DFUNet: Convolutional Neural Networks for Diabetic Foot Ulcer Classification

Manu Goyal , *Student Member, IEEE*, Neil D. Reeves , Adrian K. Davison , *Member, IEEE*, Satyan Rajbhandari, Jennifer Spragg, and Moi Hoon Yap , *Member, IEEE*

Abstract—Globally, in 2016, 1 out of 11 adults suffered from diabetes mellitus. Diabetic foot ulcers (DFU) are a major complication of this disease, which if not managed properly can lead to amputation. Current clinical approaches to DFU treatment rely on patient and clinician vigilance, which has significant limitations, such as the high cost involved in the diagnosis, treatment, and lengthy care of the DFU. We collected an extensive dataset of foot images, which contain DFU from different patients. In this DFU classification problem, we assessed the two classes as normal skin (healthy skin) and abnormal skin (DFU). In this paper, we have proposed the use of machine learning algorithms to extract the features for DFU and healthy skin patches to understand the differences in the computer vision perspective. This experiment is performed to evaluate the skin conditions of both classes that are at high risk of misclassification by computer vision algorithms. Furthermore, we used convolutional neural networks for the first time in this binary classification. We have proposed a novel convolutional neural network architecture, DFUNet, with better feature extraction to identify the feature differences between healthy skin and the DFU. Using 10-fold cross validation, DFUNet achieved an AUC score of 0.961. This outperformed both the traditional machine learning and deep learning classifiers we have tested. Here, we present the development of a novel and highly sensitive DFUNet for objectively detecting the presence of DFUs. This novel approach has the potential to deliver a paradigm shift in diabetic foot care among diabetic patients, which represent a cost-effective, remote, and convenient healthcare solution.

Index Terms—Diabetic foot ulcers, classification, deep learning, convolutional neural networks, DFUNet.

I. INTRODUCTION

DIABETES Mellitus (DM) commonly known as Diabetes, is a lifelong condition resulting from hyperglycemia (high blood sugar levels), which leads to major life-threatening complications such as cardiovascular diseases, kidney failure, blindness and lower limb amputation which is often preceded by Diabetic Foot Ulcers (DFU) [1]. According to the global report on diabetes, in 2014, there are 422 million people living with DM compared to 108 million people in 1980. Among the adults that are over 18 years of age, the global prevalence has gone up from 4.7% in 1980 to 8.5% in 2014 [2]. It is estimated by the end of 2035, the figure is expected to rise to 600 million people living with DM worldwide [3]. It is worth mentioning that about only 20% of these people will be from developed countries and the rest will be from developing countries due to poor awareness and limited healthcare facilities [4]. There is about 15%−25% chance that a diabetic patient will eventually develop DFU and if proper care is not taken, that may result in lower limb amputation [5]. Every year, more than 1 million patients who have diabetes lose part of their leg due to the failure to recognize and treat DFU appropriately [6]. A Diabetic patient with a 'high risk' foot needs periodic check-ups of doctors, continuous expensive medication, and hygienic personal care to avoid the further consequences as discussed earlier. Hence, it causes a great financial burden on the patients and their family, especially in developing countries where the cost of treating this disease can be equivalent to 5.7 years of annual income [7].

In current clinical practices, the evaluation of DFU comprises of various important tasks in early diagnosis, keeping track of development and number of lengthy actions taken in the treatment and management of DFU for each particular case: 1) the medical history of the patient is evaluated; 2) a wound or diabetic foot specialist examines the DFU thoroughly; 3) additional tests like CT scans, MRI, X-Ray may be useful to help develop a treatment plan. The patient with DFU has problem of a swollen leg, although it can be itchy and painful depending on each case. Usually, the DFU have irregular structures and uncertain outer boundaries. The visual appearance of DFU and its surrounding skin depending upon the various stages i.e. redness, callus formation, blisters, significant tissues types like granulation, slough, bleeding, scaly skin. Hence, the ulcer evaluation with computer vision algorithms would be based on the exact assessment of these visual signs as color descriptors and texture features.

The major challenges that are involved with this classification task are as follows: 1) large time in collection and expert labelling of the DFU images 2) high inter-class similarity between the normal (healthy skin) and abnormal classes (DFU) and intraclass variations depending upon the classification of DFU [8],

lighting conditions and patient's ethnicity. In this work, we propose computer vision algorithms to differentiate DFU from the healthy skin with traditional machine learning and deep learning approaches. The key contributions of this paper include:

1) This paper presents the related computerised telemedicine systems designed for DFU. We also present the DFU dataset of 397 foot images, which consists of 292 images with DFU and 105 healthy foot images. The podiatrists delineated the total of 1679 skin patches with 641 of healthy skin and 1038 of DFU. We are working with the ethical approval with NHS to make this dataset available for reproducibility of this work.

2) To the best of our knowledge, this is the first time, machine learning algorithms are used to understand and extract the computer vision features from DFU and healthy skin patches. We use convolutional Neural Networks (CNNs) to develop a fully automatic method to classify the DFU skin against the normal skin. Codes will be made publicly available after the acceptance of this work.

3) Development of a novel CNN architecture called DFU-Net, which is fine-tuned to process the input data more effectively and efficiently than other comparative state-of-the-art CNNs architecture. CNNs require substantial data to produce very accurate results, but with large filter size of DFUNet in parallel convolution blocks, it can produce good results on small dataset such as Facial Skin dataset.

II. RELATED WORK

The proliferation of information and communication technologies present both challenges and opportunities in terms of the development of new age healthcare systems. There are a number of telemedicine systems that are currently being developed a) to improve the current healthcare systems and also, decrease the cost of medical facilities; b) to improve the reach of medical facilities i.e. frequent remote assessment of patients with communication devices; c) to provide the automated solutions to deal with the shortage of expert medical professionals for these chronic diseases [9]. Over the years, researchers and doctors have developed key telemedicine systems to monitor diabetes [10], [11]. However, there are very few intelligent systems developed for assessment of diabetic foot pathologies which can be categorized into non-automated and automated telemedicine systems.

A. Telemedicine Systems for DFU

With the rapid growth in mobile telecommunications, remote communication is made possible with standalone devices like smart-phones, laptops and Internet. Nowadays, a pocket size smart-phone with the advanced mobile operating system has the capability of a personal computer that can capture and send high-resolution pictures and also, audio and video communication with advanced mobile internet like 4G. In the non-automated category, the common telemedicine systems based on these devices that are mostly set-up in the remote location for assessment of patients a) video conferencing [12]; b) three-

dimensional (3D) wound imaging [13]; c) digital photography [14]; d) optical scanner [15]. However, there is still need of specialized medical professionals on the other side for completing the assessment of the patient. Though these systems provide promising results, there is an urgent need of intelligent systems which can automatically detect the DFU pathologies remotely.

The use of automated telemedicine systems for DFU is still in its infancy. Notably, Liu *et al.* [16], [17] in 2015 developed an intelligent telemedicine system for detection of diabetic foot complications with spectral imaging, infra-red thermal images and 3D surface reconstruction. However, to implement this system, there is a requirement of several expensive devices and specialist training to use these devices. Wang *et al.* [18] have used an image capture box to capture image data and determined the area of DFU using cascaded two staged Support Vector Machine based classification. They proposed the use of a super-pixel technique for segmentation and extracted the number of features to perform two staged classification. Although this system reported a promising result, it has not been validated on a large dataset. Also, the image capture box is very impractical for data collection as there is a need for contact of the patient's feet and box surface which would not be allowed in a healthcare setting because of concerns regarding infection control. In other significant work, Goyal *et al.* [19] perform the segmentation of DFU and surrounding skin on the full foot images.

Additionally, computer methods based on manually engineered features or image processing approaches were implemented for tissue classification and segmentation of related skin lesion such as the wound. The conventional machine learning for classification task was performed by extracting various features such as texture descriptors and color descriptors on small delineated patches of wound images, followed by machine learning algorithms to classify them into normal and abnormal skin patches [20]–[24]. As in many computer vision systems, the hand-crafted features are affected by lighting conditions and skin color depending upon the ethnicity group of the patient. In general, virtually all the skin lesions related to both wound and ulcer are now termed as wound. In medical perspective, both wound and ulcer are considered differently as the wound is caused by an external problem whereas, ulcer are caused by an internal problem. Also, there are differences in appearance of the skin lesion of wound and ulcer, the cause (aetiology), the way the body responds (physiology) and disease processes (pathology) [25]. Hence, in this present study, only DFU are considered to determine how they are different from the normal healthy skin at the same place of appearance.

B. Computer Vision and Deep Learning

In recent years, there has been a rapid development in the area of computer vision, especially towards the difficult and important issues like understanding images of different domains such as spectral, medical, object and face detection, multi-class and label classification [26]–[32]. The conventional computer vision and machine learning algorithms were very limited in their ability to process the extensive image data, provide the representations of data with multiple levels of abstraction, and require a

lot of manual tuning for each input image. Deep convolutional networks as a recent machine learning algorithm comes out as an important technique to solve these kinds of computer vision problems [33], [34]. Deep convolutional networks obtain the multiple levels of representation methods by simple non-linear modules which transform the simple feature representation into the more advanced abstract representations for classification. Deep convolutional networks use images as input and start to learn features such as edges at specific directions and positions from the array of pixel values. At the higher level, it combines these edges to learn more important abstract features such as components of desirable objects and finally, these components are connected with each other to form final objects [33].

Supervised learning is one of the most common forms of machine learning. It is essential for the training of the network as the system learn the classification tasks from an extensive collection of images that are labelled differently for each category. Without training, it is not possible for the machine to detect the desired class by giving the highest score of all classes [35]–[37]. During the training stage, different images are processed by the machine to produce the output vector of scores for all classes for each image and then, the error is measured in respect of output scores versus the expected score until the desirable score for each class is obtained. After training, a validation set of data or images is used to fine tune the hyper-parameters of networks like setting the weights for each layer and the number of convolutional and pooling layers. Lastly, the system is tested with real-world test data without any expected outcome to check the performance of the system.

The remainder of the paper is structured as follows. Section III describes the methodology that we used to design classifiers based on conventional machine learning (CML) and CNNs and provides details of our proposed DFUNet. In Section IV, performance of various classifiers is tested with evaluation metrics like *Sensitivity*, *Specificity*, *Precision*, *F-Measure*, and *Area Under the receiver operating characteristic Curve (AUC)*. In Section V, the conclusion and future scope of our work are discussed.

III. METHODOLOGY

This section describes the proposed dataset containing examples of DFU of various patients. This includes expert labelling of the different regions as normal and abnormal skin patches. In addition, the feature descriptors used in experiments are detailed, including for CML, the CNNs architecture of LeNet, AlexNet, and GoogLeNet. Finally, we propose our own CNN architecture, DFUNet, to improve the way DFU are classified.

A. DFU Dataset

The first challenge was to collect a dataset of standardized color images of DFU from various patients to train the various deep learning model. We utilized an extensive database of 292 images of patient's foot with DFU over the previous five years at the Lancashire Teaching Hospitals, obtaining ethical approval from all relevant bodies and patients written informed consent. Also, we collected 105 images of the healthy foot to get the more cases for normal healthy class. Approval was obtained from the

NHS Research Ethics Committee to use these images for this research. These DFU images were captured with Nikon D3300. Whenever possible, the images were acquired with close-ups of the full foot with the distance of around 30–40 cm with the parallel orientation to the plane of an ulcer. The use of flash as the primary light source was avoided and instead, adequate room lights are used to get the consistent colors in images. To ensure the close-range focus and avoiding the blurriness in images from the close distance, a Nikon AF-S DX Micro NIKKOR 40 mm f/2.8G lens was used. We also included another test case that is captured by iPad with FootSnap application to show the robustness of algorithms over heterogeneous capture setup [38]. It consists of 20 abnormal skin patches and 32 normal skin patches in this heterogeneous test case.

B. Expert Labelling of Images

With the available annotator from Hewitt *et al.* [39], for each full image of a foot with ulcers (as illustrated in Fig. 1), the medical experts delineated the Region Of Interest (ROI) which is an important region around the ulcer comprises of significant tissues of both normal and abnormal skin. The ground truth labels are delineated by medical professionals in the form of both normal and abnormal skin patches from the ROI region. In the collection of ground truth patches, the experts only collected both classes of patches from ROI region that helped with more robust classification of the patches rather than involving the whole foot as a region. For each delineated abnormal region, the ground truth of the type of the abnormality was labelled and exported to an Extensible Markup Language (XML) file. For the annotation of 397-foot images with both ulcer and non-ulcer, there is a total of 292 ROI (Only for the foot images with ulcers). From these annotations, we produce a total of 1679 skin patches with 641 of normal and 1038 of abnormal class. Finally, we divided the dataset into training set of 1423 patches, validation set of 84 patches and testing set of 172 patches. The annotator tool which can delineate the image into different types of patches is shown in Fig. 1.

C. Data Augmentation of Training Patches

Deep networks require a lot of training image data because of the enormous number of parameters, especially weights associated with convolutional layers needed to be tuned by learning algorithms. Hence, we used data augmentation to improve the performance of the deep learning methods. We used the combination of various image processing techniques like rotation, flipping, contrast enhancement, using different color space, and random scaling to perform data augmentation. The rotation is performed by rotating the image by the angle of 90°, 180°, 270°. Then, three types of flipping (horizontal flip, vertical flip and horizontal+vertical flip) performed on the original patches. The four color space that are used for data augmentation are *Ycbcr*, *NTSC*, *HSV* and *L*a*b*. In the contrast enhancement, we used the three functions called adjust image intensity value, enhanced contrast using histogram equalization, and contrast-limited adaptive histogram equalization. We produced the two times cropped patches with random offset and random orien-
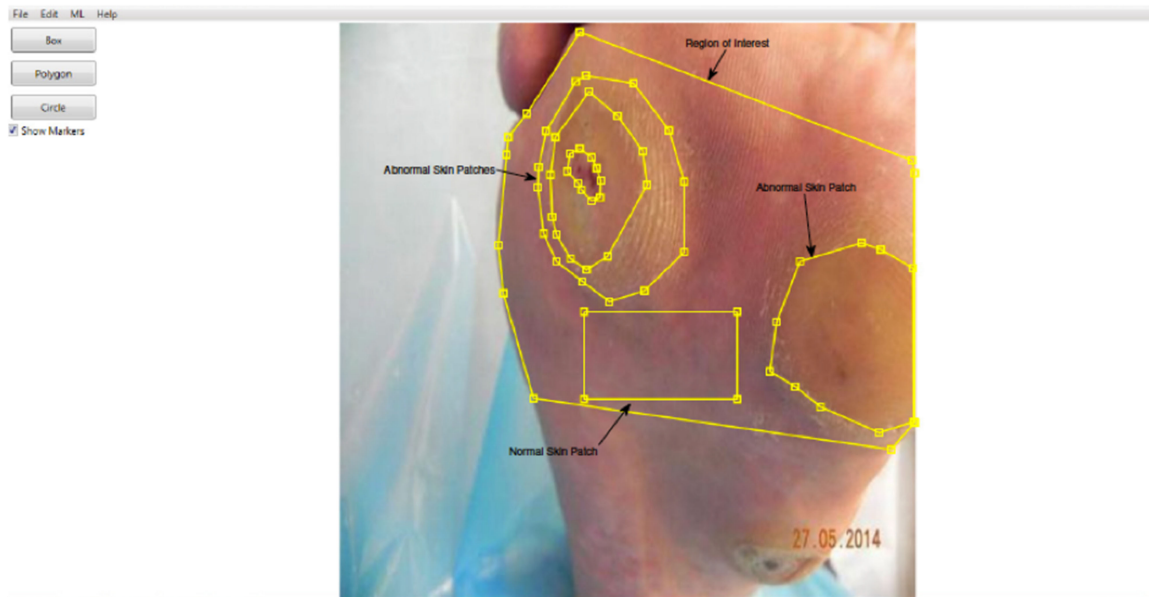
Fig. 1. An example of delineating the different regions from the whole foot image to produce abnormal and normal skin patches with annotator software [39].

tation from the original dataset of skin patches. With these techniques, we increase the number of training and validation patches by 15 times i.e. 21,345 patches for training and 1260 patches for validation.

D. Pre-Processing of Training Patches

Since, we obtained a large number of training data with data augmentation, it is essential to perform pre-processing on these patches. We used the zero-centre technique for pre-processing of these obtained patches, and then performed the normalization of every pixel.

E. Conventional Machine Learning

We investigate the use of human design features with CML on DFU and healthy skin classification. From our observation on the differences between DFU and healthy skin, the color and texture feature descriptors were the visual cues for classification. For this 2-class classification problem, the sequential minimal optimization (SMO) [40] was selected as SVM based machine learning classifier.

1) Feature Descriptors: We resize the patches of the whole dataset to 256×256 to extract the uniform color and texture feature descriptors. The three color space that we have used: *RGB*, *HSV* and *L*u*v*.

Local Binary Patterns (LBP) [41] is one of the most popular texture descriptors for the classification. In our case, the LBP features are extracted to recognize the sudden change in texture in an abnormal region of the foot for detection of DFU.

Histogram of Oriented Gradients (HOG) [42] is a manually designed feature which converts the pixel based representation into the gradient-based. In the context of this classification, HOG can be useful in terms of image gradients at an abnormal location in an image which gives you the intensity change in
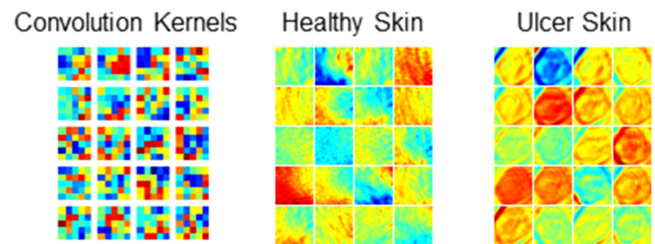

Fig. 2. The output of healthy and diabetic ulcer skin from the first convolution layer of LeNet highlight discriminative features.

that location. As the gradient is a vector quantity, it has both the magnitude and direction.

F. Convolutional Neural Networks

For comparison with the traditional features, deep learning, specifically convolutional neural networks, have been used to classify between healthy foot skin and skin with diabetic ulcerations. The first architecture we used was LeNet [43] running for 60 epochs, a learning rate of 0.01 with a step-down policy and step size of 33%, and gamma is set to 0.1. This network was originally used for recognizing digits and zip codes. These simple structures are easily recognized, even in hand-written datasets such as MNIST [44].

Using LeNet represents these structures much better than traditional features, even on a relatively small training set of 1423 patches and validation of 84 patches.

The input was 28×28 patches of skin in grayscale split into abnormal and normal skin samples. At the first convolution layer shown in Fig. 2, the kernels and activations already show the effectiveness of CNNs when highlighting important features.

We used the Caffe [45] framework to implement LeNet [43], and used the Adaptive Moment Estimation (Adam) [46] method for stochastic optimisation. This solver combines the advantages
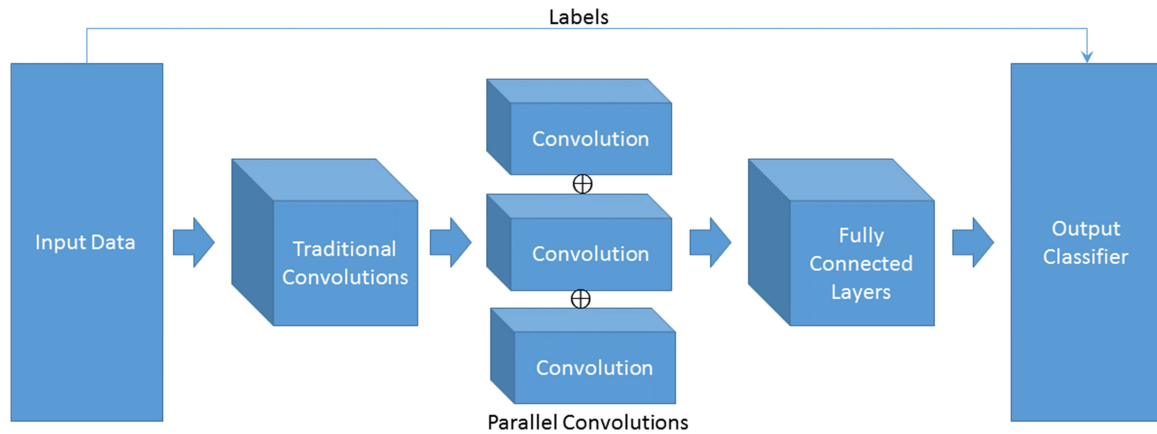
Fig. 3. An overview of the proposed DFUNet architecture. The proposed DFU architectures consists of Input Data block which consists of training and validation data, Traditional Convolution block consist of single convolutional layers, Parallel Convolution block to extract concatenated features with different convolutions, Fully Connected layers which acts as neural network and finally, Output Classifier to produce the prediction of class label.

found in AdaGrad [47], which works well with sparse gradients, and RMSProp [48], which works well in an online setting. Adam is intended for large datasets and variability in parameters. However, the results in Table IV show that smaller datasets work as effectively.

We also used popular CNN model AlexNet for classification of abnormal (DFU) and normal (healthy skin) classes. This network was originally used for classification of 1000 different objects of classes on ImageNet dataset. It emerged as the winner of ImageNet ILSVRC-2012 competition in classification category by achieving 99% confidence. There are few adjustments made in original network to work well for our 2-class classification problem. Also, a pre-trained model was used for better convergence of weights to achieve better results [34]. To train the model on Caffe framework, we used the same parameters as in LeNet i.e. 60 epochs, a learning rate of 0.01, and gamma of 0.1.

Another state-of-the-art CNN architecture that we used is GoogLeNet [27], a 22 layers deep network, with a similar experimental setting as of LeNet and AlexNet. Szegedy et al. [27] introduced a new module called inception to GoogLeNet. This acts as a multiple convolution filter inputs, which are processed on the same input and also does pooling at the same time. All the outcomes are then merged into single feature layer. This layer allows the model to take advantage of multi-level feature extraction from each input. Again, a transfer learning approach using pre-trained models to improve the performance.

G. Proposed Method - Diabetic Foot Ulcer Network

Since, this experiment focuses on the types of skin lesion which are at high risk of being misclassified by computer vision algorithms. ResNet [49], DenseNet [50], Inception [27] frameworks are very deep and computational intensive networks to work on this basic binary classification problem. The traditional CNNs such as AlexNet [33], [34] use only single type of convolutional filters popularly ranging from 1×1 to 7×7 on the input data. To improve the extraction of important features for DFU classification, we propose a new Diabetic Foot Ulcer Network

(DFUNet) architecture which is combination of important aspect of CNNs architecture - depth and parallel convolution layer. For DFUNet, we significantly decreased the depth of the network, but, we increased the size of filters in the parallel convolutional blocks to learn more features maps. DFUNet combines two types of convolutional layers i.e. traditional convolution layers at the starting of the network which use single convolutional layers followed by parallel convolutional layers, which use multiple convolutional layers for extraction of concatenated features from the same input. Detecting changes in healthy skin is a clear computer vision problem similar to malignant skin lesions, so the DFUNet is designed around convolutions to finding discriminative features for learning.

Healthy skin tends to exhibit smooth textures and DFU have many distinct features including long edges, sharp changes in intensity or color and quick changes between surrounding healthy skin and the DFU itself. DFUNet, summarised in Fig. 3, is split into three main sections: the initialisation layers inspired by GoogLeNet, parallel convolution layers to discriminate the DFU more efficiently than previous network layers and lastly, both fully-connected layers and a softmax-based output classifier. The detailed layers of the general DFUNet architecture are provided in the Table I.

The parameters used for training with DFUNet are 60 epochs, a batch size of 8, the Adam solver with a learning rate of 0.001. A step-down policy is used where the learning rate reduces with a step of 33% and gamma is set to 0.1. The Configuration of GPU Machine for Experiments is: (1) Hardware: CPU - Intel i7-6700 @ 4.00 Ghz, GPU - NVIDIA TITAN X 12 GB, RAM-32 GB DDR4 (2) Software: Caffe .

1) Input Data: The DFU training and validation images are input as 256×256 patches from areas of the feet containing DFU and healthy skin. An example of the regions of a foot cropped is shown in Fig. 4. We used the centre crop of size 224×224 and mirror as data parameters.

Initial traditional convolutional block consists of single convolution filters at each step to reduce the computational cost on feature maps. Inspired by the GoogLeNet [27] input stem, the

TABLE I
COMPLETE DESCRIPTION OF NETWORK ARCHITECTURE OF DFUNET. CONV. REFERS TO CONVOLUTIONAL LAYER, MAX-POOL. REFERS TO MAX-POOLING LAYERS. THERE ARE VARIATIONS IN FILTER SIZE OF THE PARALLEL CONVOLUTION BLOCKS OF DIFFERENT VARIANT OF DFUNET

| Layer no.	Layer type	Filter size	Stride	No. of filters	FC units	Input	Output
Layer 1	Conv.	7×7	2×2	64	-	$3\times224\times224$	$64\times112\times112$
Layer 2	Max-pool.	3×3	2×2	-	-	$64\times112\times112$	$64\times56\times56$
Layer 3	Conv.	1×1	1×1	64	-	$64\times56\times56$	$64\times56\times56$
Layer 4	Conv.	3×3	1×1	192	-	$64\times56\times56$	$192\times56\times56$
Layer 5	Max-pool.	3×3	2×2	-	-	$192\times56\times56$	$192\times28\times28$
Layer 6	Parallel conv.	$1\times1,3\times3,5\times5$	1×1	$32\oplus64\oplus128$	-	$192\times28\times28$	$224\times28\times28$
Layer 7	Max-pool.	3×3	2×2	-	-	$224\times28\times28$	$224\times14\times14$
Layer 8	Parallel conv.	$1\times1,3\times3,5\times5$	1×1	$32\oplus64\oplus128$	-	$224\times14\times14$	$224\times14\times14$
Layer 9	Parallel conv.	$1\times1,3\times3,5\times5$	1×1	$32\oplus64\oplus128$	-	$224\times14\times14$	$224\times14\times14$
Layer 10	Max-pool.	3×3	2×2	-	-	$224\times14\times14$	$224\times7\times7$
Layer 11	Parallel conv.	$1\times1,3\times3,5\times5$	1×1	$32\oplus64\oplus128$	-	$224\times7\times7$	$224\times7\times7$
Layer 12	Max-pool.	7×7	1×1	-	-	$224\times7\times7$	$224\times1\times1$
Layer 13	Fully conn.	-	-	-	1000		
Layer 14	Fully conn.	-	-	-	No. of Classes		
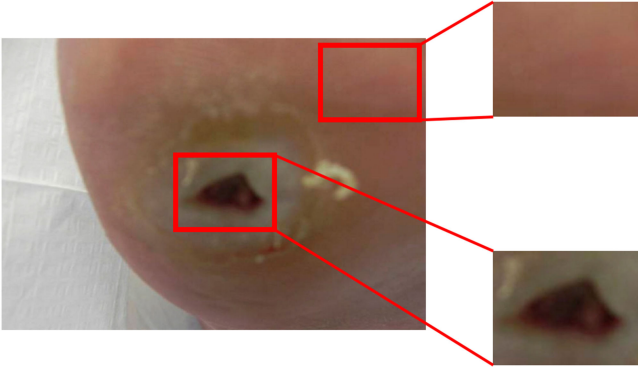
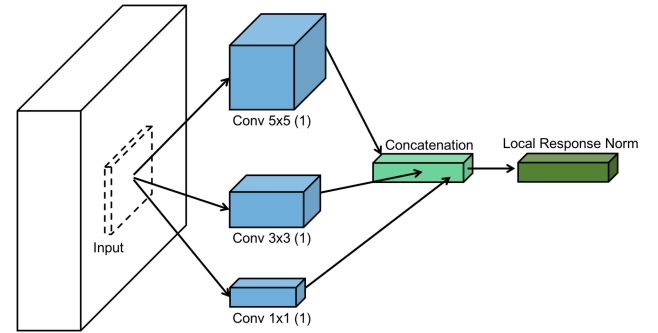Fig. 4. Healthy and ulcer patches taken from feet for training in the CNN.

Fig. 5. The structure of parallel convolution block in which three types of convolutional filters are used, concatenation layers to concatenate the features of each convolutional filters, and finally pass it local response norm layer.

input to DFUNet, begins by initial convolutions, pooling and normalisation layers in a traditional CNNs structure from layer 1 to layer 5 in Table I. Doing this step also ensures that the larger raw input image dimensionality is reduced before moving on to subsequent layers.

2) Parallel Convolutions: The idea behind using the parallel convolutional layer is basically concatenation of multiple convolution filter inputs to allow the multiple-level feature extraction and cover more spread out clusters from same input. The design of the convolutions is weighted towards creating as discriminative features as possible to highlight any DFUs in an image. Three sizes of convolution kernels are used in the parallel convolutional layers of DFUNet throughout: 5×5, 3×3 and 1×1. 1×1 convolution layer is used in the parallel convolutions to reduce the dimensionality of your input to large convolutions such as 3×3 and 5×5, thus keeping computations reasonable. These are processed in parallel to each other and finally concatenated. The core of DFUNet is the three parallel convolutions and is shown in Fig. 5. The parallel convolutional layers are key innovation in methods appears to be in the architecture of the DFUNet. As this is the one of the most significant innovation, the DFUNet is experimented with different variants of these

TABLE II
THE DESCRIPTIONS OF FILTER SIZE IN PARALLEL CONVOLUTIONAL BLOCKS OF DIFFERENT VARIANTS OF DFUNET. CONV. REFERS TO CONVOLUTIONAL LAYER AND VAR. REFERS TO VARIANT

Layers No.	DFUNet Var. 1	DFUNet Var. 2	DFUNet Var. 3	DFUNet Var. 4	DFUNet Var. 5
1st Parallel Conv.	$128\oplus256\oplus512$	$192\oplus256\oplus512$	$128\oplus128\oplus128$	$192\oplus192\oplus192$	$256\oplus256\oplus256$
2nd Parallel Conv.	$128\oplus256\oplus512$	$192\oplus256\oplus512$	$128\oplus128\oplus128$	$256\oplus256\oplus256$	$256\oplus256\oplus256$
3rd Parallel Conv.	$128\oplus256\oplus512$	$192\oplus256\oplus512$	$256\oplus256\oplus256$	$256\oplus256\oplus256$	$512\oplus512\oplus512$
4th Parallel Conv.	$128\oplus256\oplus512$	$192\oplus256\oplus512$	$256\oplus256\oplus256$	$512\oplus512\oplus512$	$512\oplus512\oplus512$

parallel sections to get the optimal architecture. We investigated the different sizes of filters in the parallel convolutions by making 5 variants to get the best variant based on the performance metrics in the Table II. We created these five variants to test the hypothesis whether increasing the size of filters improves the performance of DFUNet or not. These variants are experimented on the DFU dataset and the results are provided below in the Table III.

Each convolution provides additional discriminative power. Lower activations are present in healthy skin samples shown in

TABLE III
THE PERFORMANCE MEASURES OF VARIOUS VARIANTS OF THE DFUNET ON DFU DATASET. WHERE S.E. IS STANDARD ERROR OF AUC AND C.I. IS CONFIDENCE INTERVAL OF AUC CURVE

|	Sensitivity	Specificity	Precision	Accuracy	F-Measure	AUC Score	S.E.	95% C.I.
DFUNet Var. 1	0.923±0.029	0.910±0.037	**0.946±0.021**	0.918±0.017	0.934±0.017	0.957	0.0049	0.9481 - 0.9673
DFUNet Var. 2	0.928±0.034	0.905±0.036	0.942±0.028	0.919±0.024	0.935±0.020	0.959	0.0046	0.9499 - 0.9678
DFUNet Var. 3	0.928±0.032	0.906±0.036	0.942±0.028	0.921±0.027	0.935±0.019	0.960	0.0045	0.9518 - 0.9694
DFUNet Var. 4	0.927±0.023	0.900±0.038	0.938±0.030	0.917±0.019	0.933 ± 0.017	0.958	0.0046	0.9496 - 0.9675
DFUNet Var. 5	**0.934±0.033**	**0.911±0.044**	0.945±0.032	**0.925±0.029**	**0.939±0.024**	**0.961**	**0.0044**	**0.9520 - 0.9695**
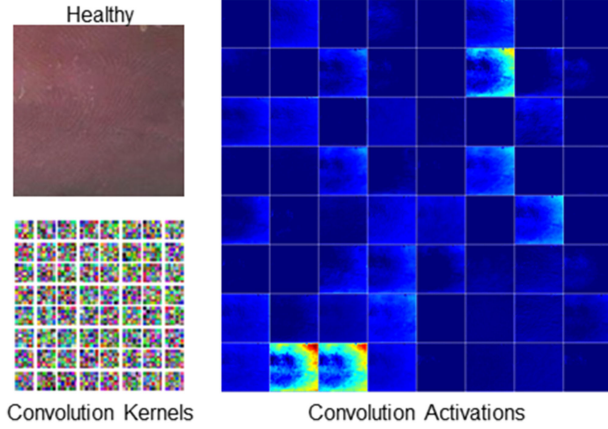
Fig. 6. The convolution activation produced by the kernels of first convolutional layer on healthy skin raw input, to highlight the features learned by convolutional layer.
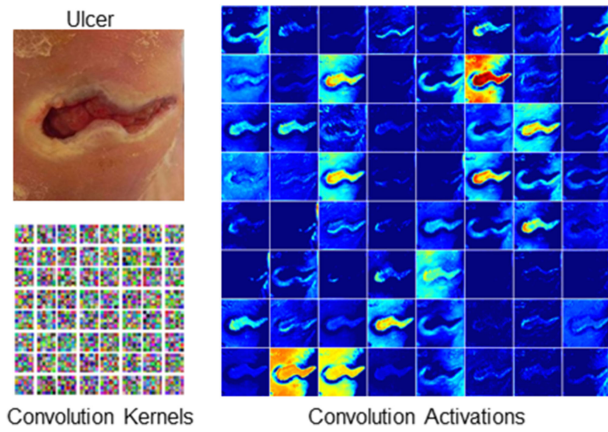

Fig. 7. The convolution activation produced by the kernels of first convolutional layer on DFU skin patch, to highlight the discriminative features learned by convolutional layer.

Fig. 6 due to the absence of skin abnormalities. Higher activations are present in skin with an ulcer as shown in Fig. 7 due to skin abnormality.

Each convolution layer uses a Rectified Linear Unit (ReLU) which is defined as

$$f(x) = \max(0, x) \tag{1}$$

where the function thresholds the activations at zero. As we use a ReLU for each convolution, they include unbounded activations,

so we use local response normalisation (LRN) to normalise these activations after each concatenation of convolutional layers. It is also proven helpful in avoiding the over-fitting problem faced by CNNs methods. Let, $a\frac{i}{x,y}$ be the source output of kernel i applied at position (x,y). Then, regularized output $b\frac{i}{x,y}$ of kernel i applied at position (x,y) is computed by

$$b\frac{i}{x,y} = a\frac{i}{x,y} \left(k + \alpha \sum_{\max(0, i-\frac{n}{2})}^{\min(N-1, i+\frac{n}{2})} (a\frac{j}{x,y})^2 \right)^{\beta} \tag{2}$$

where N is total number of kernels, n is the size of the normalization neighbourhood and $\alpha, \beta, k, (n)$ are the hyperparameters.

Further, to reduce dimensionality, a max pooling layer is included after the first and the third parallel convolutions.

3) Fully Connected Layers and Output Classifier: The final section is the softmax output of class probabilities and is a measure of how close the parameters are with respect to the ground truth labels of the training and validation data. The 2-class outputs of the DFU is healthy skin and DFU. It is formed from an average pooling layer followed by two fully connected (FC) layers with outputs of 100 for the first and 2 for the second. It is worth mentioning, the computation complexity of DFUNet is further reduced for the 2-class problem by using only 100 neurons rather than 1000 in first FC layer and last FC layer is adjusted as 2. This modification in FC layers helps in faster processing time in both training and testing phase of the DFUNet. The softmax function (cross-entropy regime) is the final layer and is defined as

$$f_j(z) = \frac{e^{z_j}}{\sum_k e^{z_k}} \tag{3}$$

where f_j is the j-th element of the vector of class scores f and z is a vector of arbitrary real-valued scores that are squashed to a vector of values between zero and one that sum to one. The loss function is defined so that having good predictions during training is equivalent to having a small loss. The final layers including fully connected layers which work as regular neural network which have connections to all activations in the previous layer, and softmax classifier, to predict the class label as either normal skin or DFU.

IV. RESULTS AND DISCUSSION

The DFU dataset was split into the 85% training, 5% validation and 10% testing sets and we adopted the 10-fold cross-

	Sensitivity	Specificity	Precision	Accuracy	F-Measure	AUC Score	S.E.	95% C.I.
LBP	0.919±0.029	0.764±0.052	0.878±0.038	0.865±0.038	0.898±0.033	0.932	0.0061	0.9202 - 0.9443
LBP + HOG	0.881±0.022	0.841±0.032	0.906±0.027	0.866±0.042	0.893±0.022	0.931	0.0060	0.9190 - 0.9427
LBP + HOG + Colour Descriptors	0.902±0.027	0.845±0.027	0.904±0.025	0.880±0.034	0.904±0.024	0.943	0.0054	0.9324 - 0.9537
LeNet (CNN)[42]	0.912±0.026	0.810±0.063	0.871±0.038	0.872±0.041	0.893±0.019	0.929	0.0050	0.9405 - 0.9603
Alexnet (CNN)[33]	0.895±0.024	0.886±0.029	0.933±0.032	0.893±0.021	0.914±0.022	0.950	0.0050	0.9405 - 0.9603
GoogLeNet (CNN)[27]	0.905±0.027	**0.912±0.052**	**0.949±0.038**	0.907±0.022	0.927±0.019	0.960	0.0045	0.9514 - 0.9690
Proposed DFUNet	**0.934±0.033**	0.911±0.044	0.945±0.032	**0.925±0.029**	**0.939±0.024**	**0.961**	**0.0044**	**0.9520 - 0.9695**

validation technique. Hence, for training, validation, and testing set using the proposed DFUNet architecture, we used approximately 1423 patches (including 882 abnormal cases), 84 patches (including 52 abnormal cases), and 172 patches (104 abnormal cases) respectively from the 397 original foot images. As mentioned previously, we used both CML models and CNNs models to do the classification task. LeNet was the only architecture that worked on 28×28 gray scale patches rather than 256×256 RGB images as input used by GoogLeNet, AlexNet, DFUNet and CML. It was included to show how the basic deep learning works on this new classification problem.

In Table IV, we report *Sensitivity*, *Specificity*, *Precision*, *Accuracy*, *F-Measure* and *Area under curve of ROC (AUC)* as our evaluation metrics. In medical imaging, *Sensitivity* and *Specificity* are considered reliable evaluation metrics for classifier completeness.

$$Sensitivity = \frac{TP}{TP + FN} \quad (4)$$

$$Specificity = \frac{TN}{FP + TN} \quad (5)$$

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

$$Accuracy = \frac{TP + FN}{TP + TN + FP + FN} \quad (7)$$

$$F - Measure = \frac{2 * TP}{2 * TP + FP + FN} \quad (8)$$

In Table III, we report the performance measures of various DFUNet variants with different parameters as explained in the architecture of DFUNet in the previous section. There was not much gap in performances between all the models. But, overall, the DFUNet variant 5 performed best in every evaluation metrics except *Precision* in which DFUNet variant 1 performed the best. It also proved the earlier hypothesis correct as increasing the size of filters in the parallel convolution layers improved the performance of DFUNet. Hence, DFUNet variant 5 which uses the much larger filter sizes than other variants in last two parallel convolutional layers produced better results. Hence, with best results achieved by DFUNet variant, we used it as a proposed DFUNet to compare the performance with other traditional machine learning and deep learning models. ROC curve for all the variants is illustrated by Fig. 8.
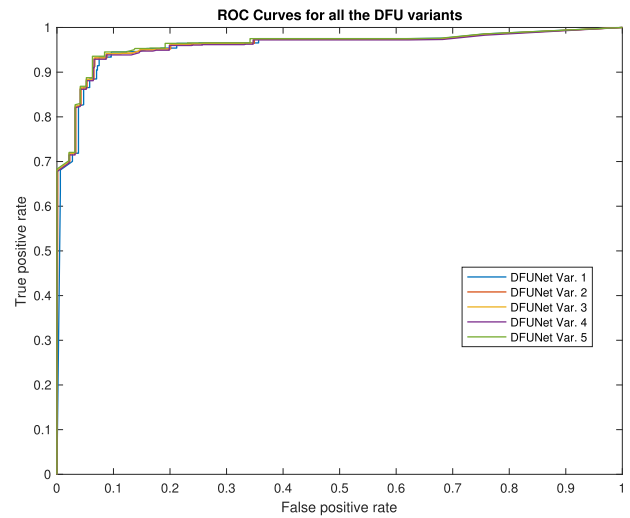

Fig. 8. The ROC curve for all DFUNet models as mentioned in Table III, DFUNet var. 5 performed best with an AUC score of 0.961. Var. refers to variant.

There are three CML models and three CNNs models used for classification. In CML, we used the combination of LBP, HOG and Colour descriptors (*RGB*, *HSV* and *L*u*v*) as feature vectors and then, we trained an SMO for our classification problem. For each CNN, LeNet, AlexNet, GoogLeNet and our proposed DFUNet are the chosen architectures used for classification. Each classifier performed well for *Sensitivity* with less than 1.4% margin between the highest result (DFUNet) and the lowest result (LBP + HOG). There is a more significant gap of 7.7% in *Specificity* for the CML models performance measure, with results ranging from 0.835 to 0.845.

For the CNNs approaches, LeNet achieved the lowest score of 0.81 for *Specificity*, whereas the AlexNet, GoogLeNet and DFUNet performed best in this category, with 0.892, 0.912, and 0.908 respectively. AUC is considered to be a viable performance measure for the different machine learning approaches for classification, with DFUNet and GoogLeNet achieving 0.961 and 0.960 respectively.

Overall, we showed that using CNNs can outperform the more traditional CML features by a large margin. All CNN architectures achieved higher results than any of the CML results in most cases. GoogLeNet and DFUNet were the best perform-
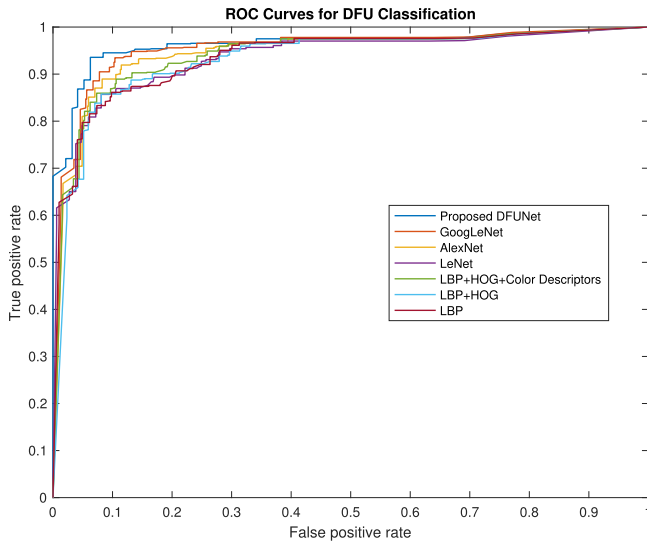
Fig. 9.   ROC Curve for all the Models Including CML and CNNs Mentioned in Table IV in Which our Proposed DFUNet Method Achieved the Best AUC Score.

ers for various evaluation metrics among all the classifiers. The ROC curve for all the models is demonstrated by the Fig. 9. The details of AUC performance for each method is described in Table IV.

We received better results than GoogLeNet on various evaluation metrics. The reason behind using the DFUNet rather than conventional CNNs architecture, in particular, GoogLeNet is to speed up the best results with lesser layers i.e. 14 layers architecture compared to the 22 layers architecture of GoogLeNet. We also reduce the number of neurons in the FC layers to improve processing time of DFUNet according to the 2-class problem. With the 10-fold cross validation, on the same machine configuration and input batch size on Caffe framework, DFUNet took an average of 3 minutes 32 seconds where as GoogLeNet took average of 16 minutes 27 seconds to train a model with the same amount of training and validation data. For testing, DFU-Net took an average of 49 seconds whereas GoogLeNet took an average of 72 seconds to classify the same test data. Therefore, we demonstrated how reducing the number of layers using the bespoke architecture of DFUNet markedly reduced processing time, while also achieving higher sensitivity and specificity with introduction of parallel convolution layers with increased number of filter input.

Our proposed DFUNet has highest performance measures in *Sensitivity*, with a score of 0.934, *F-measure* with 0.939 and *AUC* with 0.962. Whereas, GoogLeNet has highest score in *Specificity* and *Precision* due to it's robust nature of being able to find more subtle changes using the inception architecture [27].

With data augmentation technique, these patches were made 15 times for both training and validation. But, when we tested the data augmentation training in our experiment, there was no differences found in performance metrics with all the models. Hence, we did not include the data augmentation results in Table III and Table IV as it did not improve the results. The main reasons behind the failure of data augmentation were overall

performance metrics recorded without data augmentation was quite high and there was only small number of misclassification cases which were not corrected even with models trained with data augmentation. For example, we found in evaluations, some ulcer conditions with similar skin tone and ulcers of small size shown in second row in Fig. 10 have too subtle features to be detected as ulcer regardless of any pre-processing with data-augmentation. Hence, we did not use data augmentation to produce final results as the training with data augmentation become more computational expensive because of 15 times more data than the normal dataset. Also, the focus of this work is to determine the skin lesions are at high risk to be detected as misclassification.

There is no evidence of an influence of factors such as lighting conditions and skin tone due to patient's ethnicity on DFU classification. As ulcer and surrounding skin has quite distinctive texture and color features from the normal healthy skin irrespective of above-mentioned factors. In our experiments, these factors result in very few misclassified instances in testing set when there is very high red skin tone as shown in Fig. 10.

A. Experimental Analysis and Discussion

Diagnosis and detection of DFU by the computerized method has been an emerging research area with the evolution of computer vision, especially deep learning methods. This preliminary experiment of binary classification of DFU and healthy skin is performed to learn the distinctive features of both types of skin lesions. Also, the main motivation of this experiment to find the type of skin lesions which are at high risk of being misclassified by algorithms. In this experiment, we proposed a new lightweight deep learning architecture which can classify DFU and healthy skin lesions with high accuracy. There are a few examples of correctly and incorrectly classified cases in both abnormal and normal classes by DFUNet as illustrated in Fig. 10. The computer vision algorithms struggle to classify the very subtle DFU with similar skin tone correctly. They are detected as normal with high percentage as illustrated by example 1 and 2 of misclassification cases of abnormal class in Fig. 10. Also, DFU that are very small in size is misclassified as normal as shown by example 3 and 4 of misclassification cases of abnormal class Fig. 10. In normal skin, the patches with toe, highly wrinkled skin, and very high red tone skin are classified wrongly by the proposed method as illustrated by the examples of misclassified cases of normal classes in Fig. 10.

V. PERFORMANCE EVALUATION ON HETEROGENEOUS TEST CASE

Since, DFU dataset is captured with the same DSLR camera as mentioned in above section. With computer vision techniques, it is preferable to have heterogeneous capture to form dataset. But, strict medical ethical approval does not allow to use different cameras to capture the pictures of DFU in the healthcare setting. Hence, we collected another heterogeneous dataset of standardized DFU images with FootSnap application. These images are captured with iPad camera. We tested our algorithm on this heterogeneous dataset and received good performance with
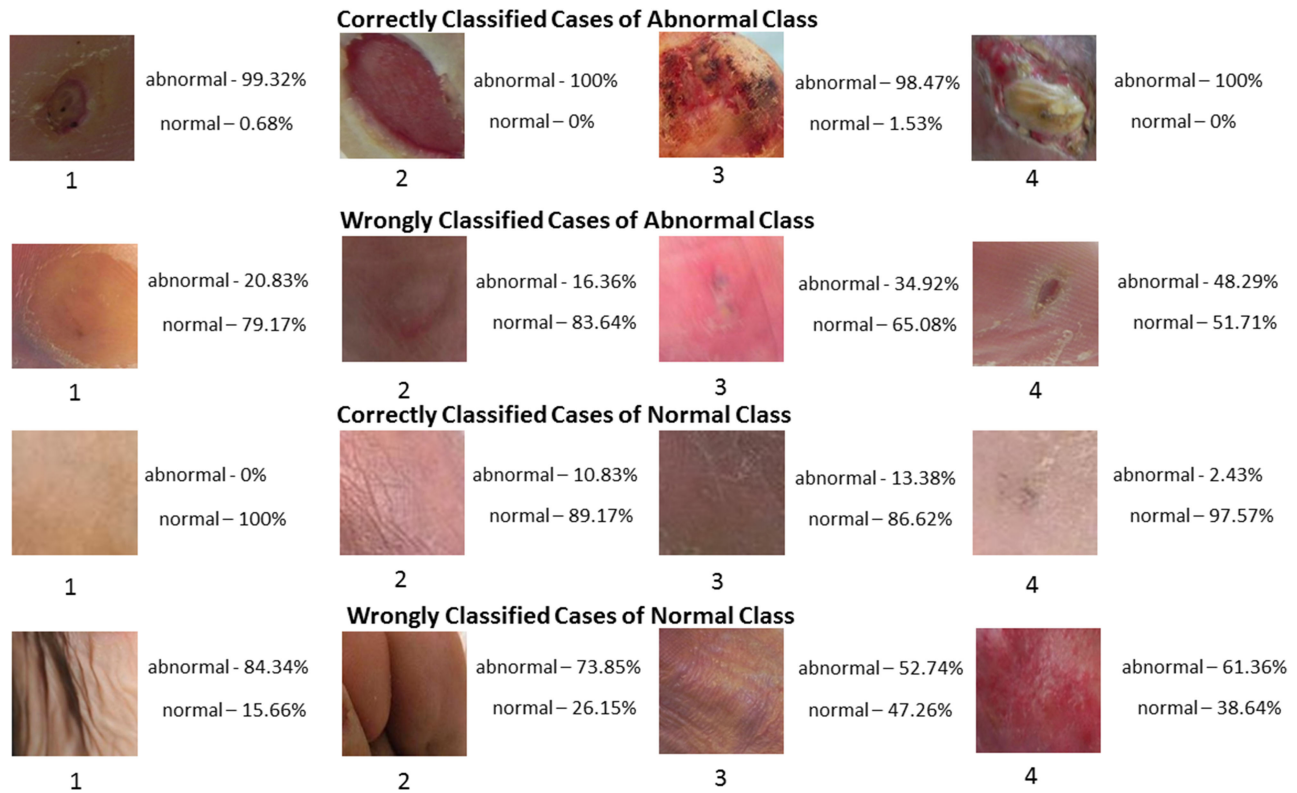
Fig. 10.   Few examples of accurate and inaccurate classified cases for both abnormal and normal classes with DFUNet.

Fig. 11.   The examples of three classes in facial skin dataset.

TABLE V

FACIAL SKIN CLASSIFICATION TASK WITH THREE CLASSES AS NORMAL SKIN, SPOT, WRINKLE. THE PROPOSED DFUNET OUTPERFORMED GOOGLENET IN EVERY PERFORMANCE METRICS ON THIS DATASET

	Sensitivity	Specificity	Precision	Accuracy	F-Measure	MCC
GoogLenet	0.783	0.882	0.784	0.846	0.784	0.665
Proposed DFUNet	**0.867**	**0.930**	**0.867**	**0.907**	**0.867**	**0.796**

Sensitivity score of 0.929, *F-measure* with 0.931, *Specificity* of 0.908, *Precision* with 0.942 and *AUC* with 0.950 score.

VI.   PERFORMANCE EVALUATION ON FACIAL SKIN DATASET

Since, DFUNet performed well on the classification of DFU skin patches, to test the robustness of DFUNet on other skin lesion datasets, we run the experiment of 3-class classification of facial skin patches i.e. normal, spot and wrinkles as shown in the Fig. 11. It is worth mentioning, there is no public skin lesion dataset available for research without prior written consent. In this derma dataset, we delineated the equal number of skin patches i.e. 110 patches for each class. We used only two best performing CNN architectures in Table. IV i.e. GoogLeNet and DFUNet for this experiment. With the same experimental settings, DFUNet outperforms GoogLeNet in each evaluation metrics for 10-fold cross-validation data as shown in Table V. This is due to the deep learning models does not work well with smaller dataset even with full training [37]. But, DFUNet

uses larger filter sizes in the later parallel convolution layers to extract more multiple features which help DFUNet outperforms GoogLeNet in this experiment.

VII.   CONCLUSION AND FUTURE WORK

In this work, we trained various classifiers based on traditional machine learning algorithms, CNNs and proposed a new CNN architecture, DFUNet on DFU classification which discriminates the DFU skin from healthy skin. With high-performance measures in classification, DFUNet allows the accurate automated detection of DFU in foot images and make it an innovative technique for DFU evaluation and medical treatment. For the detection of DFU, it is vital to understand the difference between DFU and healthy skin to know the features differences between these two classes in computer vision perspective. This work has a potential for technology that may transform the detection and treatment of diabetic foot ulcers and lead to a paradigm shift in the clinical care of the diabetic foot. This work has formed the basis to achieve future targets that include: 1) developing the automatic annotator that can automatically

delineate and classify the foot images without the help of clinicians; 2) developing the automatic ulcer detection, recognition and segmentation with the help of these classifiers; 3) implementing the method to determine the various pathologies of DFU as multi-class classification similar to the Texas classification and other grading scales; 4) implementing the various user-friendly software tools including mobile applications for ulcer recognition [51]. Since DFUNet worked well for DFU classification, this proposed framework will likely be used for classifying the other skin lesions such as wound classification, infections like chicken pox or shingles, other skin lesions like moles and freckles, spotting marks and pimples [52] against the normal skin. For classification, DFUNet is a light-weight CNN framework that is used for DFU dataset consists of two classes (ulcer and normal skin) and facial skin dataset consists of three classes (spot, wrinkles and normal skin), it will be further tested in the future to include many more classes. Therefore, we demonstrated how reducing the number of layers and number of neurons in FC layers using the bespoke architecture of DFUNet markedly reduced processing time, while also achieving higher sensitivity and specificity.

REFERENCES

[1] S. Wild, G. Roglic, A. Green, R. Sicree, and H. King, "Global prevalence of diabetes estimates for the year 2000 and projections for 2030," *Diabetes Care*, vol. 27, no. 5, pp. 1047–1053, 2004.

[2] W. H. Organization *et al.*, Global report on diabetes who Geneva, Geneva, Switzerland, 2016.

[3] K. Bakker, J. Apelqvist, B. Lipsky, J. Van Netten, and N. Schaper, "The 2015 IWGDF guidance documents on prevention and management of foot problems in diabetes: Development of an evidence-based global consensus," *Diabetes/Metabolism Res. Rev.*, vol. 32, no. S1, pp. 2–6, 2016.

[4] A. J. Boulton, L. Vileikyte, G. Ragnarson-Tennvall, and J. Apelqvist, "The global burden of diabetic foot disease," *Lancet*, vol. 366, no. 9498, pp. 1719–1724, 2005.

[5] F. Aguiree *et al.*, *IDF Diabetes Atlas*, 6th ed., L. Guariguata, T. Nolan, J. Beagley, U. Linnenkamp, and O. Jacqmain, Eds. Brussels, Belgium: International Diabetes Federation, 2013.

[6] D. G. Armstrong, L. A. Lavery, and L. B. Harkless, "Validation of a diabetic wound classification system: The contribution of depth, infection, and ischemia to risk of amputation," *Diabetes Care*, vol. 21, no. 5, pp. 855–859, 1998.

[7] P. Cavanagh, C. Attinger, Z. Abbas, A. Bal, N. Rojas, and Z.-R. Xu, "Cost of treating diabetic foot ulcers in five different countries," *Diabetes/Metabolism Res. Rev.*, vol. 28, no. S1, pp. 107–111, 2012.

[8] L. A. Lavery, D. G. Armstrong, and L. B. Harkless, "Classification of diabetic foot wounds," *J. Foot Ankle Surg.*, vol. 35, no. 6, pp. 528–531, 1996.

[9] S. Franc *et al.*, "Telemedicine: What more is needed for its integration in everyday life?" *Diabetes Metabolism*, vol. 37, pp. S71–S77, 2011.

[10] O. El-Gayar, P. Timsina, N. Nawar, and W. Eid, "A systematic review of it for diabetes self-management: Are we there yet?" *Int. J. Med. Inf.*, vol. 82, no. 8, pp. 637–652, 2013.

[11] O. Erkaymaz, M. Ozer, and M. Perc, "Performance of small-world feedforward neural networks for the diagnosis of diabetes," *Appl. Math. Comput.*, vol. 311, pp. 22–28, 2017.

[12] J. Clemensen, S. B. Larsen, M. Kirkevold, and N. Ejskjaer, "Treatment of diabetic foot ulcers in the home: Video consultations as an alternative to outpatient hospital care," *Int. J. Telemed. Appl.*, vol. 2008, 2008, Art. no. 1.

[13] F. L. Bowling *et al.*, "Remote assessment of diabetic foot ulcers using a novel wound imaging system," *Wound Repair Regeneration*, vol. 19, no. 1, pp. 25–30, 2011.

[14] C. E. Hazenberg, J. J. van Netten, S. G. van Baal, and S. A. Bus, "Assessment of signs of foot infection in diabetes patients using photographic foot imaging and infrared thermography," *Diabetes Technol. Therapeutics*, vol. 16, no. 6, pp. 370–377, 2014.

[15] P. Foltynski *et al.*, "Monitoring of diabetic foot syndrome treatment: Some new perspectives," *Artif. Organs*, vol. 35, no. 2, pp. 176–182, 2011.

[16] C. Liu, J. J. van Netten, J. G. van Baal, S. A. Bus, and F. van Der Heijden, "Automatic detection of diabetic foot complications with infrared thermography by asymmetric analysis," *J. Biomed. Opt.*, vol. 20, no. 2, 2015, Art. no. 026003.

[17] J. J. van Netten, M. Prijs, J. G. van Baal, C. Liu, F. van Der Heijden, and S. A. Bus, "Diagnostic values for skin temperature assessment to detect diabetes-related foot complications," *Diabetes Technol. Therapeutics*, vol. 16, no. 11, pp. 714–721, 2014.

[18] L. Wang, P. Pedersen, E. Agu, D. Strong, and B. Tulu, "Area determination of diabetic foot ulcer images using a cascaded two-stage SVM based classification," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 9, pp. 2098–2109, Sep. 2017.

[19] M. Goyal, M. H. Yap, N. D. Reeves, S. Rajbhandari, and J. Spragg, "Fully convolutional networks for diabetic foot ulcer segmentation," in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, 2017, pp. 618–623.

[20] H. Wannous, Y. Lucas, and S. Treuillet, "Enhanced assessment of the wound-healing process by accurate multiview tissue classification," *IEEE Trans. Med. Imag.*, vol. 30, no. 2, pp. 315–326, Feb. 2011.

[21] M. Kolesnik and A. Fexa, "Multidimensional color histograms for segmentation of wounds in images," in *Proc. Image Anal. Recognit.*, 2005, pp. 1014–1022.

[22] M. Kolesnik and A. Fexa, "How robust is the SVM wound segmentation?" in *Proc. 7th Nordic Signal Process. Symp.*, Jun. 2006, pp. 50–53, doi: 10.1109/NORSIG.2006.275274.

[23] E. S. Papazoglou, L. Zubkov, X. Mao, M. Neidrauer, N. Rannou, and M. S. Weingarten, "Image analysis of chronic wounds for determining the surface area," *Wound Repair Regeneration*, vol. 18, no. 4, pp. 349–358, 2010.

[24] F. Veredas, H. Mesa, and L. Morente, "Binary tissue classification on wound images with neural networks and Bayesian classifiers," *IEEE Trans. Med. Imag.*, vol. 29, no. 2, pp. 410–427, Feb. 2010.

[25] M. H. Hermans, "Wounds and ulcers: Back to the old nomenclature," *Wounds*, vol. 22, no. 11, pp. 289–293, 2010.

[26] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 818–833.

[27] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 1–9.

[28] M. Anthimopoulos, S. Christodoulidis, L. Ebner, A. Christe, and S. Mougiakakou, "Lung pattern classification for interstitial lung diseases using a deep convolutional neural network," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1207–1216, May 2016.

[29] H.-C. Shin *et al.*, "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1285–1298, May 2016.

[30] M. H. Yap *et al.*, "End-to-end breast ultrasound lesions recognition with a deep learning approach," *Proc. SPIE Med. Imag.: Biomed. Appl. Mol., Struct., Functional Imag.*, 2018, vol. 10578, Art. no. 1057819.

[31] M. H. Yap *et al.*, "Automated breast ultrasound lesions detection using convolutional neural networks," *IEEE J. Biomed. Health Informat.*, vol. 22, no. 4, pp. 1218–1226, Jul. 2018.

[32] M. Goyal and M. H. Yap, "Multiclass semantic segmentation of skin lesions via fully convolutional networks," *arXiv preprint arXiv:1711.10449*, 2017.

[33] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[34] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[35] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[36] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 3431–3440.

[37] N. Tajbakhsh *et al.*, "Convolutional neural networks for medical image analysis: Full training or fine tuning?" *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1299–1312, May 2016.

[38] M. H. Yap *et al.*, "Computer vision algorithms in the detection of diabetic foot ulceration a new paradigm for diabetic foot care?" *J. Diabetes Sci. Technol.*, vol. 10, pp. 612–613, 2016.

[39] B. M. Hewitt, M. H. Yap, and R. A. Grant, "Manual whisker annotator (MWA): A modular open-source tool," *J. Open Res. Softw.*, vol. 4, no. 1, 2016, p. e16, doi: http://doi.org/10.5334/jors.93.

[40] J. C. Platt, "Fast training of support vector machines using sequential minimal optimization," in *Advances in Kernel Methods*, 1999, pp. 185–208.

[41] D.-C. He and L. Wang, "Texture unit, texture spectrum, and texture analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 28, no. 4, pp. 509–512, Jul. 1990.

[42] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 1, 2005, pp. 886–893.

[43] Y. LeCun *et al.*, "Convolutional networks for images, speech, and time series," in *The Handbook of Brain Theory and Neural Networks*, vol. 3361, 1995.

[44] Y. LeCun, C. Cortes, and C. J. Burges, "The MNIST database of handwritten digits," 1998. [Online]. Available: http://yann.lecun.com/exdb/mnist/

[45] Y. Jia *et al.*, "Caffe: Convolutional architecture for fast feature embedding," in *Proc. 22nd ACM Int. Conf. Multimedia*, 2014, pp. 675–678.

[46] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[47] J. Duchi, E. Hazan, and Y. Singer, "Adaptive subgradient methods for online learning and stochastic optimization," *J. Mach. Learn. Res.*, vol. 12, pp. 2121–2159, 2011.

[48] T. Tieleman and G. Hinton, "Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude," *COURSERA: Neural Netw. Mach. Learn.*, vol. 4, no. 2, pp. 26–31, 2012.

[49] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.

[50] F. Iandola, M. Moskewicz, S. Karayev, R. Girshick, T. Darrell, and K. Keutzer, "Densenet: Implementing efficient convnet descriptor pyramids," *arXiv preprint arXiv:1404.1869*, 2014.

[51] M. H. Yap *et al.*, "A new mobile application for standardizing diabetic foot images," *J. Diabetes Sci. Technol.*, vol. 12, no. 1, pp. 169–173, 2018.

[52] J. Alarifi *et al.*, "Facial skin classification using convolutional neural networks," in *Proc. 14th Int. Conf. Image Anal. Recognit.*, Montreal, QC, Canada, 2017, vol. 10317, pp. 479–485.

Adrian K. Davison received the B.Sc. (Hons.) degree in multimedia computing in 2012, and the Ph.D. degree from Manchester Metropolitan University, Manchester, U.K., in 2016. He is a Research Associate with the University of Manchester. He had an active role as a student representative within MMU, Co-Chairing the internal MMU Science and Engineering Symposium 2015. His current role is the medical imaging analysis of children's hip disease. His research expertise is in facial expression and microfacial expression analysis, deep learning, and medical image analysis.

Satyan Rajbhandari received the M.D. degree from FRCP, Sheffield and Edinburgh, U.K. He is a Consultant Physician and a Honorary Clinical Professor with Lancashire Teaching Hospitals, Preston, U.K. His specialist areas are Diabetes, endocrinology, and general medicine. His research interests inlcude diabetes treatment, cardiovascular risks in diabetes, diabetic neuropathy, diabetic foot, and epidemiology. He was the recipient of various awards and prizes, such as the Arun Baksi Award for close co-operation between primary and secondary care diabetes service, the Young Researcher Award, European Association of Study of Diabetic Neuropathy (2005), the Sir Ernest Finch Travel Award for Young Researcher (2000), the Award by Nepal Govt., 1988 for management of mass casualty, Colombo Plan Scholarship (1981–1986), Mahendra Ratna Scholarship (1978–1980), and highest scorer in national GCSE level (Nepal).

Jennifer Spragg received the Graduate degree from the University of Salford, Manchester, U.K., in 2005, and Registered with the Health Professions Council (HPC). She is a Podiatrist/ Chiropodist with the Rossendale Practice, Rawtenstall, U.K. She is also a member of the Society of Chiropodists and Podiatrists. She has 25 years experience as a Registered General Nurse (RGN)—last 15 years as Practice Nurse in GP surgeries and has a wide range of knowledge of medical conditions and treatments.

Manu Goyal received the Master of Technology degree in computer science and applications from Thapar University, Patiala, India. He is a Research Scholar with the Manchester Metropolitan University, Manchester, U.K. His research expertise is in medical imaging analysis, computer vision, deep learning, wireless sensor networks, and Internet of things.

Moi Hoon Yap received the Ph.D. degree in computer science from Loughborough University, Loughborough, U.K., in 2009. She is a Reader (Associate Professor) in computer vision with the Manchester Metropolitan University, Manchester, U.K., and a Royal Society Industry Fellow with Image Metrics Ltd., Manchester, U.K. After her Ph.D., she was a Postdoctoral Research Assistant (April 09–October 11) with the Centre for Visual Computing, the University of Bradford. Her research expertise is computer vision, machine learning, and deep learning. She was an Associate Editor for the *Journal of Open Research Software* and a reviewer for the IEEE TRANSACTIONS ON IMAGE PROCESSING, IEEE TRANSACTIONS ON MULTIMEDIA, IEEE TRANSACTIONS ON CYBERNETICS, and IEEE JOURNAL OF BIOMEDICAL HEALTH AND INFORMATICS.

Neil D. Reeves is a Professor of musculoskeletal biomechanics with the Research Centre for Musculoskeletal Science & Sports Mediciné, School of Healthcare Science. He conducts research into gait, movement impairment, and foot pathology in diabetes and has published in journals such as *Diabetes Care* and *Diabetic Medicine*. He has led research projects on these topics supported by the European and U.K. funding bodies.