

# Fusion of Hand-crafted and Deep Features for Automatic Diabetic Foot Ulcer Classification

Nora Al-Garaawi<sup>1</sup>, Zainab Harbi<sup>1</sup>, Tim Morris<sup>2</sup>

<sup>1</sup>Department of Computer Science, Faculty of Education for Girls, University of Kufa, Najaf, Iraq

<sup>2</sup>Department of Computer Science, University of Manchester, Manchester, UK

**Abstract** – This paper proposes to combine both the texture and deep features to build a robust diabetic foot ulcer recognition system since both features represent valuable information about the disease. The proposed system consists of three stages: feature extraction, feature fusion, and DFU classification. The feature extraction is performed by extracting the handcrafted and deep features. The feature fusion is performed by concatenating both feature vectors into a single vector. The DFU classification is performed by training a random forest classifier on the fusion vectors and the resulting classifier is used then for classification. Experimental results showed that the proposed approach provides satisfactory performance in DFU, ischaemia, and infection classification.

**Keywords** – diabetic foot ulcer, classification, ischaemia and infection, hand-crafted features, deep features, fusion features.

## 1. Introduction and Background

DFU such as ischaemia and infection which happen due to elevated blood glucose can lead to amputation of the lower limb if not diagnosed early and properly [1].

Recently, computer-based systems using artificial intelligence algorithms such as image processing, computer vision, and machine learning algorithms have begun to play a significant role in many medical and non-medical applications using various imaging modalities such as MRI, CT, and ultrasound, and DFU classification is no exception. Building such a system has several advantages such as developing patient care, early diagnosis, and reducing mistakes in the traditional care systems. Despite the existence of several DFU detection and classification systems [2], [3], [4], [5], the development of a fully automated DFU diagnosis system is still in the earlier stages. In this paper, we introduce a new fully automatic DFU diagnosis system.

Generally, the building of the computer-based DFU automatic diagnosis system can be performed using either conventional or convolutional approaches. In the conventional approach, the handcrafted features regarding the meaningful information (e.g. DFU region) in the input image can be extracted using any texture descriptor. Texture descriptors are a widely used approach for feature extraction in a wide range of medical and non-medical applications as in [6], [7], [8], [9], [10], [11]. The extracted features can be used further to train one of the machine learning approaches for final classification (e.g. DFU recognition) [2]. Another approach for building the DFU diagnosis system is to use a deep representation for building an end-to-end diagnosis system as in [2], [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17].

Although the literature showed that the deep-based DFU classification systems achieved better performance than handcrafted-based DFU classification [2], we argue that both deep and handcrafted features are important to distinguish between healthy and unhealthy skin regions. This is obvious because in medical images texture features are important because they reflect the distribution of pixels and underlying structure of different diseases [18], [19], [20], [21]. Moreover, deep features are also important as demonstrated in achievements in several medical and non-medical applications (e.g. DFU diagnosis) [6], [7], [8], [9], [13], [14], [15], [16], [17], [21], [22].

---

DOI: 10.18421/TEM113-10

<https://doi.org/10.18421/TEM113-10>

**Corresponding author:** Nora Al-Garaawi,  
Department of Computer Science, Faculty of Education for Girls, University of Kufa, Najaf, Iraq.

**Email:** [noora.alagaraawi@uokuf.edu.iq](mailto:noora.alagaraawi@uokuf.edu.iq)

Received: 08 March 2022.

Revised: 12 June 2022.

Accepted: 17 June 2022.

Published: 29 August 2022.

 © 2022 Nora Al-Garaawi, Zainab Harbi & Tim Morris; published by UIKTEN. This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 License.

The article is published with Open Access at <https://www.temjournal.com/>

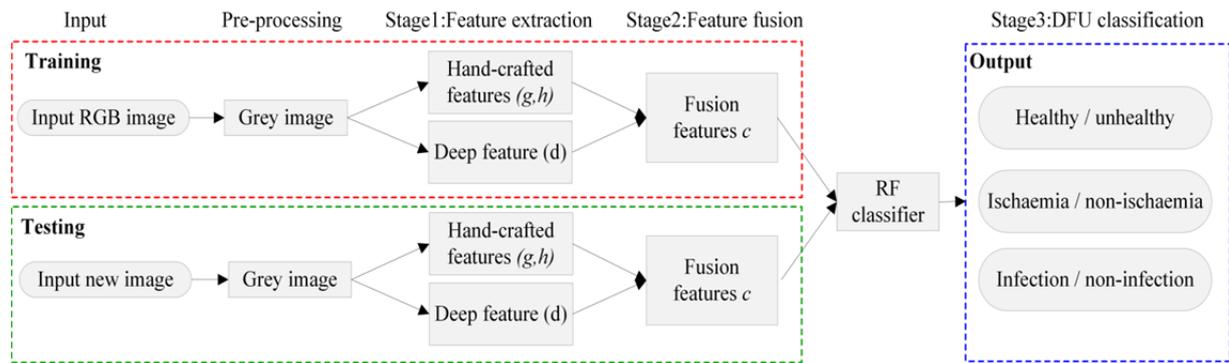


Figure 1. An overview of the proposed DFU classification system

Interestingly, in recent years the researchers' focus has been shifted to combine handcrafted features with deep features which have shown impressive performance in several computer vision tasks and applications such as face detection [23], facial expression recognition [23], age estimation [23], and image recognition [24]. In other words, the models that were trained on the fused features from both the handcrafted and deep features achieved significantly higher accuracy than those trained on the handcrafted features only or deep features only as in [22], [23], [24], [25], [26], [27]. Therefore, in this study, we propose to use the fused features from both the handcrafted and deep features for the recognition of DFU with the presence of ischaemia and infection. The underlying hypothesis of this feature fusion is that usually, the same pattern (e.g. DFU image) can be represented using several approaches and each representation (feature vector) reflects different characteristics of the same pattern. Therefore, when combining all feature vectors in a single feature vector, there is no doubt that the effective discriminant information of each vector will be kept to a certain degree in the resulting vector which is very important to increase the classification and recognition rate.

**Contribution:** for compact representation and accurate DFU recognition, in this paper, we present a new and automatic DFU recognition system in which the advantages of both handcrafted and deep features are combined together to represent the input image. To this end, in addition to the deep features, we use two of the widely used texture descriptors: HOG [28] and Gabor [29] descriptors for hand-crafted feature extraction. The extracted feature vectors from the HOG and Gabor descriptors are concatenated with the deep feature vector and the resulting vector is used for training an RF classifier which can then be used for DFU classification of a new image. Figure 1. shows an overview of the proposed method.

Experimental results on DFU datasets of healthy, unhealthy, ischaemic, non-ischaemic, infection, and non-infection classes showed that our proposed approach of using several feature vectors jointly achieved higher performance than that of using those feature vectors separately.

The rest of this paper is structured as follows. In Section 2, a brief background about the used methods and a description of the proposed methodology are given. In Section 3, the databases used to evaluate the proposed system are described. Section 4 reports experimental design and results. Section 5 presents conclusions of the proposed method.

## 2. The Proposed Method

The aim of this study is to investigate the advantages of fusing different feature vectors (handcrafted features and deep features) in the performance of DFU classification. To this end, we present a fully automatic DFU recognition system for classifying the input image into healthy, unhealthy, ischaemia, non-ischaemia, infection, and non-infection classes. The proposed system is divided into three main stages. These stages are feature extraction, feature fusion, and DFU classification. In the feature extraction stage, the hand-crafted (Gabor, HOG), and deep texture descriptors are used to extract the DFU features. In the feature fusion stage, the hand-crafted (Gabor, HOG) and the deep feature vectors are concatenated together to form the fusion feature vector. In the DUF classification stage, the fusion feature vectors are used to train a RF classifier which can be used then to classify the DFU into healthy, unhealthy, ischaemia, non-ischaemia, infection, and non-infection classes. Figure 1. describes the stages of the proposed system. See text in sections 2.1, 2.2, and 2.3 for more details about stage1, stage2, and stage3 of the proposed system respectively.

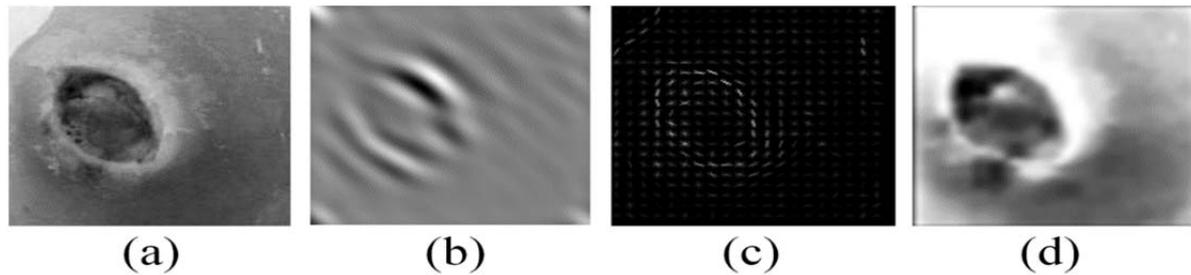


Figure 2. Example of hand-crafted and deep texture descriptors responses: (a) original image, (b) Gabor-based response, (c) HOG-based response, and (d) deep-based response

### 2.1. Stage 1: Feature Extraction

In the proposed study, three types of texture features are extracted: HOG feature [28], Gabor features [29], and deep features [30]. In the following, we describe those features briefly.

**HOG-based features:** we used the HOG texture descriptor developed in [28] to extract the DFU's HOG features  $h$ . HOG is one of the widely used feature extraction methods in the field of object detection and classification. It uses the distribution of the local gradient in the image to capture the structure and shape of an object with its gradient magnitude and direction. See reference [28] for more details regarding the HOG descriptor. In this study to extract the HOG features, we first convert the input image of the DFU to a grey image; we then divided the grey image into  $8 * 8$  pixel cells from which the HOG features are extracted. Figure 2. (c) shows an example of the HOG response of a DFU image which shows the texture that corresponds to the DFU area.

**Gabor-based features:** we used the Gabor Filter developed in [29] to extract the DFU's Gabor features  $g$ . The Gabor filter represents one of the commonly used texture descriptors in the field of image classification. It is a linear filter similar to the human visual system used to capture any specific frequency content found in a specific direction of an image. Mathematically, a 2D Gabor filter is a Gaussian function modulated by a sinusoidal wave and it consists of real and imaginary parts. The real part can be calculated as:

$$g_{\lambda,\theta,\sigma,\gamma}(x,y) = \exp\left(-\frac{x'^2 + \gamma y'^2}{2\sigma^2}\right) \cos\left(2\pi \frac{x'}{\lambda} + \phi\right)$$

Where  $\hat{x} = x\cos\theta + y\sin\theta$ ,  $\hat{y} = x\sin\theta + y\cos\theta$ ,  $\lambda$  is the wavelength of Gabor filter,  $\theta$  is the orientation of the normal to the stripes of the function,  $\phi$  is the phase offset,  $\gamma$  is the spatial ratio and  $\sigma$  is the standard deviation of the Gaussian envelope. See Reference [29] for more details about the Gabor filter. In this study, to calculate the Gabor features vector and the Gabor filter responses, we

first convert the RGB image to the grey image and then calculate eight filter banks of four orientations  $\theta: (0, \pi/4, \pi/2, 3\pi/4)$ , and two-phase offset  $\phi: (0, \pi/2)$ . The values of  $\sigma, \lambda, \gamma$  are set to 2, 2.5, and 0.3 respectively which are experimentally determined. Figure 2. (b) shows the response of the Gabor filter to the DFU image which reflects the texture that corresponds to the disease area.

**Deep-based features:** we used GoogLeNet CNN architecture developed in [30] to extract the deep features  $d$ . GoogLeNet was the winner in the ILSVRC-2014 challenge. Using the GoogLeNet model, both the multi-size inputs can be utilized and the pooling can be performed on the same input and at the same time. After that, the results will combine into a single feature layer in order to allow for the model to benefit from multi-level feature extraction of every input. The input layer of GoogLeNet takes an image size of  $224*224*3$  and the last layer is a softmax layer for classifying 1000 different classes. In this study, to extract the deep features, we first resize the DFU training images to  $224*224$  in order to fit the shape input of the GoogLeNet models. Second, we train the classification layer (last layer) on the two classes (binary classification). Third, the features of the last convolutional layer before the fully connected layers were extracted and used to train an RF classifier which can be used for DFU versus normal classification. Figure 2. (d) shows an example of deep features of the DFU image which show the texture that corresponds to the DFU area.

### 2.2. Stage 2: Feature fusion

Feature fusion is the process of combining several feature vectors into a single feature vector. The underlying hypothesis of the feature fusion is that usually, the same pattern (e.g. DFU image) can be represented using several approaches and each representation (feature vector) reflects different characteristics of the same pattern. Therefore, when combining all feature vectors in a single feature vector, there is no doubt that the effective discriminant information of each vector will be kept to a certain degree in the resulting vector which is

very important to increase the classification and recognition rate. Therefore, in this study we propose to combine the extracted features vectors  $g$ ,  $h$ ,  $d$  into a single feature vector  $c$  as is equation (2).

$$c = (g^T|h^T|d^T)^T \quad (2)$$

### 2.3. Stage 3: DFU classification

In this study, DFU, ischaemia, and infection classification were performed using the RF classifier developed in [31]. The RF classifier is chosen since it is one of the commonly and successfully used classifiers with many binary and non-binary classification problems, and it has maximum-margin separation with linearly separable and non-separable data such as DFU. See references [31] for more and complete details about the RF classifier, its parameters, and its application in several applications. In this study, the number of trees was set to 100 in all the experiments which gave the best results. Each tree makes a prediction and then the prediction of all trees will be combined into a single forest prediction and the average is calculated by:

$$P(y|v) = \frac{1}{R} \sum_{r=1}^R P(y|v) \quad (3)$$

Where  $R$  is the number of trees in the forest,  $p(y|v)$  is the probability of the class  $y$  given the feature vector  $v$ .

### 3. Model Formal

Suppose we have  $N$  labelled training images, where for each image  $I_i$  we have an index  $y_i$  indicating the DFU class and a feature vector  $v_i(h, g, d, c)$ . Thus  $y \in \{0, 1\}$  for healthy and unhealthy classes, ischaemic and non-ischaemic classes, or infected and non-infected classes. The  $h, g, d$ , and  $c$  indicate HOG features vector  $h$ , the Gabor features vector  $g$ , deep features vector  $d$  and their combination vector  $c$  respectively. Given the extracted features  $h, g, d, c$  for each image  $I_i$  and their corresponding class  $y_i$ , training an RF classifier function  $P$  on each feature vector  $v_i$  separately is one approach to predict the class  $y$  of new images  $I$  as follows:

$$y(I|h) = \max_y P(y|h) \text{ or} \quad (4)$$

$$y(I|g) = \max_y P(y|g) \text{ or} \quad (5)$$

$$y(I|d) = \max_y P(y|d) \quad (6)$$

Where  $y(I)$  is the most probable class of the input image being class  $y$ .  $P(y|h)$  is the probability of the input image being class type  $y$  given the HOG feature vector  $h$ .  $P(y|g)$  is the probability of the input image being class type  $y$  given the Gabor feature vector  $g$ .  $P(y|d)$  is the probability of the input image being class type  $y$  given the deep feature vector  $d$ . Another approach is to train the classifier function  $P$  on the fusion vector  $c$  as in equation (7):

$$y(I|c) = \max_y P(y|c) \quad (7)$$

Where  $P(y|c)$  is the probability of the input image being class type  $y$  given the fusion feature vector  $c$ . Fusing all the feature vectors into a single feature vector in this way is important since different information regarding the disease can be gained and the unhealthy region whose appearance looks like a normal region can be dealt with more effectively leading to a robust recognition.

### 4. Database

In this paper, Part-A [2] and Part-B [3] DFU datasets that were collected in Lancashire teaching hospitals in the United Kingdom (UK) were used to evaluate the performance of the proposed method. The Part-A dataset contains 1,679 images divided into 641 foot images of healthy class and 1038 foot images with Ulcer (DFU) [2]. The Part-B dataset contains ischaemic parts and infected parts which are two types of DFU diseases. The ischaemic part contains 9870 foot images divided equally into ischaemic and non-ischaemic foot images. The infection part contains 5892 foot images also divided equally into images of infected and non-infected feet. Some examples of the Part-A and Part-B datasets are shown in Figure 3. Table 1. describes these databases and their particularities [3].

Table 1. Datasets description

Database	classes	Images	labels
Part-A	Healthy	641	0
Part-A	Unhealthy	1038	1
Part-B	Ischaemia	4935	0
Part-B	Non-ischaemia	4935	1
Part-B	Infection	2946	0
Part-B	Non-infection	2946	1



Figure 3. Sample images of part-A and Part-B datasets

## 5. Experiments Design and Results

In this section, we report a series of experiments to evaluate and compare the performance of using handcrafted features and deep features jointly to the performance of using them separately in the DFU recognition task with the existence of ischaemia and infection classes. The first experiment is focused on assessing the proposed method in recognizing healthy and unhealthy (DFU) classes. The second experiment is focused on assessing the proposed method in recognizing ischaemic versus non-ischaemic classes. The third experiment is focused on evaluating the proposed method in recognizing infected versus non-infected classes.

Evaluation Metric: in this study, the area under the curve (AUC) of the receiver operating characteristic (ROC) was used to evaluate the performance of the proposed approach. The process of calculating the ROC is performed by determining the accuracy of the classifier between two classes (binary classification) and then plotting the true positive rate versus the false positive rate. In addition to the ROC, we calculate sensitivity (SEN), specificity (SPE), precision (PRE), accuracy (ACC), and F-measure (F) calculated using Eq. (8), Eq. (9), Eq. (10), Eq. (11), and Eq. (12) respectively.

$$\text{sensitivity} = \frac{TP}{TP + FN} \quad (8)$$

$$\text{specificity} = \frac{TN}{FP + FN} \quad (9)$$

$$\text{precision} = \frac{TP}{TP + FP} \quad (10)$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (11)$$

$$F - \text{measure} = \frac{2 * TP}{2 * TP + FP + FN} \quad (12)$$

Where TP, FN, FP, and TN referred to a true positive, false negative, false negative, and true negative respectively. In all experiments, we performed 5-fold cross-validation experiments and then calculated a mean accuracy, standard deviation, and AUC of the 5 folds. During each fold, images of each dataset are divided into the training set (80% images), validation set (10% images), and testing set (10% images). After that, we interchanged the validation and testing sets in order to make sure that we tested every image in the dataset exactly once.

### 5.1. Healthy Versus Unhealthy Classification Results

Using the Part-A dataset of healthy and unhealthy classes, we extracted four different feature vectors: Hog feature vector  $h$ , Gabor feature vector  $g$ , deep feature vector  $d$ , and fusion feature vector  $c$ . Given the extracted features, we trained four RF classifiers: HOG-based RF classifier, Gabor-based RF classifier, deep-based RF classifier and fusion-based RF classifier using equations 4, 5, 6, and 7 respectively which are used then to estimate the probability of healthy versus unhealthy classes. Figure 4. and Table 2. show ROC curves and cross-validation mean results. These results demonstrated that training the RF classifier using HOG, Gabor, and deep features jointly have better performance than using those features separately (see ROC curves in Figure 4.). Results in Table 2. showed that training the RF classifier on the fusion features vector significantly increased the AUC from 0.88, 0.90, and 0.92 using HOG features, Gabor features, and deep features, respectively, to 0.94 using those features jointly.

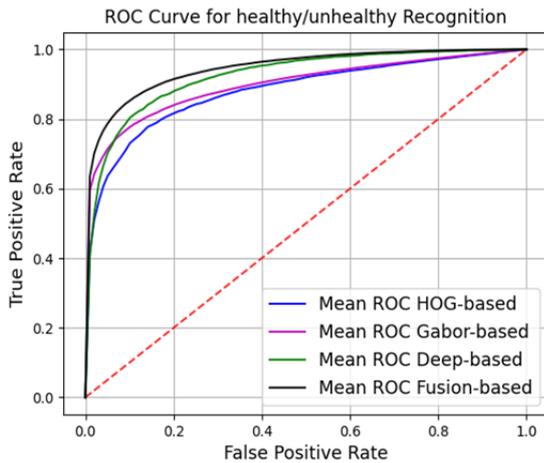


Figure 4. ROC curves comparison of four different features in healthy versus non-healthy classification: HOG features, Gabor features, deep features and fusion features. Overall the fusion features method achieved the best results

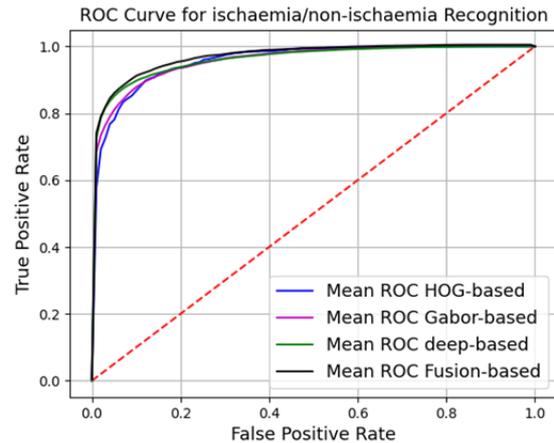


Figure 5. ROC curves comparison of four different features in ischaemia versus non-ischaemia classification: HOG features, Gabor features, deep features and fusion features. Overall the fusion features method achieved the best results

Moreover, the recognition results of the proposed system using combined features using the part-A dataset achieved improvements of 1%, 1%, 1%, 2%, 2%, and 2% in SEN, SPE, PRE, ACC, and AUC respectively, as illustrated in Table 2. when compared to the HOG, Gabor, or deep features only.

### 5.2. Ischaemia Versus non-ischaemia Classification Results

Using the part-B dataset, practically ischaemia/non-ischaemia part, we extracted four different feature vectors: HOG feature vector  $h$ , Gabor feature vector  $g$ , deep feature vector  $d$ , and fusion feature vector  $c$ . Given the extracted features, we trained four RF classifiers: HOG-based RF classifier, Gabor-based RF classifier, deep-based RF classifier and fusion-based RF classifier using equations 4, 5, 6, and 7 respectively which are used then to estimate the probability of ischaemia versus non-ischaemia classes. Figure 5. and Table 3. show ROC curves and cross-validation mean results. These results demonstrated that training the RF classifier using HOG, Gabor, and deep features jointly have better performance than using those features separately (see ROC curves in Figure 5.). Results in Table 3. showed that training the RF classifier on the fusion feature vector significantly increased the AUC from 0.95 using HOG features, 0.96 using Gabor features, and 0.96 using deep features to 0.97 using those features jointly. Moreover, the recognition results of the proposed system using combined features on ischaemia dataset achieved improvements of 2%, 1%, 1%, 2%, 1%, and 1% in SEN, SPE, PRE, ACC, and

AUC respectively, as illustrated in Table 3. when compared to the HOG, Gabor, or deep features only.

### 5.3. Infection Versus non-infection Classification Results

Using part-B dataset, the infection/non-infection part, we extracted four different feature vectors: HOG feature vector  $h$ , Gabor feature vector  $g$ , deep feature vector  $d$ , and fusion feature vector  $c$ . Given the extracted features, we trained four RF classifiers: HOG-based RF classifier, Gabor-based RF classifier, deep-based RF classifier and fusion-based RF classifier using equations 4, 5, 6, and 7 respectively which are used then to estimate the probability of infection versus non-infection. Figure 6. and Table 4. show ROC curves and cross-validation mean results. These results demonstrated that training the RF classifier using HOG, Gabor, and deep features jointly have better performance than using those features separately (see ROC curves in Figure 6.). Results in Table 4. showed that training the RF classifier on the fusion feature vector significantly increased the AUC from 0.73% using HOG features, 0.77% using Gabor features, and 0.78 using deep features to 0.81% using those features jointly. Moreover, the recognition results of the proposed system using combined features on infection dataset achieved improvements of 5%, 3%, 5%, 5%, 1%, and 3% in SEN, SPE, PRE, ACC, and AUC respectively, as illustrated in Table 4. when compared to the HOG, Gabor, or deep features only.

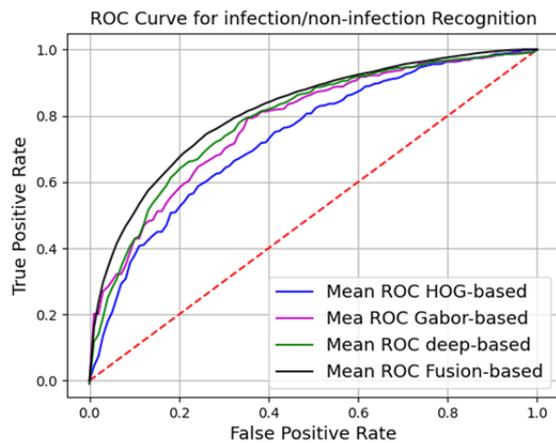


Figure 6. ROC curves comparison of four different features in infection versus non-infection classification: HOG features, Gabor features, deep features and fusion features. Overall the fusion features method achieved the best results

Results in Table 4. showed that training the RF classifier on the fusion feature vector significantly increased the AUC from 0.73% using HOG features, 0.77% using Gabor features, and 0.78 using deep features to 0.81% using those features jointly. Moreover, the recognition results of the proposed system using combined features on infection dataset achieved improvements of 5%, 3%, 5%, 5%, 1%, and 3% in SEN, SPE, PRE, ACC, and AUC respectively, as illustrated in Table 4. when compared to the HOG, Gabor, or deep features only.

#### 5.4. Comparison with Baseline

In Table 5., we compare our results to the results of other approaches using the same dataset. These results demonstrate that the results of our approach are satisfactory and higher than others especially with the presence of ischaemia and infection classes.

## 6. Discussion

In this study, a fully automatic DFU classification system is proposed which can be used for early DFU detection and to reduce the negative complications of the disease. In the proposed system since the texture and deep features of the DFU image hold valuable information regarding the DFU disease, we argue that using all the image features by combining both the texture and deep features is important for an accurate and robust DFU classification. In this study, a fully automatic DFU classification system is proposed which can be used for early DFU detection

and to reduce the negative complications of the disease. In the proposed system since the texture and deep features of the DFU image hold valuable information regarding the DFU disease, we argue that using all the image features by combining both the texture and deep features is important for an accurate and robust DFU classification. In section 5, we reported extensive experiments using two of the recently publish DFU datasets to demonstrate the effectiveness of the proposed hypothesis in the DFU classification problem. We demonstrated that training the classifier on the fusion features of both the hand-crafted and deep features greatly improved the recognition rate of DFU classification. This is clearly due to the participation of several features which helped to increase the discriminant information regarding the diseases. In addition, it can be seen from experimental results that the recognition result of deep-based DFU recognition is slightly better than that of both Gabor-based and HOG-based features.

## 7. Conclusion

Recently, most DFU classification systems use deep features or hand-crafted features for DFU classification. Since both features are extracted from the same pattern and contain valuable information of the DFU disease, in this study, we performed several experiments to investigate the advantages of using both the hand-crafted and deep features jointly on the performance of automatic DFU recognition. The hand-crafted features are extracted using two of the widely used texture descriptors: HOG and Gabor, the deep features are extracted using GoogLNet network. We have shown that using the fusion features of the hand-crafted and deep features have a significant effect on the accuracy of DFU recognition. The proposed system achieved satisfactory performance that can generalize well across different DFU classes including healthy vs. unhealthy, ischaemic vs non-ischaemic, infected vs non-infected classes. Experimental results using two hand-crafted features, deep features, and their combination on two DFU datasets demonstrated that using the fusion of hand-crafted and deep features reported better performance than using those features separately. That performance gain is due to the discriminant information of the same pattern that is extracted using different feature vectors.

Table 2. DFU classification results: Comparison among Gabor-based, HOG-based, deep-based, and fusion-based RF classifiers

Features	SEN	SPE	PRE	ACC	F	AUC
HOG-based	0.89± 0.07	0.73 ± 0.20	0.86 ± 0.03	0.84 ± 0.04	0.86 ± 0.04	0.88 ± 0.15
Gabor-based	0.89 ± 0.03	0.71 ± 0.34	0.87 ± 0.19	0.85 ± 0.19	0.89 ± 0.07	0.90 ± 0.16
Deep-based	0.86 ± 0.01	0.83 ± 0.04	0.87 ± 0.04	0.86 ± 0.05	0.87 ± 0.04	0.92 ± 0.01
Fusion-based	0.90 ± 0.04	0.84 ± 0.05	0.88 ± 0.05	0.88 ± 0.08	0.89 ± 0.03	0.94 ± 0.020

Table 3. Ischaemia classification results: Comparison among Gabor-based, HOG-based, deep-based, and fusion-based RF classifiers

Features	SEN	SPE	PRE	ACC	F	AUC
HOG-based	0.89± 0.12	0.88 ± 0.03	0.92 ± 0.01	0.89 ± 0.04	0.90 ± 0.06	0.95 ± 0.01
Gabor-based	0.91 ± 0.05	0.87 ± 0.04	0.93 ± 0.09	0.90 ± 0.07	0.92 ± 0.05	0.96 ± 0.06
Deep-based	0.91 ± 0.03	0.88 ± 0.02	0.93 ± 0.03	0.90 ± 0.02	0.92 ± 0.01	0.96 ± 0.01
Fusion-based	0.93 ± 0.02	0.90 ± 0.05	0.94 ± 0.03	0.92 ± 0.02	0.93 ± 0.08	0.97 ± 0.02

Table 4. Infection classification results: Comparison among Gabor-based, HOG-based, deep-based, and fusion-based RF classifiers

Features	SEN	SPE	PRE	ACC	F	AUC
HOG-based	0.68± 0.01	0.66 ± 0.06	0.67 ± 0.02	0.67 ± 0.04	0.69 ± 0.06	0.73 ± 0.02
Gabor-based	0.68 ± 0.08	0.67 ± 0.04	0.68 ± 0.07	0.67 ± 0.08	0.75 ± 0.03	0.77 ± 0.03
Deep-based	0.69 ± 0.01	0.68 ± 0.05	0.68 ± 0.02	0.68 ± 0.03	0.75 ± 0.05	0.78 ± 0.01
Fusion-based	0.74 ± 0.02	0.71 ± 0.05	0.73 ± 0.02	0.73 ± 0.01	0.76 ± 0.03	0.81 ± 0.06

Table 5. Comparison with other method

Dataset	Reference	Method	Sen	Spe	Pre	Acc	F-Measure	AUC
DFU	[2]	LBP	0.92	0.76	0.88	0.87	0.90	0.93
		LBP + HOG	0.88	0.84	0.91	0.87	0.89	0.93
		LBP + HOG + Colour	0.90	0.85	0.90	0.88	0.90	0.94
		LeNet (CNN)	0.91	0.81	0.87	0.87	0.89	0.93
	[5]	VGG16	0.90	-	0.92	-	0.91	-
		Alexnet (CNN)	0.87	-	0.91	-	0.89	-
		GoogLeNet (CNN)	0.91	-	0.96	-	0.93	-
		DFU-QUTNet	0.94	-	0.95	-	0.95	-
		Ours	0.90	0.84	0.88	0.88	0.89	0.94
Ischaemia	[3]	Ensemble (CNN)	0.89	0.92	0.92	0.90	0.90	0.90
		Ours	0.93	0.90	0.94	0.92	0.93	0.97
Infection	[3]	Ensemble (CNN)	0.80	0.74	0.74	0.73	0.72	0.73
		Ours	0.74	0.71	0.73	0.73	0.76	0.81

## References

- [1]. 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(5), 1047-1053.
- [2]. Goyal, M., Reeves, N. D., Davison, A. K., Rajbhandari, S., Spragg, J., & Yap, M. H. (2018). Dfunet: Convolutional neural networks for diabetic foot ulcer classification. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 4(5), 728-739.
- [3]. 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.
- [4]. 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(6), 1762.
- [5]. Alzubaidi, L., Fadhel, M. A., Olewi, 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(21), 15655-15677.
- [6]. Al-Garaawi, N., Wu, Q., & Morris, T. (2021). BRIEF-based face descriptor: an application to automatic facial expression recognition (AFER). *Signal, Image and Video Processing*, 15(2), 371-379.
- [7]. Algaraawi, N. (2019). *Modelling of human ageing, compound emotions, and intensity for automatic facial expression recognition*. The University of Manchester (United Kingdom).
- [8]. Nailon, W. H. (2010). Texture analysis methods for medical image characterisation. *Biomedical imaging*, 75, 100.
- [9]. 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(1), 151-165.
- [10]. Erickson, B. J., Korfiatis, P., Akkus, Z., & Kline, T. L. (2017). Machine learning for medical imaging. *Radiographics*, 37(2), 505.
- [11]. Algaraawi, N., & Morris, T. (2016). Study on aging effect on facial expression recognition. In *Proceedings of the World Congress on Engineering* (Vol. 1).
- [12]. 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.
- [13]. 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*, 8, 228586-228597.
- [14]. Eid, M. M., Yousef, R. N., & Mohamed, M. A. (2018). A proposed automated system to classify diabetic foot from thermography. *International Journal of Scientific & Engineering Research*, 9(12), 371-381.
- [15]. Goyal, M., Reeves, N. D., Rajbhandari, S., & Yap, M. H. (2018). Robust methods for real-time diabetic foot ulcer detection and localization on mobile devices. *IEEE journal of biomedical and health informatics*, 23(4), 1730-1741.
- [16]. Jawahar, M., Anbarasi, L. J., Jasmine, S. G., & Narendra, M. (2020, June). Diabetic foot ulcer segmentation using color space models. In *2020 5th international conference on communication and electronics systems (ICCES)* (pp. 742-747). IEEE.
- [17]. Song, A., Zhu, H., Huang, X., Xu, X., Liu, L., & Chen, Y. Cascade Attention DetNet: Object Detection For Diabetic Foot Ulcer.
- [18]. Castellano, G., Bonilha, L., Li, L. M., & Cendes, F. (2004). Texture analysis of medical images. *Clinical radiology*, 59(12), 1061-1069.
- [19]. Chabat, F., Yang, G. Z., & Hansell, D. M. (2003). Obstructive lung diseases: texture classification for differentiation at CT. *Radiology*, 228(3), 871-877.
- [20]. He, D. C., & Wang, L. (1990). Texture unit, texture spectrum, and texture analysis. *IEEE transactions on Geoscience and Remote Sensing*, 28(4), 509-512.
- [21]. Albawi, S., Mohammed, T. A., & Al-Zawi, S. (2017, August). Understanding of a convolutional neural network. In *2017 international conference on engineering and technology (ICET)* (pp. 1-6). IEEE.
- [22]. Al-Garaawi, N., Ebsim, R., Alharan, A. F., & Yap, M. H. (2022). Diabetic foot ulcer classification using mapped binary patterns and convolutional neural networks. *Computers in Biology and Medicine*, 140, 105055.
- [23]. 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*.
- [24]. Sun, Q. S., Zeng, S. G., Liu, Y., Heng, P. A., & Xia, D. S. (2005). A new method of feature fusion and its application in image recognition. *Pattern Recognition*, 38(12), 2437-2448.
- [25]. Levi, G., & Hassner, T. (2015, November). 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).
- [26]. Muhammad Anwer, R., Khan, F. S., van de Weijer, J., & Laaksonen, J. (2017, June). Tex-nets: Binary patterns encoded convolutional neural networks for texture recognition. In *Proceedings of the 2017 ACM on International Conference on Multimedia Retrieval* (pp. 125-132).
- [27]. 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.

- [28]. Dalal, N., & Triggs, B. (2005, June). Histograms of oriented gradients for human detection. In *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)* (Vol. 1, pp. 886-893). IEEE.
- [29]. Petkov, N. (1995). Biologically motivated computationally intensive approaches to image pattern recognition. *Future Generation Computer Systems*, 11(4-5), 451-465.
- [30]. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... & Rabinovich, A. (2015). Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1-9).
- [31]. Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5-32.