Diabetic Foot Ulcer Classification using Mapped Binary Patterns and Convolutional Neural Networks

Nora Al-Garaawi[a,*], Raja Ebsim[b], Abbas Alharan[a] and Moi Hoon Yap[c]

[a]Department of Computer Science, Faculty of Education for Girls, University of Kufa, Najaf, Iraq
[b]Centre for Imaging Sciences, The University of Manchester, Manchester, UK
[c]Centre for Advanced Computational Science, Manchester Metropolitan University, Manchester, UK

ARTICLE INFO

Keywords:
Diabetic foot ulcers
Convolutional neural networks
Diabetic foot ulcers classification
Mapped local binary patterns
Feature fusion

ABSTRACT

Diabetic Foot Ulcer (DFU) is one of the major complications of Diabetes, which leads to lower limb amputation if not treated early and properly. In addition to the clinical diagnostic services provided by all medical staff, recently, research in automation using computer vision and machine learning methods play an important role in DFU classification. Most recent automatic approaches to DFU classification are based on Convolutional Neural Networks (CNNs) using RGB images only as input. In this paper, we present a new CNN-based DFU classification method, in which we show that feeding an appropriate feature (texture information) to the CNN models provide a complementary performance to the standard RGB-based deep models of DFU classification task and better performance can be obtained if both RGB images and its texture features are combined together and used as input to the CNN. To this end, our proposed method consists of two main stages. The first stage extracts the texture information from the RGB image using Mapped Binary Patterns technique. The obtained mapped image is used to aid the second stage in recognizing DFU as it contains texture information of ulcer. The stack of RGB image and Mapped Binary Patterns can be fed to the CNN as a tensor input, or as a fused image which is a linear combination of RGB and Mapped Binary Patterns images. Extensive experiments on DFU dataset show that the proposed methods provide better performance than the state-of-the-art CNN-based methods that use the RGB images only with an area under the receiver operator characteristic curve (AUC) of 0.981 and F-Measure of 0.952.

1. Introduction and Background

Diabetic Foot Ulcers (DFU) may results in lower extremity amputation, which is a major complication of Diabetes [46]. One way of helping to identify and treat DFU is to build an automatic (computer-based) DFU diagnostic system which can be beneficial not only in early DFU detection but also in reducing clinical workloads, cost-efficacy, standardizing treatment, improving patient care and could reduce the number of misdiagnoses. Building such a diagnostic system can be done using several computer vision and machine learning approaches including convolutional neural networks (CNNs) and non-CNNs approaches. Although both approaches are applied to the DFU problem and the former provided better performance than the later, the automatic DFU diagnostic system is still in its infancy. In this paper, we show that non-CNNs approaches, where the appropriate information regarding the DFU disease can be found, may still important and fuse it with CNNs approaches can enhance the performance of the automatic DFU classification system.

Broadly speaking there are many researchers that focused on diagnosis the DFU from digital images using several computer vision and machine learning algorithms into two schemas. The first schema is by using conventional approaches, where handcrafted such as texture features are used for image representation. Many texture feature extraction techniques were introduced in the literature and used for texture image classification issues including medical and non-medical applications [2, 3, 32, 21, 16]. Local Binary Patterns (LBP) [33] is one of the most successful texture description approaches in several medical and non-medical applications, because of its easy implementation, invariance to rotation and robustness to monotonic illumination changes. The process of capturing the spatial structure of a texture pattern using LBP method is done by describing the pixel neighbourhood using its binary codes which are used then to form a local binary pattern code. In DFU classification, the development of conventional-based DFU classification system was first proposed by Goyal et al. [18]. The authors measured the variation between DFU and healthy skin by training a classifier on the texture and colour information. In their study, they used LBP [22] and Histogram of Oriented Gradients (HOG) [13] for feature extraction and Support Vector Machine (SVM) for healthy versus unhealthy skin classification. The SVMs were trained and test using sequential minimal optimization (SMO) algorithm [34].

The second schema is by using CNNs approaches, where deep features are used for the image representation leading to the development of CNNs-based DFU detection and classification methods [5, 9, 14, 20, 25, 43, 4, 12, 18, 19]. Some of the previously mentioned studies have focused only on using deep features for building DFU detection methods [5, 9, 20, 25, 43, 47]. Such a model is a critical requirement for medical problems (e.g.DFU) since it is used to localize the appropriate unhealthy (DFU) region for subsequent modules such as feature extraction and classification. The other studies [4, 5, 12, 18, 19] have focused on using deep representa-

*Corresponding author
nora.algaaawi@uokufa.edu.iq (N. Al-Garaawi)
ORCID(s): 0000-0001-7039-6037 (N. Al-Garaawi)

tion for the development of DFU classification for classifying healthy versus unhealthy skin region, which is the focus of this paper. Goyal et al. [18] proposed a novel CNN architecture called DFUNet and shared a dataset. The authors first used a manual whisker annotator, an open source annotator (MWA) [23], to outline the healthy and DFU patches from the original foot image. These patches were then used to train the proposed network for binary classification of healthy versus unhealthy (DFU) classes. Experimental results on their DFU dataset demonstrated that the proposed network has successfully outperformed the performance of some popular networks, including GoogLeNet [44] and AlexNet [29] with 94.5%, 93.4% and 93.9% of precision, recall and F-Measure respectively. Later in 2020, Alzubaid et al. [4] designed a new CNN architecture called DFU-QTNet for automatic recognition of healthy class versus unhealthy (DFU) class. The authors first cropped the healthy and unhealthy (DFU) patches from the original foot images and labelled them manually by medical Experts. These patches were used to train the proposed DFU-QTNet. Due to gradient, they found that increasing the width of the network with keeping the depth comparable to the traditional neural networks helped to increase the overall performance, whereas, increasing the number of network layers has decreased the overall performance. Experimental results on DFUNet dataset [18] demonstrated that the proposed network has successfully outperformed the performance of GoogLeNet [44], AlexNet [29] and DFUNet [18] with 95.4%, 93.6% and 94.5% of precision, recall and F-Measure, respectively.

The CNN approaches are gaining more and more attention as they are successfully applied to many image processing and computer vision tasks, providing better performance than the non-CNN approaches. DFU classification tasks are not the exceptions, for example, the CNNs in [18] provide better DFU classification performance than the conventional methods such as LBP [22] and HOG [13]. To the best of the authors' knowledge, the majority of CNN-based studies in DFU classification used the RGB image (not the handcrafted features) as the input, and they learn and extract the features from the training data without human intervention.

In spite of the usefulness of deep features in DFU classification, texture analysis on the other hand is a useful way of increasing the information obtainable from medical images as it contains texture features regarding the disease. These features are, in fact, mathematical parameters computed from the distribution of pixels, which characterize the texture type and thus the underlying structure of the objects shown in the image. This information is an important feature for any diagnostic system since different types of disease have different characteristics in texture distribution [7, 10, 11, 22]. In case of DFU problem, the skin of healthy foot usually display smooth textures whereas the skin of diabetic foot with ulcer tends to exhibit distinct features including skin color changes, intensity changes, large edges and quick changes between surrounding normal skin and the ulcer.

Recently, hybrid approaches, the combination of convolutional and CNN approaches is applied in some computer vision applications such as medical and non-medical images. The hybrid approaches attempt to utilize the advantages that come from both approaches. The convolutional approaches is often used to capture the discriminative features that characterize the texture type, these feature can be used to enhance the performance of CNN system by trying to feed more information of the texture type. A common way of this combination is by extract the hand crafted feature vector using any convolutional method and combine it with last convolutional layer features and the obtained feature vector is used as a input feature for a classifier to classify the texture type as in [38, 26, 1, 36]. Because each part of this method is individually trained, thus it cannot be benefited from end-to-end learning.

Interestingly, in a recent performance evaluation for several other computer vision tasks [6, 24, 30], the CNNs models trained on the handcrafted features only or in combination with the RGB images were shown to achieve competitive performance than the CNNs models trained on the RGB images only. For example, an ensemble of CNN models trained on RGB and LBP mapped coded images for emotion recognition [30], texture recognition [6], and remote sensing scene classification [6]. In addition, [24] used a weighted sum of RGB images and Gabor responses images is fed the CNN for age estimation, gender classification, face detection, and facial expression recognition. Inspired by these research, and with the great recent success of deep learning and the importance of the texture features in medical imaging, we propose to combine the importance of texture coded images within the deep learning framework to investigate its potential in DFU classification.

Contribution: Motivated by the above observations, this paper introduces a new framework for DFU classification. Since the texture features contain special and important information regarding the DFU disease and for compact representation, our contribution is achieved by training the CNN model of our system on both the texture features and RGB images jointly. Precisely, we propose a method to get the benefits of LBP, together with the features that are learned by CNN with the input images. In other words, we extract several LBP responses and concatenate them with the input image. This can also be considered as a fusion of input image and LBP responses, which looks like an image with enhanced textures and the fused image is fed to the CNN. The advantage of using a combination of all representations is that more information about the disease can be obtained, and the disease whose appearance similar to healthy skin are dealt with more effectively. We train several CNN models to distinguish between healthy and unhealthy (DFU), and we demonstrate that a combination of the two independent measures leads to better overall discrimination.

The remainder of this paper is organized as follows. Section 2 describes the framework of the proposed DFU classification method. The dataset used in the current work are then described in section 3. Section 4 reports the experimental design and results, and Section 5 discusses the results. Finally, Section 6 presents conclusions and future works.

2. Methodology

In this work, we investigate the benefits of using LBP codes as input to CNNs models in DFU classification. We design a CNN architecture with three different input: DFU-RGB-Net using original RGB images, DFU-TEX-Net using texture coded mapped LBP images, and DFU-RGB-TEX-Net using both RGB and texture coded mapped LBP images.

The proposed DFU classification method classifies the input image in healthy and DFU classes. The stages of the process are (i) extract the texture features (LBP codes) using Basic LBP method [33], and then convert the extracted LBP codes to the 3D space using the method described in [30] in order to make the LBP codes suited as a CNN's input, and (ii) train several CNNs models on the RGB images and the mapped LBP codes separately and in combination. The first aim is to investigate the ability of the CNNs model trained on texture features only in DFU classification in comparison to the CNNs models trained on RGB images only. The second aim is to demonstrate which are the best features for the problem of DFU classification. The third aim is to investigate the importance of textures features when fused with the RGB images in DFU recognition. In the following, we describe the proposed method in detail.

LBP codes Extraction: As mentioned in the beginning, LBP is one of the most successful and widely used descriptor in analysing images and texture classification. LBP descriptor capture local image micro-textures and it works by labelling each pixel in the image with a binary number. This number results from thresholding the grey level intensity of neighbourhoods of each pixel with the intensity of the centre pixel. Thresholded values are coded as 0 or 1 and are read systematically to form a binary number which is used then to label each pixel with a decimal number, called an LBP code, which represents the local structure around each pixel. The formal definition of LBP is given in Equations 1 and 2. Given an image I of width w and height h, the LBP code $LBPC_{n,r}(x_c, y_c)$ of the centre pixel $I(x_c, y_c)$ from any patch in the image is computed as:

$$LBPC_{n,R}(x_c, y_c) = \sum_{n=0}^{n-1} s(I(x_n, y_n) - I(x_c, y_c))2^n \quad (1)$$

where n is the total number of involved neighbours, R is the radius of the local neighborhood, $I(x_n, y_n)$ is the gray value of the neighbors, $I(x_c, y_c)$ is the gray value of the central pixel and $s(t)$ is the thresholding function and is computed as:

$$s(t) = \begin{cases} 1 & \text{if } t >= 0 \\ 0 & otherwise. \end{cases} \quad (2)$$

The final LBP code from (1) is 2^n bit string. For example, in case of $n = 8$ pixel neighbourhoods, the final output will be 8 bit number between 0 and 255. Generally the occurrence of LBP code is then used to generate a histogram to represent an entire image region. This histogram is then used for classification among several classes either by computing the histogram's similarities or by training a classifier such as Support Vector Machines (SVMs) to use on the test image. Owing to the recent extreme success that achieved by deep learning, we seek to investigate the strength of training the CNN architecture instead of learning a classifier on the LBP feature in the problem of DFU classification. Training the CNN model directly on the LBP code is not applicable owing to the unordered nature of the LBP codes which are not suited for the convolution operations, which equivalent to a weighted average of the input values, performed within CNN. This problem can be solved by mapping the LBP codes to 3D metric space as already done in [30] with the context of emotion recognition.

Mapped LBP codes: The work of [30] provides a solution to the unordered nature of LBP codes. They propose to map the LBP codes to points in a 3D metric space in which the Euclidean distance approximates the distance between the LBP codes. After the transformation of the LBP codes they can be averaged together using convolution operations within CNN models. In [30], the authors have successfully solved the problem of unordered nature of the LBP codes by mapping the LBP codes to points in 3D metric space where the distance between the LBP codes is approximated using Euclidean distance. Once the LBP codes are transformed, they can average together using convolution operators of CNN models. The method is accomplished by defining a distance $\delta_{m,n}$ between the LBP codes $LBPC_m$ and $LBPC_n$. In [30], the authors choose the Earth Movers Distance (EMD) [37] since it accounts for both the different bit values and their locations. Once the distance between LBP codes is defined, it is possible to obtain a mapping of the LBP codes into D-dimensional space which approximately preserves this distance. Multi Dimensional Scaling (MDS) method [8, 39] can be used to obtain this mapping as follow:

$$\delta_{m,n} \approx \|L_m - L_n\| = \|MDS(LBPC_m) - MDS(LBPC_n)\| \quad (3)$$

where $L_m = MDS(LBPC_m)$ and $L_n = MDS(LBPC_n)$ are the mapping of code m and n into the D-dimensional space. Based n this mapping, we can transfer the LBP codes into a representation that suitable to be used as CNN's input. Authors in [30] also proposed a mapped method to calculate a cyclic distance to accounts for cyclic nature of the LBP code. For more details about the mapped LBP code see [30]. In this paper, the texture coded mapped images are obtained by first extracting the LBP values. Since the best results for LBP values are obtained using three different values of radius parameter: 1, 5 and 10 [30]. Therefore, in this paper, we use the same parameter settings of three different values of radius parameter: 1, 5 and 10. These convert the values of pixels intensity of an image to one of the range from 1 to 256 LBP values (see the third row of Figure 1).

Those three LBP codes are then processed using the encoding from Equation 3 with the Regular (R) and Cyclic (C)

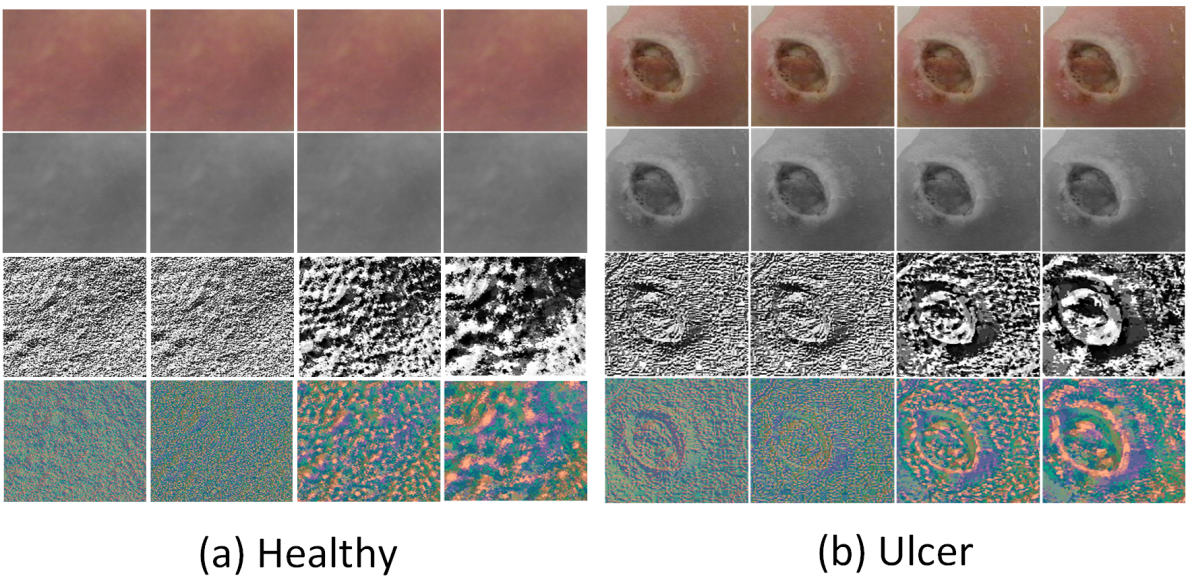

Figure 1: Example of the mapped LBP coded images using (a) healthy image and (b) image with ulcer: original RGB (first row), converted to gray value (second row), converted to LBP codes (third row from left to right with 1 radius, 5 radius and 10 radius) and the LBP codes are mapped to a 3D metric space (fourth row from left to right LBP-1-R, LBP-1-C, LBP-5-C and LBP-10-C).

distance (see fourth row of Figure 1). In total, each image was represented using both RGB values and by extraction four mapped LBP codes: LBP-1-R, LBP-1-C, LBP-5-C and LBP-10-C. Figure 1 shows an example of a healthy and DFU images mapped to a 3D metric space. The resulting texture coded mapped image I^* with 3-channel are then used as input to CNN models as described below.

Fusing RGB image and Mapped LBP codes: Once the mapped LBP coded image I^* are computed from the original RGB image I as described above, where the micro texture pattern regarding the diseases can represented, the fused image I^{**} can be computed as follows:

$$I^{**}(x, y) = I(x, y) + I^*(x, y) \tag{4}$$

where the value of each pixel in the output image I^{**} is a linear combination of the corresponding pixel values in the input images I and its mapped LBP coded image I^*. The advantage of fusing the images this way is to obtain more information about the ulcer and normal skin. Thus this information can help to maximize the sharpness of the object (DFU) and the robustness of local image features in the merged image since both the RGB and mapped LBP images contain overlapping information and combined them together lead to better discrimination of image data into two pixel groups. Therefore, more features regarding the disease (DFU) can be learned by CNN learning. Figure 2 shows an example of combining RGB images with the mapped LBP coded images.

CNN Model:

Our proposed network consist of four convolution layers and two fully connected layers without padding. Each convolution layers consists of one Rectifier Linear Unit (ReLU)

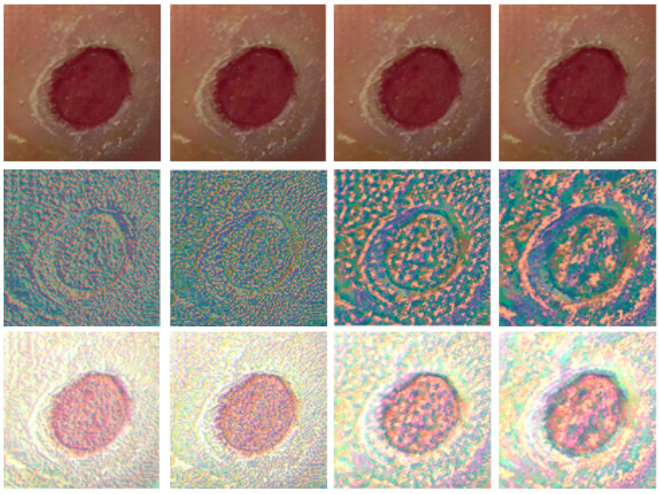

Figure 2: Illustration of fusing RGB and mapped LBP coded images: (top row) RGB image, (middle row) mapped LBP coded image and (bottom row) the resulting fusion of both the RGB and the mapped LBP coded images.

activation function and followed by one pooling layer. The last fully connected layer consists of one Sigmoid function which is usually used for binary classification problems. Binary cross entropy was used as a loss function optimized with Adaptive Moment Estimation (Adam) [27]. Weights and biases were initialised with the Xavier uniform kernel initializer [17] and zeros respectively. The input patch size was 151×151. The optimal size of the proposed CNN architecture is found empirically where the number of the convolution and max-pooling is increased gradually and subsequently, the number of the filters are adjusted gradually as well, and then the network with the best performance was

Table 1
Summary of the proposed CNN architecture.

Layer Type	Kernel Size	Stride Size	Activation Function	No. of filter	FC units Layer	Input Shape	Output Shape
Convolution	3×3	1	ReLU	32	-	(151,151,3)	(149,149,32)
Max pooling 2D	2×2	2	-	-	-	(149,149,32)	(74,74,32)
Convolution	3×3	1	ReLU	64	-	(74,74,32)	(72,72,64)
Max pooling 2D	2×2	2	-	-	-	(72,72,64)	(36,36,64)
Convolution	3×3	1	ReLU	128	-	(36,36,64)	(34,34,128)
Max pooling 2D	2×2	2	-	-	-	(34,34,128)	(17,17,128)
Convolution	3×3	1	ReLU	256	-	(17,17,128)	(15,15,256)
Max pooling 2D	2×2	2	-	-	-	(15,15,256)	(7,7,256)
Flatten	-	-	-	-	-	-	12544
Fully connected	3×3	1	ReLU	-	-	-	-
Dropout (rate=0.5)	-	-	-	-	250	-	-
Fully connected	-	-	Sigmoid	-	2	-	-

chosen. Table 1 summarized the details of the CNN architecture proposed in this paper.

Our goal is to classify the input image into DFU and healthy classes using CNNs. One way to predict the class y of the new image is to train a CNN classifier f on the RGB images I as follows:

$$y = f(I) \tag{5}$$

Another way is to train the f function on the the mapped LBP codes image I^* as follows:

$$y = f(I^*) \tag{6}$$

An alternative is to train the f function on the fused image I^{**} which obtained from Equation 4 as follows:

$$y = f(I^{**}) \tag{7}$$

Figure 3 illustrates the deep architecture proposed in this paper. In the beginning, the performance of the proposed network is investigated using RGB images only as input, what we referred to as DFU-RGB-Net. Secondly, the performance is investigated using LBP codes only as input, what we referred to as DFU-TEX-Net. Finally, the information of the original image value and the texture features are combined before the first convolution layer and both RGB images and the LBP codes together are used as input to the CNN, what we referred to as DFU-RGB-TEX-Net.

For benchmark algorithms, we implemented two of the existing state-of-the-art CNN architecture including AlexNet and GoogLeNet [44] for the classification of healthy and unhealthy (DFU) classes. These networks are widely and successfully used in several computer vision problems such as classification of medical [31, 35, 41] and non-medical applications in general and in particular for DFU classification [18, 4]. AlexNet CNN architecture was developed by [29] and emerged as a winner of ImageNet ILSVRC-2012 competition in the classification category by achieving 99%. This network consists of eight layers. The first five layers are convolution layers, some of them are followed by max-pooling layers. The last three layers are fully connected layers. The input layer of AlexNet takes an image size of 227 x 227 x 3 dimension and the output from the last fully connected layer is fed to a softmax function to produce the probabilities of 1000 different classes. See [29] for more details about the AlexNet network.

GoogLeNet CNN architecture was developed by [44] and was the winner in the ILSVRC-2014 challenge. This network consists of 22 layers deep, nine of them are inception modules with three different convolutional kernels (1 x 1, 3 x 3 and 5 x 5) in each. Using such a module, multiple convolution filter inputs can be processed on the same input and do pooling at the same time. All the results are then combined into a single feature layer which allows the model to take advantage of multi-level feature extraction from every input. The input layer of GoogLeNet takes an image size of 224 x 224 x 3 dimension and the last layer is a softmax layer for classifying 1000 different classes.

In this paper, we resize the images of the DFU dataset to 227 x 227 and 224 x 224 in order to fit the shape input of AlexNet and GoogLeNet respectively. Furthermore, the final layer of both networks was adjusted to work well with our binary (2 classes) classification problem.

3. Dataset

We evaluate our approach by performing experiments on the recently introduced DFU classification dataset [18]. The dataset consists of 1,679 images divided in 641 healthy (normal) and 1038 Ulcer (abnormal)foot images. Since the Deep networks require a huge amount of data to train, validate, test and obtain a convincing conclusion, the number of images

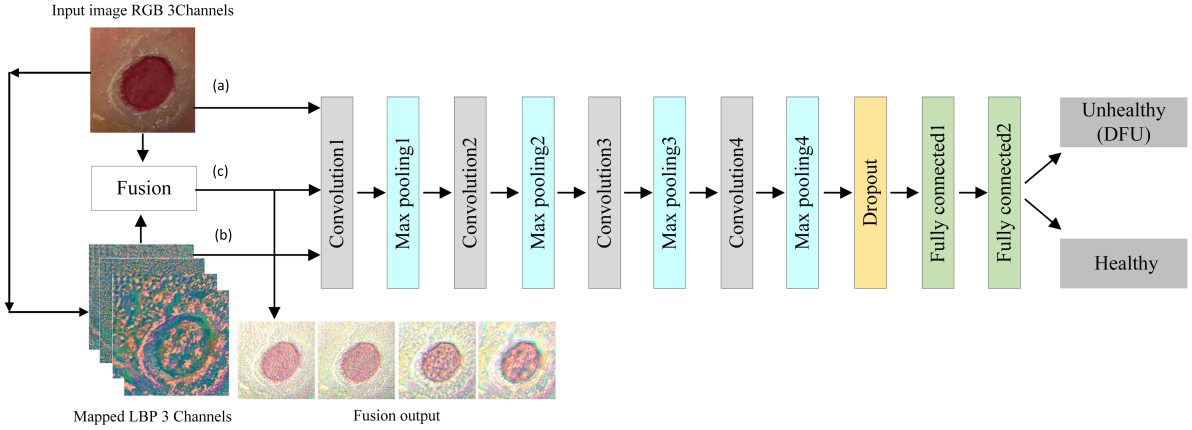

Figure 3: Three-stream deep CNN architectures: (a) DFU-TEX-Net trained using RGB images, (b) DFU-RGB-Net trained using Mapped LBP coded imagesand (c) DFU-RGB-TEX-Net trained on the combination of RGB and mapped LBP coded images before the first fully connected layer.

is increased by 10 times using data augmentation techniques similar to that in [18] and in total the data is increased from 1679 images to 16790 images. The data augmentation is performed by using a combination of several techniques including: rotation, flipping, and using different color space. The rotation is accomplished by rotating the original images with several angles: 90° , 180° , 270° . The flipping is performed by flipping the original image with horizontal flip, vertical flip and horizontal+vertical flip. Then the the color space augmentation is performed by using four color space including YCbCr, NTSC, HSV and L*a*b on the original images. Before the augmentations, we have split the dataset into 10 folds, each fold represents 10% from the original dataset. Then, the augmentations techniques are performed in each part separately to make sure all the augmented images originated from the original images. We further perform 10-fold cross validation experiments in 5 iterations. In each iteration 8 folds (80%) (13,432 images) from the dataset used as a training set, 1 fold (10%) (1,679 images) from the dataset used as a validation set, 1 fold (10%) (1,679 images) from the dataset used as a testing set. In the first iteration, the first and second folds are used to test and validate the models respectively and the rest are used to train the model. In the second iteration, the third and fourth folds are used as the testing and validation sets respectively while the rest are used as the training set. In the third iteration, the fifth and sixth folds are used to test and validate the models respectively and the rest are used to train the model. In the fourth iteration, the seventh and eighth folds are used to test and validate the models respectively and the rest are used to train the model. In the fifth iteration, the ninth and tenth folds are used to test and validate the models respectively and the rest are used to train the model. In each iteration, the validation and testing sets are swapped to ensure that each fold of the 10 folds have been used to test the model and every image was tested exactly once. Figure 4 illustrates the data pipeline and cross validation protocol used in this paper.

4. Experiments and Results

We performed a series of experiments to investigate the effect of feeding texture features (i.e. mapped LBP) to the CNN models along with/without the RGB images on the performance of the automatic DFU recognition method. We then compare the performance of the proposed approach with the recently publicised results on DFU recognition.

The performances were evaluated by calculating area under the Receiver Operating Characteristic curve (AUC) [15]. The Receiver Operating Characteristic (ROC) curve plots the performance of the binary classification between positive and negative classes by plotting the true positive rate (number of correctly classified true images) (Sensitivity) against the false positive rate (number of misclassified true images) (1-Specificity). The AUC is a measure of how well a classifier can distinguish between two groups (binary classification) (e.g.DFU/normal) and is in the range [0,1]. We also report Sensitivity, Specificity, Precision, Accuracy and F-Measure as our evaluation metrics. In medical imaging, Sensitivity and Specificity are considered reliable evaluation metrics for classifier completeness.

In all experiments, the results are reported as average of 5-fold cross validation as described in Figure 4 to give a mean accuracy, standard deviation and AUC. During the training of the network, the learning rate with Adam solver was set 0.001,the epoch and batch size were set to 30 and 32 respectively and then we selected the model with lowest validation loss. At the start of each epoch, the training data was randomly shuffled in order to produce different batches each time. In every experiment, the image was represented using five different representations including both RGB values and four extracted mapped LBP codes: LBP-1-R, LBP-1-C, LBP-5-C, LBP-10-C.

Implementation: LBP encoding and mapping, as described in Section 2, was implemented in Matlab R2014a as in [30]. Training and testing the CNN models were done using the Keras 2.3.1 open source framework for Deep Convolutional Neural Networks library written in Python 3.8.3 that runs on

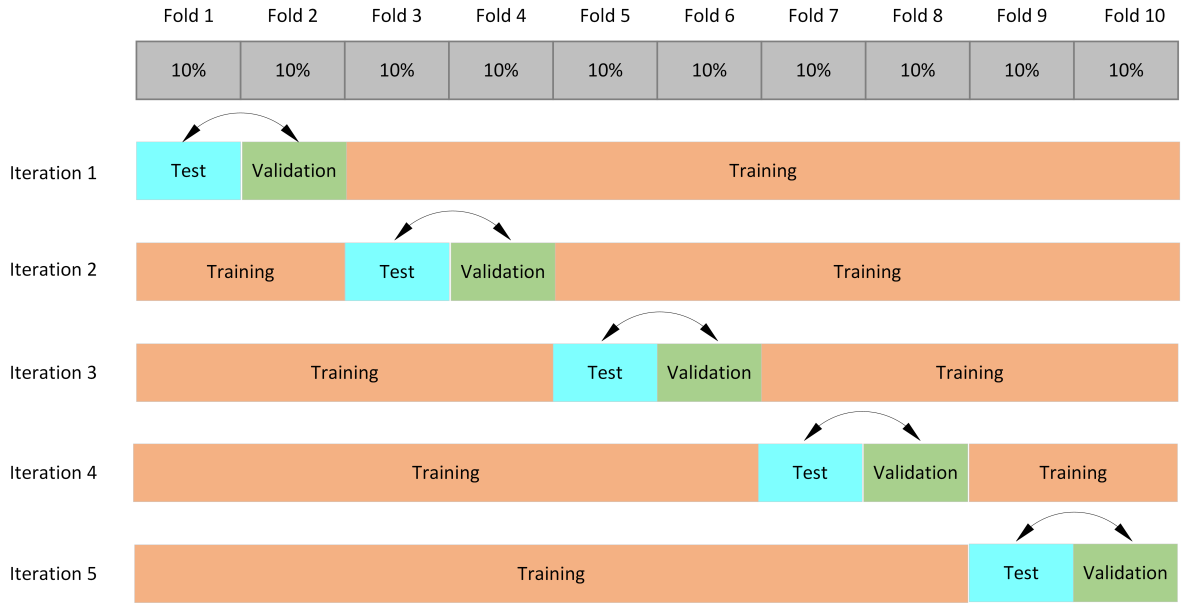

Figure 4: The illustration of our cross validation protocol.

Table 2
DFU classification results. Overall, the texture features (mapped LBP codes) achieved approximately similar performance to that of using RGB values and our proposed network provide better results than AlexNet and GoogLeNet.

Network	Features	Sensitivity	Specificity	Precision	Accuracy	F-Measure	mean AUC (stdev.)
DUF-RGB-Net (ours)	RGB	0.912	0.897	0.936	0.906	**0.952**	0.961 (± 0.005)
DUF-TEx-Net (ours)	LBP-1-R	0.918	0.877	0.922	0.902	0.920	0.960 (± 0.007)
DUF-TEx-Net (ours)	LBP-1-C	0.897	0.882	0.925	0.892	0.911	0.950 (± 0.006)
DUF-TEx-Net (ours)	LBP-5-C	**0.921**	**0.931**	**0.956**	**0.925**	0.938	0.953 (± 0.015)
DUF-TEx-Net (ours)	LBP-10-C	0.912	0.899	0.937	0.907	0.924	**0.965** (± 0.016)
Alexnet	RGB	0.898	0.778	0.869	0.852	0.892	0.929 (± 0.020)
Alexnet	LBP-1-R	0.912	0.825	0.894	0.878	0.902	0.945 (± 0.023)
Alexnet	LBP-1-C	0.923	0.722	0.873	0.846	0.889	0.898 (± 0.155)
Alexnet	LBP-5-C	0.904	0.726	0.855	0.836	0.874	0.890 (± 0.132)
Alexnet	LBP-10-C	0.870	0.837	0.896	0.857	0.883	0.922 (± 0.011)
GoogLeNet	RGB	0.890	0.826	0.893	0.866	0.891	0.930 (± 0.014)
GoogLeNet	LBP-1-R	0.915	0.835	0.902	0.884	0.908	0.952 (± 0.035)
GoogLeNet	LBP-1-C	0.853	0.727	0.837	0.804	0.908	0.892 (± 0.021)
GoogLeNet	LBP-5-C	0.904	0.726	0.855	0.836	0.874	0.895 (± 0.133)
GoogLeNet	LBP-10-C	0.890	0.826	0.893	0.866	0.891	0.929 (± 0.017)

Tensorflow 2.1.0.

4.1. RGB Images vs. Mapped LBP Coded Images

In this experiment, the performance of CNNs models that used texture features (mapped LBP codes) as input calculated using Equation 3 is compared with the performance of CNNs models that used RGB images only as input in order to investigate whether RGB images and mapped LBP coded images have complementary features. Using the proposed CNN architecture described in section 2 and Figure 3(a,b), five CNNs models (one for each representation) were trained using similar configurations. The cross validation mean AUC results are shown in Table 2 and ROC curves in Figure 5 (a). These results show that using the mapped LBP codes only provides satisfactory and approximately similar performance to that of using RGB images only. The results also show that the best LBP codes in splitting the data were

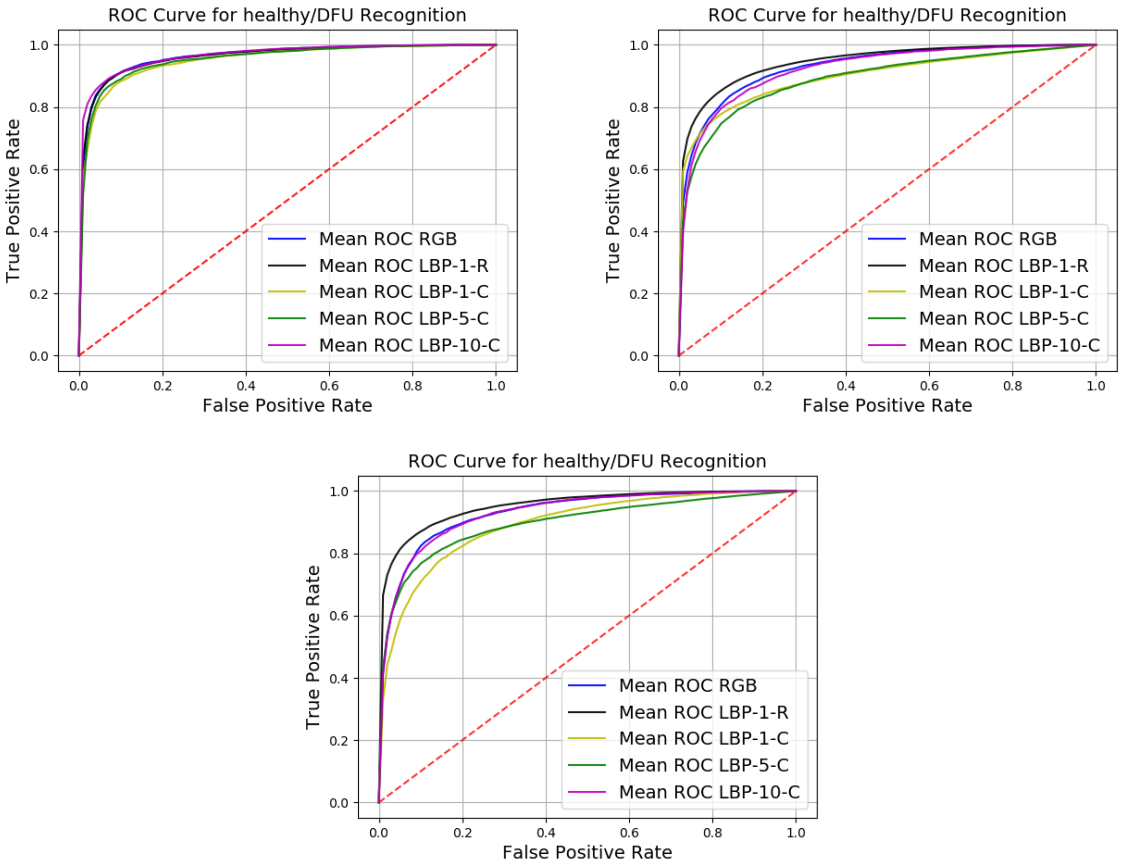

Figure 5: ROC curves of all representations: RGB images, LBP-1-R images, LBP-1-C images, LBP-5-C images and LBP-10-C images in detecting DFU vs. non-DFU (a) Our proposed DFU-RGB-TEX-Net, (b) AlexNet, and (c)GoogLeNet. Overall, the mapped LBP features achieved approximately similar performance to that of using RGB values and our proposed net (top) provides better results than AlexNet (middle) and GoogLeNet (bottom).

LBP-1-R and LBP-10-C, with an AUC of 0.960 ± 0.007 and 0.965 ± 0.016 respectively.

For more validation regarding the effectiveness of using texture coded images as an alternative to the RGB images as input to the CNN and in particular, in the DFU classification, we used two different existing network architectures: the AlexNet networks [29] and the GoogLeNet network [45] (see original papers for more details regarding design and architecture of each network) and our adjustment on these networks described in 2. The results again show that using the mapped LBP codes only provides satisfactory and approximately similar performance to that of using RGB values only (see the cross validation mean AUC results illustrated in Table 2 and ROC curves described in Figure 5 (b and c). The best extracted features again are the LBP-1-R and LBP-10-C features, with an AUC of 0.945 ± 0.023 and 0.922 ± 0.011 using AlexNet and 0.952 ± 0.035 and 0.929 ± 0.017 using GoogLeNet respectively.

Overall, we showed that feeding Mapped LBP codes images to the three different CNN architectures (GoogLeNet, AlexNet and our proposed method) as an alternative to RGB images can achieve better or comparable performance to that of using RGB images in terms of Sensitivity, Specificity, Precision, Accuracy and F-Measure as shown in Table 2. The results in Table 2 also showed that our proposed networks (DFU-RGB-Net and DFU-TEX-Net) provide significantly better performance than AlexNet and GoogLeNet that why we use our proposed CNN architecture in the rest of the experiments. The other reason behind using our proposed CNN architecture rather than using AlexNet and GoogLeNet architecture is to speed up the calculations using 4 layers with the aid of the handcrafted information

4.2. Fusion of RGB images and Mapped LBP coded images

In this experiment, the performance of CNNs models that used the fused images (the combination of RGB and mapped LBP coded images) as input calculated using Equation 4 is compared with the performance of CNNs models that used RGB images only as input in order to investigate whether the coded texture features can improve the overall performance of the network. We train four DFU-RGB-TEX-Net networks using similar configurations described in section 2 and Figure 3(c). Every network was trained using RGB images fused with one of mapped LBP codes (LBP-1-R, LBP-1-C, LBP-5-C, LBP-10-C). The cross-validation

Table 3
DFU classification results. Comparison of RGB results to Combined (RGB and mapped LBP codes) results.

Network	Features	Sensitivity	Specificity	Precision	Accuracy	F-Measure	mean AUC (stdev.)
DFU-RGB-Net	RGB	0.912	0.897	0.936	0.906	0.952	0.961 (± 0.006)
DFU-RGB-TEX-Net	RGB+ (LBP-1-R)	**0.943**	0.939	0.962	**0.941**	**0.952**	**0.981** (± 0.006)
DFU-RGB-TEX-Net	RGB+ (LBP-1-C)	0.925	0.908	0.943	0.919	0.934	0.970 (± 0.014)
DFU-RGB-TEX-Net	RGB+ (LBP-5-C)	0.921	0.931	0.956	0.925	0.938	0.971 (± 0.020)
DFU-RGB-TEX-Net	RGB+ (LBP-10-C)	0.921	**0.946**	**0.965**	0.930	0.942	0.977 (± 0.006)

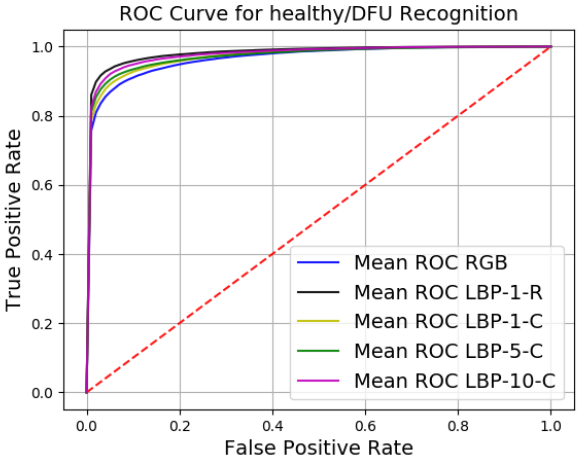

Figure 6: ROC curves comparing the performance of using the combined RGB images and mapped LBP codes to the RGB images only in detecting DFU vs. non-DFU. Overall our proposed combined both RGB images and LBP codes achieved the best result.

mean AUC results are shown in Table 3 and ROC curves in Figure 6. These results show that feeding the mapped LBP codes to the CNN significantly increases the AUC from 0.961, 0.960, 0.950, 0.953 and 0.965 using RGB, LBP-1-R, LBP-1-C, LBP-5-C and LBP-10-C respectively to 0.981, 0.970, 0.971 and 0.977 of using fused features of (RGB + (LBP-1-R)), (RGB + (LBP-1-C)), (RGB + (LBP-5-C)) and (RGB + (LBP-10-C)) respectively.

In summary, we showed that feeding the combination of mapped LBP codes and RGB images can significantly increase the performance by large gap of 2.2% in sensitivity, 2% in specificity, 1% in precision, 2% in accuracy and 2% in AUC. The increase could come from the specific measurements of texture features regarding the disease adding to the RGB information captured by mapped LBP features.

4.3. Comparison to the other methods

The above experiments have shown that our DFU-RGB-TEX-Net, where the model train on both the RGB and the hand-crafted (LBP codes) features jointly, achieves better accuracy in DFU classification than DFU-RGB-Net (trained on RGB value only) or DFU-TEX-Net (trained on LBP coded images only). This is likely because the specific information of the disease adding to the RGB information captured by LBP codes. Here we compare the results of our approach to the state of the art results published by [18] and [4] in DFU classification applied on the same dataset.We have added these results(adopted from [18] and [4]) to Table4. These comparisons are summarized in Table 4. The results show that using the combined RGB and texture features achieve better than using RGB features only or texture features only as in [18] by large gap of 1% in sensitivity, 4% in specificity, 2% in precision, 2% in accuracy, 1.3% in f-measure and 2% in AUC (see Table 4). The results of our method also outperformed the results of other methods as in [4] by small gap of 0.7 in sensitivity, 1.1% in precision and 0.7 in f-measure (see Table 4).

Our proposed method performed better than the methods used in DFUNet [18] and DFU-QUTNet [4] because our method combined the texture features with RGB images (compact representation), while in [18], texture features and RGB images were used separately and in [4],the RGB values was used only.

4.4. Model Interpretability

In this section, we use both visualizing the intermediate convolution layer output technique and Grad-CAM [40] technique to interpret and understand how the model make prediction. The former technique help to understand how different filters are learned by the model and how the input is transferred by the layers whereas Grad-CAM technique visualizing where a CNN model is looking.

For the analysis of the effects of feeding the LBP responses, we compare some feature maps from both the DFU-RGB-Net where the CNN trained on RGB images only (see Figure 7 (a)) and the DFU-RGB-TEX-Net where the CNN trained on the fusion of RGB and LBP responses (see Figure 7 (b)). Figure 8 shows the Grad-CAM for a DFU and healthy skin class. The regions in red show the areas which activate more units (neurons) in the last convolutional layer before the classification. From the results, we can see that the features from the DFU-RGB-TEX-Net contain more strong DFU features than the DFU-RGB-Net, which is believed to be the cause of better performance.

Table 4
Comparison of our approach with the state-of-the-art approaches in DFU classification, in which our proposed method achieved the best result.

Ref	Method	features	Sensitivity	Specificity	Precision	Accuracy	F-Measure	AUC
[18]	LBP	texture	0.919	0.764	0.878	0.865	0.898	0.932
	LBP + HOG	texture	0.881	0.841	0.906	0.866	0.893	0.931
	LBP + HOG + Colour	texture + colour	0.902	0.845	0.904	0.880	0.904	0.943
	LeNet (CNN)	RGB	0.912	0.810	0.871	0.872	0.893	0.929
	Alexnet (CNN)	RGB	0.895	0.886	0.933	0.893	0.914	0.950
	GoogLeNet (CNN)	RGB	0.905	0.912	0.949	0.907	0.927	0.960
	DFUNet	RGB	0.934	0.911	0.945	0.925	0.939	0.961
[4]	VGG16 [42]	RGB	0.897	-	0.923	-	0.909	-
	Alexnet (CNN) [28]	RGB	0.872	-	0.911	-	0.891	-
	GoogLeNet (CNN) [44]	RGB	0.905	-	0.956	-	0.929	-
	DFU-QUTNet [4]	RGB	0.936	-	0.954	-	0.945	-
This Work	DFU-RGB-TEX-Net	RGB + (LBP-1-R)	**0.943**	0.939	0.962	**0.941**	**0.952**	**0.981**
	DFU-RGB-TEX-Net	RGB+ (LBP-1-C)	0.925	0.908	0.943	0.919	0.934	0.970
	DFU-RGB-TEX-Net	RGB+ (LBP-5-C)	0.921	0.931	0.956	0.925	0.938	0.971
	DFU-RGB-TEX-Net	RGB+ (LBP-10-C)	0.921	**0.946**	**0.965**	0.930	0.942	0.977

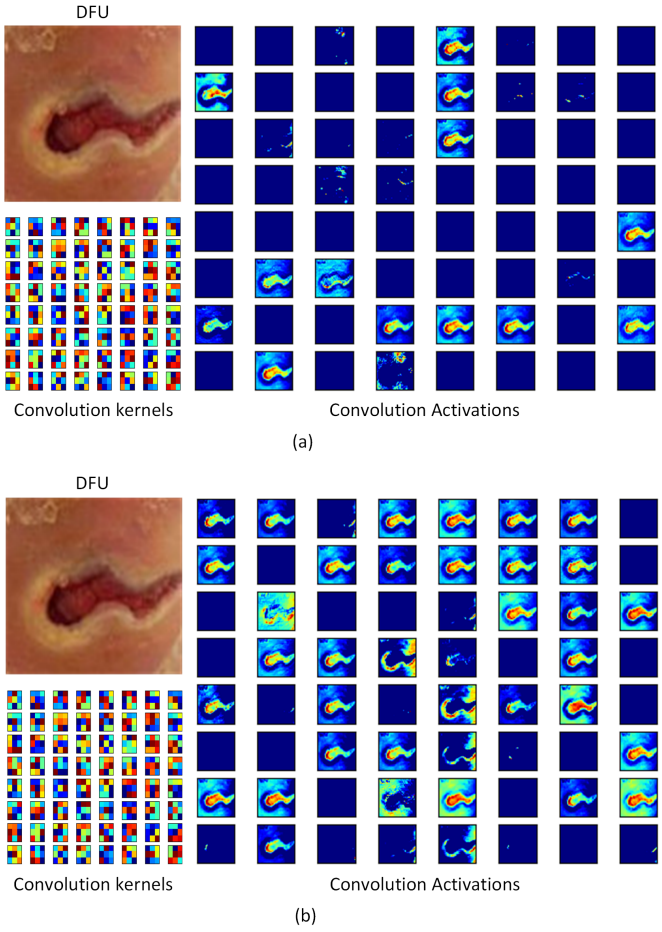

Figure 7: Illustration of model interpretability Comparison using feature maps technique: (a) Features from CNN model which trained using RGB images as input and, (b) Features from CNN model which trained using the fused of Mapped LBP coded images and RGB images as input.

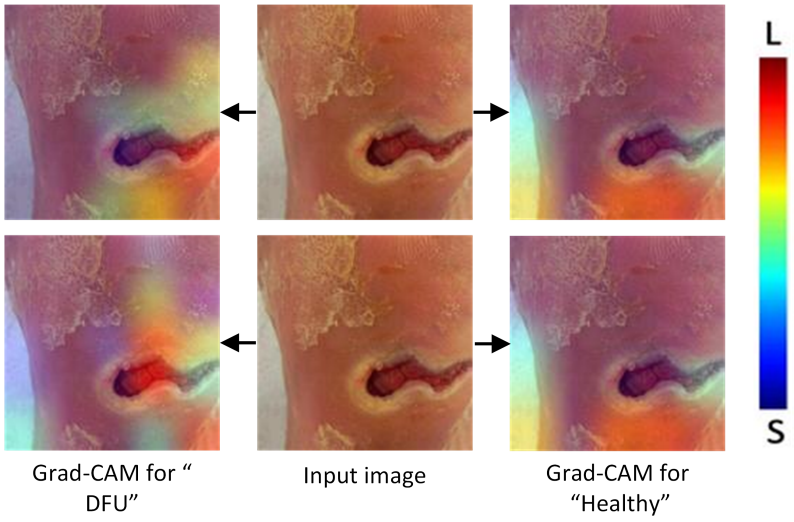

Figure 8: Illustration of model interpretability Comparison using Grad-CAM technique for a DFU (left) and healthy (right) skin classes show localizing map of the important feature in the image which activate more units (neurons) in the last convolutional layer before the classification: (top) activation from CNN model which trained using RGB images as input and, (bottom) activation from CNN which model trained using the fused of Mapped LBP coded images and RGB images as input. L and S refereed for large and small.

5. Discussion

Diabetic Foot Ulcer (DFU) is one of the major complications of Diabetes, which leads to lower limb amputation if not treated early and properly. Therefore, early prediction and early treatment of patients with mild symptoms at a high risk of DFU progression to a severe/critical stage are important ways to reduce lower limb amputation.

In this study, since texture features contain important information regarding the disease and subsequently such features are important for disease's classes classification (e.g. DFU verse healthy), we argue that fuse the handcrafted features with the original RGB images in order to utilize the advantages of both and then use the fused images as input to CNN instead of using RGB images only is important for an accurate DFU recognition. We have conducted extensive experiments to demonstrate that our system, which effectively fuses the two complementary representations regarding the disease, achieves better performance (0.981% AUC) than using either representation as input to the CNN separately (e.g., 0.965% vs. 0.961% in AUC when using the texture features and RGB values). More importantly, due to the capability of our system in learning texture information in combination with the RGB values, our system reports a higher sensitivity, specificity, precision, accuracy and f-measure and AUC compared with the counterparts that consider the RGB images only as illustrated in Table 4.

Since traditional approaches for fusing handcrafted features with CNN learning cannot benefit from end-to-end learning because each part of this method is individually trained where the handcrafted feature vector is extracted using any convolutional method and then it combined with the features of the last convolutional layer and the obtained feature vector is used as an input for a classifier. In this study, we attempted to explore ways to fuse handcrafted features with deep learning to improve DFU predicting in an end-to-end manner. Experimental results show that both handcrafted features and deep features of paramount importance to this problem. Furthermore, the texture features contributed significantly to the prediction of DFU as it reveals the change of the foot's skin appearance.

To investigate the interpretability of the DFU patterns learned by our model, we used visualizing the inter-mediate convolution layer output technique to understand how different filters are learned by the model and how the input is transferred by the layers. Experimental results showed that the model trained on the fused texture and RGB images contain stronger DFU features7 (b)) than the model trained on the RGB images only (see Figure 7 (a)). We also showed the activation maps using Grad-CAM. Experimental results described in Figure 7 illustrate that the DFU region has the greatest influence on the prediction task. Hence, they are valuable for DFU classification.

6. Conclusion

Recently, most of the DFU classification method trained the CNN models on the RGB images, with the belief that the CNN will automatically capture the appropriate features regarding the object (DFU) from the data. In this paper, we conducted extensive experiments to investigate the benefit of training the CNN model on fusion of Mapped LBP coded images (handcrafted features) and RGB images as input on the performance of DFU recognition tasks. We have noticed that using Mapped LBP coded images instead of RGB images as input to the CNN provided comparable performance on the similar experimental settings and dataset. We observed that train the CNN model on the RGB im-

ages and handcrafted features jointly can enhance the overall performance of the DFU classification. Specifically, we have shown that feeding the mapped LBP coded images with the RGB images provided better results in DFU classification. Our proposed DFU classifier outperformed the existing methods with AUC of 0.981 on cross-validation experiments on DFUNet dataset [18].

Finally, the results of our approach (using mapped LBP and CNN) are encouraging and will lead to further investigations in designing a robust solution for DFU pathology recognition. Therefore, for future works, we will revise the present work using different hand-crafted descriptors such as Gabor filter response and Histogram of Gradient. We will expand our work to other DFU pathology, including the recognition of infection and ischaemia. Future research will be benefited by investigating appropriate handcrafted features as the input or convolutional layers of the CNNs. Additionally, the newer deep learning models, such as DenseNet and EfficientNet, will be explored in the future.

References

[1] Abdellatef, E., Omran, E.M., Soliman, R.F., Ismail, N.A., Abd Elrahman, S.E.S., Ismail, K.N., Rihan, M., Abd El-Samie, F.E., Eisa, A.A., 2020. Fusion of deep-learned and hand-crafted features for cancelable recognition systems. Soft Computing 24, 15189–15208.

[2] Al-Garaawi, N., Wu, Q., Morris, T., 2020. Brief-based face descriptor: an application to automatic facial expression recognition (afer). Signal, Image and Video Processing , 1–9.

[3] Algaraawi, N., 2019. Modelling of Human Ageing, Compound Emotions, and Intensity for Automatic Facial Expression Recognition. The University of Manchester (United Kingdom).

[4] Alzubaidi, L., Fadhel, M.A., Oleiwi, S.R., Al-Shamma, O., Zhang, J., 2020. Dfu_qutnet: diabetic foot ulcer classification using novel deep convolutional neural network. Multimedia Tools and Applications 79, 15655–15677.

[5] Amin, J., Sharif, M., Anjum, M.A., Khan, H.U., Malik, M.S.A., Kadry, S., 2020. An integrated design for classification and localization of diabetic foot ulcer based on cnn and yolov2-dfu models. IEEE Access .

[6] Anwer, R.M., Khan, F.S., van de Weijer, J., Molinier, M., Laaksonen, J., 2018. Binary patterns encoded convolutional neural networks for texture recognition and remote sensing scene classification. ISPRS journal of photogrammetry and remote sensing 138, 74–85.

[7] Bernasconi, A., Antel, S.B., Collins, D.L., Bernasconi, N., Olivier, A., Dubeau, F., Pike, G.B., Andermann, F., Arnold, D.L., 2001. Texture analysis and morphological processing of magnetic resonance imaging assist detection of focal cortical dysplasia in extra-temporal partial epilepsy. Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society 49, 770–775.

[8] Borg, I., Groenen, P.J., 2005. Modern multidimensional scaling: Theory and applications. Springer Science & Business Media.

[9] Cassidy, B., Reeves, N.D., Joseph, P., Gillespie, D., O'Shea, C., Rajbhandari, S., Maiya, A.G., Frank, E., Boulton, A., Armstrong, D., et al., 2020. Dfuc2020: Analysis towards diabetic foot ulcer detection. arXiv preprint arXiv:2004.11853 .

[10] Castellano, G., Bonilha, L., Li, L., Cendes, F., 2004. Texture analysis of medical images. Clinical radiology 59, 1061–1069.

[11] Chabat, F., Yang, G.Z., Hansell, D.M., 2003. Obstructive lung diseases: texture classification for differentiation at ct. Radiology 228, 871–877.

[12] Cruz-Vega, I., Hernandez-Contreras, D., Peregrina-Barreto, H., Rangel-Magdaleno, J.d.J., Ramirez-Cortes, J.M., 2020. Deep learning classification for diabetic foot thermograms. Sensors 20, 1762.

[13] Dalal, N., Triggs, B., 2005. Histograms of oriented gradients for human detection, in: 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05), Ieee. pp. 886–893.

[14] Eid, M.M., Yousef, R.N., Mohamed, M.A., . A proposed automated system to classify diabetic foot from thermography. International Journal of Scientific & Engineering Research 9, 371–381.

[15] Fawcett, T., 2006. An introduction to roc analysis. Pattern recognition letters 27, 861–874.

[16] Giger, M.L., 2018. Machine learning in medical imaging. Journal of the American College of Radiology 15, 512–520.

[17] Glorot, X., Bengio, Y., 2010. Understanding the difficulty of training deep feedforward neural networks, in: Proceedings of the thirteenth international conference on artificial intelligence and statistics, JMLR Workshop and Conference Proceedings. pp. 249–256.

[18] Goyal, M., Reeves, N.D., Davison, A.K., Rajbhandari, S., Spragg, J., Yap, M.H., 2018a. Dfunet: Convolutional neural networks for diabetic foot ulcer classification. IEEE Transactions on Emerging Topics in Computational Intelligence 4, 728–739.

[19] Goyal, M., Reeves, N.D., Rajbhandari, S., Ahmad, N., Wang, C., Yap, M.H., 2020. Recognition of ischaemia and infection in diabetic foot ulcers: Dataset and techniques. Computers in biology and medicine 117, 103616.

[20] Goyal, M., Reeves, N.D., Rajbhandari, S., Yap, M.H., 2018b. Robust methods for real-time diabetic foot ulcer detection and localization on mobile devices. IEEE journal of biomedical and health informatics 23, 1730–1741.

[21] Hatt, M., Tixier, F., Pierce, L., Kinahan, P.E., Le Rest, C.C., Visvikis, D., 2017. Characterization of pet/ct images using texture analysis: the past, the present… any future? European journal of nuclear medicine and molecular imaging 44, 151–165.

[22] He, D.C., Wang, L., 1990. Texture unit, texture spectrum, and texture analysis. IEEE transactions on Geoscience and Remote Sensing 28, 509–512.

[23] Hewitt, B., Yap, M.H., Grant, R.A., 2016. Manual whisker annotator (mwa): A modular open-source tool. Journal of Open Research Software 4.

[24] Hosseini, S., Lee, S.H., Cho, N.I., 2018. Feeding hand-crafted features for enhancing the performance of convolutional neural networks. arXiv preprint arXiv:1801.07848 .

[25] Jawahar, M., Anbarasi, L.J., Jasmine, S.G., Narendra, M., 2020. Diabetic foot ulcer segmentation using color space models, in: 2020 5th International Conference on Communication and Electronics Systems (ICCES), IEEE. pp. 742–747.

[26] Khan, M.A., Sharif, M., Akram, T., Raza, M., Saba, T., Rehman, A., 2020. Hand-crafted and deep convolutional neural network features fusion and selection strategy: an application to intelligent human action recognition. Applied Soft Computing 87, 105986.

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

[28] Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems 25, 1097–1105.

[29] Krizhevsky, A., Sutskever, I., Hinton, G.E., 2017. Imagenet classification with deep convolutional neural networks. Communications of the ACM 60, 84–90.

[30] Levi, G., Hassner, T., 2015. Emotion recognition in the wild via convolutional neural networks and mapped binary patterns, in: Proceedings of the 2015 ACM on international conference on multimodal interaction, pp. 503–510.

[31] Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., Van Der Laak, J.A., Van Ginneken, B., Sánchez, C.I., 2017. A survey on deep learning in medical image analysis. Medical image analysis 42, 60–88.

[32] Nailon, W.H., 2010. Texture analysis methods for medical image characterisation. Biomedical imaging 75, 100.

[33] Ojala, T., Pietikainen, M., Maenpaa, T., 2002. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Transactions on pattern analysis and machine intelligence

24, 971–987.

[34] Platt, J., 1998. Fast training of support vector machines using sequential minimal optimization, in: Advances in Kernel Methods - Support Vector Learning, MIT Press.

[35] Razzak, M.I., Naz, S., Zaib, A., 2018. Deep learning for medical image processing: Overview, challenges and the future. Classification in BioApps , 323–350.

[36] Reddy, G.V., Savarni, C.D., Mukherjee, S., 2020. Facial expression recognition in the wild, by fusion of deep learnt and hand-crafted features. Cognitive Systems Research 62, 23–34.

[37] Rubner, Y., Tomasi, C., Guibas, L.J., 2000. The earth mover's distance as a metric for image retrieval. International journal of computer vision 40, 99–121.

[38] Saba, T., Mohamed, A.S., El-Affendi, M., Amin, J., Sharif, M., 2020. Brain tumor detection using fusion of hand crafted and deep learning features. Cognitive Systems Research 59, 221–230.

[39] Seber, G.A., 2009. Multivariate observations. volume 252. John Wiley & Sons.

[40] Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D., 2017. Grad-cam: Visual explanations from deep networks via gradient-based localization, in: Proceedings of the IEEE international conference on computer vision, pp. 618–626.

[41] Shen, D., Wu, G., Suk, H.I., 2017. Deep learning in medical image analysis. Annual review of biomedical engineering 19, 221–248.

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

[43] Song, A., Zhu, H., Huang, X., Xu, X., Liu, L., Chen, Y., . Cascade attention detnet: Object detection for diabetic foot ulcer .

[44] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., 2015. Going deeper with convolutions, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1–9.

[45] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z., 2016. Rethinking the inception architecture for computer vision, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2818–2826.

[46] Wild, S., Roglic, G., Green, A., Sicree, R., King, H., 2004. Global prevalence of diabetes: estimates for the year 2000 and projections for 2030. Diabetes care 27, 1047–1053.

[47] Yap, M.H., Hachiuma, R., Alavi, A., Brungel, R., Goyal, M., Zhu, H., Cassidy, B., Ruckert, J., Olshansky, M., Huang, X., et al., 2020. Deep learning in diabetic foot ulcers detection: A comprehensive evaluation. arXiv preprint arXiv:2010.03341 .