] P. N. Thotad, G. R. Bharamagoudar, and B. S. Anami, "Diabetic foot ulcer detection using deep learning approaches," *Sensors International*, vol. 4, pp. 100–110, 2023.
- [42] C. J. Lan T, Li Z, "Fusionsegnet: Fusing global foot features and local wound features to diagnose diabetic foot," *Comput Biol Med*, vol. 152, 2023.
- [43] H. Maldonado, R. Bayareh, I. Torres, A. Vera, J. Gutiérrez, and L. Leija, "Automatic detection of risk zones in diabetic foot soles by processing thermographic images taken in an uncontrolled environment," *Infrared Physics and Technology*, vol. 105, 2020.
- [44] S. Madhava Prabhu and S. Verma, "Automatic segmentation of plantar thermograms using adaptive c means technique," *International Journal of Electrical and Computer Engineering*, vol. 11, no. 2, pp. 1250–1258, 2021.
- [45] S. K. Das, P. Roy, and A. K. Mishra, "Fusion of hand-crafted and deep convolutional neural network features for effective identification of diabetic foot ulcer," *Concurrency and Computation: Practice and Experience*, vol. 34, no. 5, 2022.
- [46] C. Venkatesan, M. Sumithra, and M. Murugappan, "Nfu-net: An automated framework for the detection of neurotrophic foot ulcer using deep convolutional neural network," *Neural Processing Letters*, vol. 54, no. 5, pp. 3705–3726, 2022.
- [47] S. Das, P. Roy, and A. Mishra, "Dfu_spnet: A stacked parallel convolution layers based cnn to improve diabetic foot ulcer classification," *ICT Express*, vol. 8, no. 2, pp. 271–275, 2022.
- [48] L. Alzubaidi, M. Fadhel, S. Oleiwi, O. Al-Shamma, and J. Zhang, "Dfu_qutnet: diabetic foot ulcer classification using novel deep convolutional neural network," *Multimedia Tools and Applications*, vol. 79, no. 21-22, pp. 15 655–15 677, 2020.
- [49] S. K. Das, P. Roy, and A. K. Mishra, "Recognition of ischaemia and infection in diabetic foot ulcer: A deep convolutional neural network based approach," *International Journal of Imaging Systems and Technology*, vol. 32, no. 1, pp. 192–208, 2022.
- [50] Z. Liu, J. John, and E. Agu, "Diabetic foot ulcer ischemia and infection classification using efficientnet deep learning models," *IEEE Open Journal of Engineering in Medicine and Biology*, pp. 1–14, 2022.
- [51] N. Al-Garaawi, R. Ebsim, A. F. Alharan, and M. H. Yap, "Diabetic foot ulcer classification using mapped binary patterns and convolutional neural networks," *Computers in Biology and Medicine*, vol. 140, 2022.
- [52] M. H. Yap, B. Cassidy, J. M. Pappachan, C. O'Shea, D. Gillespie, and N. D. Reeves, "Analysis towards classification of infection and ischaemia of diabetic foot ulcers," *CoRR*, vol. 1, 2021.
- [53] B. Cassidy, C. Kendrick, N. D. Reeves, J. M. Pappachan, C. O'Shea, D. G. Armstrong, and M. H. Yap, "Diabetic foot ulcer grand challenge 2021: Evaluation and summary," in *Diabetic Foot Ulcers Grand Challenge*. Springer International Publishing, 2022, pp. 90–105.
- [54] L. Ai, M. Yang, and Z. Xie, "Actnet: asymmetric convolutional transformer network for diabetic foot ulcers classification," *Physical and Engineering Sciences in Medicine*, vol. 45, no. 4, pp. 1175–1181, 2022.
- [55] A. Da Costa Oliveira, A. De Carvalho, and D. Dantas, "Faster r-cnn approach for diabetic foot ulcer detection," in *VISIGRAPP 2021 - Proceedings of the 16th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, vol. 4, 2021, pp. 677–684.
- [56] B. Cassidy, N. Reeves, J. Pappachan, D. Gillespie, C. O'Shea, S. Rajbhandari, A. Maiya, E. Frank, A. Boulton, D. Armstrong, B. Najafi, J. Wu, R. Kochhar, and M. Yap, "The dfuc 2020 dataset: Anal-

- ysis towards diabetic foot ulcer detection,” *European Endocrinology*, vol. 1, no. 1, pp. 5–11, 2021.
- [57] H.-N. Huang, T. Zhang, C.-T. Yang, Y.-J. Sheen, H.-M. Chen, C.-J. Chen, and M.-W. Tseng, “Image segmentation using transfer learning and fast r-cnn for diabetic foot wound treatments,” *Frontiers in Public Health*, vol. 10, 2022.
- [58] H. Eldem, E. Ülker, and O. Y. Işıklı, “Encoder–decoder semantic segmentation models for pressure wound images,” *The Imaging Science Journal*, vol. 0, no. 0, pp. 1–12, 2023.
- [59] 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*, 2022.
- [60] C. Kendrick, B. Cassidy, J. M. Pappachan, C. O’Shea, C. J. Fernandez, E. Chacko, K. Jacob, N. D. Reeves, and M. H. Yap, “Translating clinical delineation of diabetic foot ulcers into machine interpretable segmentation,” 2022.
- [61] C. Wang, D. M. Anisuzzaman, V. Williamson, M. K. Dhar, B. Rostami, J. Niezgodá, S. Gopalakrishnan, and Z. Yu, “Fully automatic wound segmentation with deep convolutional neural networks,” *Scientific Reports*, vol. 10, pp. 21–31, Jan. 2020.
- [62] G. D. Finlayson and E. Trezzi, “Shades of gray and colour constancy,” in *Color Imaging Conference*. IS&T - The Society for Imaging Science and Technology, 2004, pp. 37–41.
- [63] B. Cassidy, N. D. Reeves, J. M. Pappachan, N. Ahmad, S. Haycocks, D. Gillespie, and M. H. Yap, “A cloud-based deep learning framework for remote detection of diabetic foot ulcers,” *IEEE Pervasive Computing*, vol. 21, no. 2, pp. 78–86, 2022.
- [64] D. Y. Chino, L. C. Scabora, M. T. Cazzolato, A. E. Jorge, C. Traina-Jr., and A. J. Traina, “Segmenting skin ulcers and measuring the wound area using deep convolutional networks,” *Computer Methods and Programs in Biomedicine*, vol. 191, pp. 105–115, 2020.
- [65] C. Wang, X. Yan, M. Smith, K. Kochhar, M. Rubin, S. M. Warren, J. Wrobel, and H. Lee, “A unified framework for automatic wound segmentation and analysis with deep convolutional neural networks,” in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2015, pp. 2415–2418.
- [66] M. Goyal, M. H. Yap, N. D. Reeves, S. Rajbhandari, and J. Spragg, “Fully convolutional networks for diabetic foot ulcer segmentation,” in *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2017, pp. 618–623.
- [67] C. Cui, K. Thurnhofer-Hemsi, R. Soroushmehr, A. Mishra, J. Gryak, E. Domínguez, K. Najarian, and E. López-Rubio, “Diabetic wound segmentation using convolutional neural networks,” in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2019, pp. 1002–1005.
- [68] M. Goyal, N. D. Reeves, S. Rajbhandari, and M. H. Yap, “Robust Methods for Real-Time Diabetic Foot Ulcer Detection and Localization on Mobile Devices,” *IEEE J Biomed Health Inform*, vol. 23, no. 4, pp. 1730–1741, 07 2019.
- [69] B. Cassidy, N. D. Reeves, J. M. Pappachan, D. Gillespie, C. O’Shea, S. Rajbhandari, A. G. Maiya, E. Frank, A. J. Boulton, D. G. Armstrong, B. Najafi, J. Wu, R. S. Kochhar, and M. H. Yap, “The DFUC 2020 Dataset: Analysis Towards Diabetic Foot Ulcer Detection,” *touchREV Endocrinol*, vol. 17, no. 1, pp. 5–11, Apr 2021.
- [70] M. Goyal and S. Hassanpour, “A refined deep learning architecture for diabetic foot ulcers detection,” *CoRR*, vol. abs/2007.07922, 2020.
- [71] G. Vilone and L. Longo, “Classification of explainable artificial intelligence methods through their output formats,” *Machine Learning and Knowledge Extraction*, vol. 3, no. 3, pp. 615–661, 2021.