Recognition of Ischaemia and Infection in Diabetic Foot Ulcers: Dataset and Techniques

Manu Goyal[a], Neil D. Reeves[b], Satyan Rajbhandari[c], Naseer Ahmad[d], Chuan Wang[e], Moi Hoon Yap[a,*]

[a] *Centre for Advanced Computational Sciences, Manchester Metropolitan University, M1 5GD, Manchester, UK*
[b] *Research Centre for Musculoskeletal Science & Sports Medicine, Manchester Metropolitan University, M1 5GD, Manchester, UK.*
[c] *Lancashire Teaching Hospital, PR2 9HT, Preston, UK.*
[d] *University of Manchester and Manchester Royal Infirmary, M13 9WL, Manchester, UK.*
[e] *Department of Endocrinology, Sun Yat-sen Memorial Hospital, Sun Yat-sen University, Guangzhou 510120, P.R. China.*

Abstract

Recognition and analysis of Diabetic Foot Ulcers (DFU) using computerized methods is an emerging research area with the evolution of image-based machine learning algorithms. Existing research using visual computerized methods mainly focuses on recognition, detection, and segmentation of the visual appearance of the DFU as well as tissue classification. According to DFU medical classification systems, the presence of infection (bacteria in the wound) and ischaemia (inadequate blood supply) has important clinical implications for DFU assessment, which are used to predict the risk of amputation. In this work, we propose a new dataset and computer vision techniques to identify the presence of infection and ischaemia in DFU. This is the first time a DFU dataset with ground truth labels of ischaemia and infection cases is introduced for research purposes. For the handcrafted machine learning approach, we propose a new feature descriptor, namely the Superpixel Color Descriptor. Then we use the Ensemble Convolutional Neural Network (CNN) model for more effective recognition of ischaemia and infection. We propose to use a natural data-augmentation method, which

*Corresponding author: Tel.: +44 161 247 1503;
Email address: `M.Yap@mmu.ac.uk` (Moi Hoon Yap)

identifies the region of interest on foot images and focuses on finding the salient features existing in this area. Finally, we evaluate the performance of our proposed techniques on binary classification, i.e. ischaemia versus non-ischaemia and infection versus non-infection. Overall, our method performed better in the classification of ischaemia than infection. We found that our proposed Ensemble CNN deep learning algorithms performed better for both classification tasks as compared to handcrafted machine learning algorithms, with 90% accuracy in ischaemia classification and 73% in infection classification.

Keywords:
Diabetic foot ulcers, deep learning, ischaemia, infection, machine learning.

1. Introduction

Diabetic Foot Ulcers (DFUs) are a major complication of diabetes which can lead to amputation of the foot or limb. Treatment of Diabetic foot ulcers is a global major health care problem resulting in high care costs and mortality rate. Recognition of infection and ischaemia is very important to determine factors that predict the healing progress of DFU and the risk of amputation. Ischaemia, the lack of blood circulation, develops due to chronic complications of diabetes. This can result in gangrene of the diabetic foot ulcer, which may require amputation of the part of the foot or leg if not recognised and treated early. Detailed knowledge of the vascular anatomy of the leg, and particularly ischaemia enables medical experts make better decisions in estimating the possibility of DFU healing, given the existing blood supply [1]. In previous studies, it is estimated that patients with critical ischaemia have a three-year limb loss rate of about 40% [2]. Patients with an active DFU and particularly those with ischaemia or gangrene should be checked for the presence of infection. Approximately, 56% of DFU become infected and 20% of DFU infections lead to amputation of a foot or limb [3, 4, 5]. In one recent study, 785 million patients with diabetes in the US between 2007 and 2013 suggested that DFU and associated infections constitute a powerful risk factor for emergency department visits and hospital admission [6].

There are a number of DFU classification systems such as Wagner, University of Texas, and SINBAD Classification systems, which include information on the site of DFU, area, depth, presence of neuropathy, presence of ischaemia, and infection [7, 8, 9]. SINBAD stands for S (Site), I (Ischaemia), N (Neuropathy), B (Bacterial infection), A (Area), D (Depth). This paper focuses on ischaemia and infection, which are defined as follow:

1. Ischaemia: Inadequate blood supply that could affect DFU healing. Ischaemia is diagnosed by palpating foot pulses and measuring blood pressure in the foot and toes. The visual appearance of ischaemia might be indicated by the presence of poor reperfusion to the foot, or black gangrenous toes (tissues death to part of the foot). From a computer vision perspective, these might be important hints of the presence of ischaemia in the DFU.
2. Bacterial Infection: Infection is defined as bacterial soft tissue or bone infection in the DFU, which is based on the presence of at least two classic findings of inflammation or purulence. It is very hard to determine the presence of diabetic foot infections from DFU images, but increased redness in and around ulcer and coloured purulent could provide indications. In the medical system, blood testing is performed as the gold standard diagnostic test. Also, in the present dataset, the images were captured after the debridement of necrotic and devitalized tissues which removes an important indication of the presence of infection in DFU.

In related work, Netten et al. [10] find that clinicians achieved low validity and reliability for remote assessment of DFU in foot images. Hence, it is clear that analysing these conditions from images is a difficult task for clinicians. In various image recognition applications, such as medical imaging and natural language processing tasks, machine learning algorithms performed better than skilled humans including clinicians [11, 12, 13].

The previous state-of-the-art image-based computer-aided diagnosis of DFU is composed of multiple stages, including image pre-processing, image segmentation, feature extraction, and classification. Veredas et al. [14] proposed the use of color and texture features from the segmented area and multi-layer neural network to perform the tissue classification to distinguish between healing-tissue and skin for healing prediction. Wannous et al. [15] performed tissue classification from color and texture region descriptors on a 3-D model for the wound. Wang et al. [16] used a cascaded two-stage classifier to determine the DFU boundaries for area determination of DFU. Major progress in the field of image-based machine learning, especially deep learning algorithms, allows the extensive use of medical imaging data with end-to-end models to provide better diagnosis, treatment, and prediction of diseases [17, 18]. Deep learning models for DFU, predominantly led by works from our laboratory have achieved high accuracy in the recognition of DFUs with machine learning algorithms [19, 20, 21, 22].

The major issues and challenges involved with the assessment of DFU using machine learning methods from foot images are as follows: 1) a major time-burden involved in data collection and expert labelling of the DFU images; 2) high inter-class similarity and intra-class variations are dependent upon the different classification of DFU; 3) non-standardization of the DFU dataset, such as distance of the camera from the foot, orientation of the image and lighting conditions; 4) lack of meta-data, such as patient ethnicity, age, sex and foot size.

Accurate diagnosis of ischaemia and infection requires establishing a good clinical history, physical examination, blood tests, bacteriological study and Doppler study of leg blood vessels. These tests and resources are not always available to clinicians across the world and hence the need for a solution to inform diagnosis, such as the one we proposed in this paper. Experts working in the field of diabetic foot ulceration have good experience of predicting the presence of underlying ischaemia or infection simply by looking at the ulcer. We aim to replicate that in machine learning. To increase the reliability of the annotation, two experts predict the presence of ischaemia and infection from DFU images. Due to high risks of infection and ischaemia in DFU leading to patient's hospital admission, and amputation [23], recognition of infection and ischaemia in DFU with cost-effective machine learning methods is a very important step towards the development of complete computerized DFU assessment system for remote monitoring in the future.

2. DFU Dataset and Expert Labelling

For binary classification of ischaemia and infection in DFU, we introduce a dataset of 1459 images of patient's foot with DFU over the previous five years at the Lancashire Teaching Hospitals, obtaining ethical approval from all relevant bodies and patients written informed consent. Approval was obtained from the NHS Research Ethics Committee to use these images for this research. These DFU images were captured with different cameras (Kodak DX4530, Nikon D3300, and Nikon COOLPIX P100). The current dataset

(a) Infection

(b) No Infection

(c) Ischaemia

(d) No Ischaemia

Figure 1: Examples of foot images with DFU used for binary expert annotations for infection and ischaemia.

we received with the ethical approval from NHS did not contain any records or meta-data about these conditions or any medical classification.

Since there is no clinical meta-data regarding this DFU dataset, the experiment is performed on the images with handcrafted traditional machine learning and deep learning. This is the first time, recognition of ischaemia and infection in DFU is performed based on images, hence, there is no publicly available dataset. Here, we introduce the first DFU dataset with ground truth labels of ischaemia and infection cases. Expert labelling of each DFU

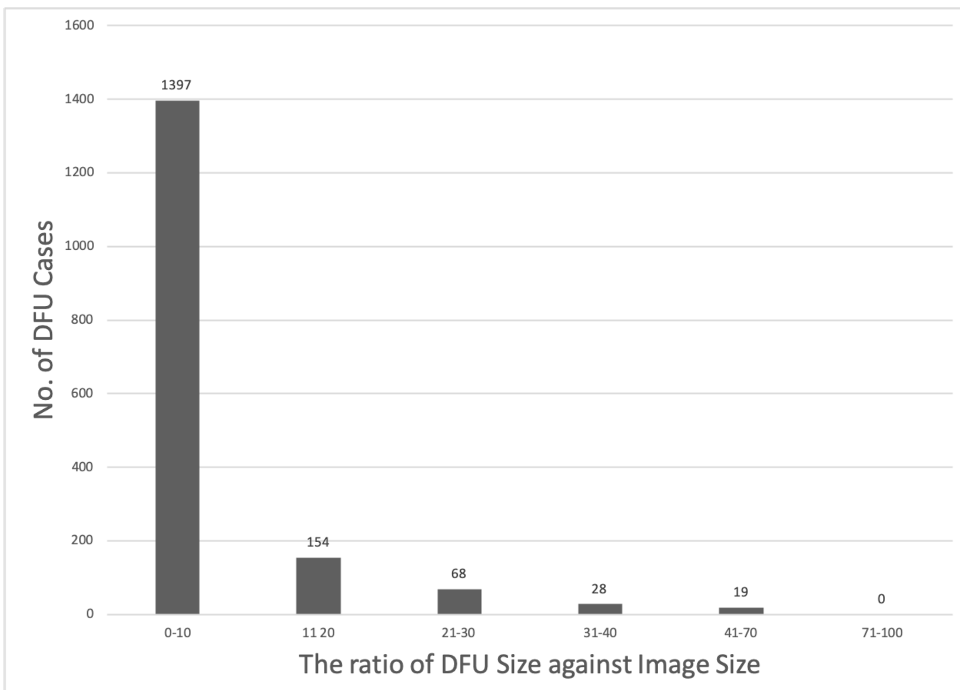


Figure 2: The number of DFU cases according to the area of DFU in full foot image of the DFU dataset.

according to the different conditions present in DFU according to the popular medical classification system on this DFU dataset is particularly important for this task. The ground truth was produced by two healthcare professionals (consultant physicians with specialisation in the diabetic foot) on the visual inspection of DFU images. Where there was disagreement for the ground truth, the final decision was made by the more senior physician. These ground truths are used for the binary classification of infection and ischaemia of DFU. A few examples of foot images with DFU used for binary expert annotation are shown in Fig. 1. The complete number of cases of expert annotation of each condition is detailed in Table 1. The dataset, alongside its ground truth labels, will be made available upon acceptance of this article.

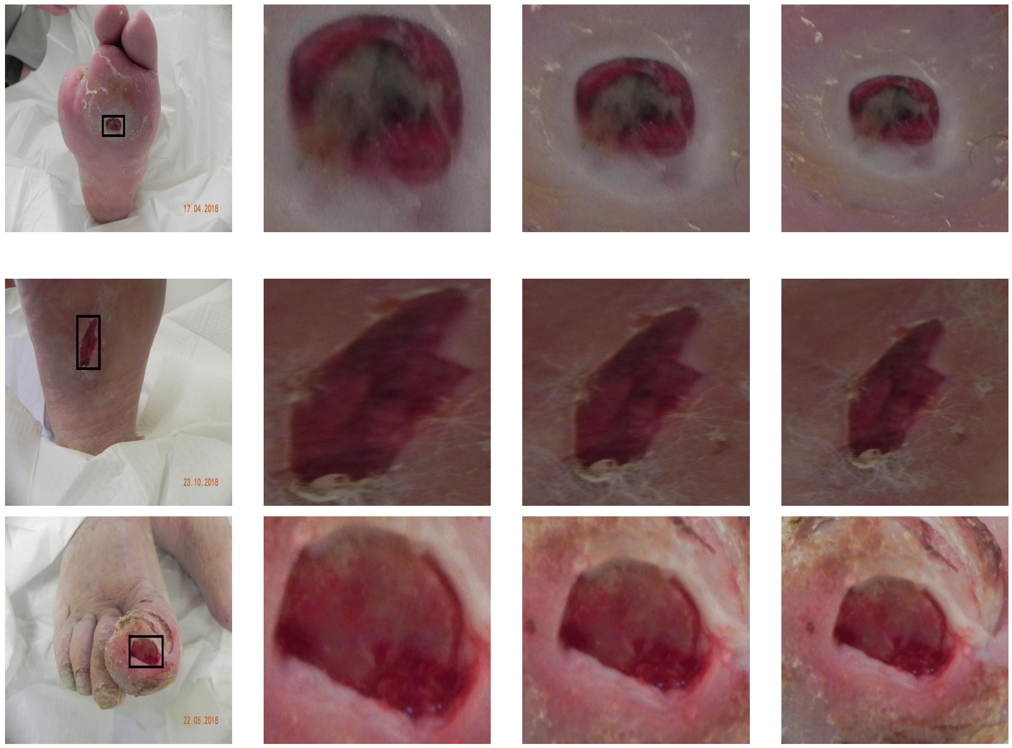
3. Methodology

This section describes our proposed techniques for the recognition of ischaemia and infection of the DFU diagnosis system. The preparation of a balanced dataset, handcrafted features, and machine learning methods (handcrafted machine learning and deep learning approaches) used for binary classification of ischaemia and infection are detailed in this section.

3.1. Natural Data-Augmentation Technique based on Deep Learning Algorithm

This section describes our proposed data augmentation method, called Natural Data-augmentation, which is based on deep DFU localization algorithm (Faster R-CNN).

In the DFU dataset, the images (size)varies between 1600×1200 and 3648×2736) depending on the cameras used to capture the data. In deep learning, data augmentation is envisioned as an important tool to improve the performance of algorithms. As shown in Fig. 2, approximately 92% of DFU cases have area between 0% to 20% on foot images. In common data-augmentation, the number of techniques used such as flip, rotation, random scale, random crop, translation, and Gaussian noise to perform augment in the dataset. Since DFU occupies a very small percentage of the total area of foot images, there is a risk of missing the region of interests by using important augmentation technique such as random scale, crop, and translation. Hence, Natural Data-augmentation is more suitable for the DFU evaluation rather than common data-augmentation. This augmentation technique helps

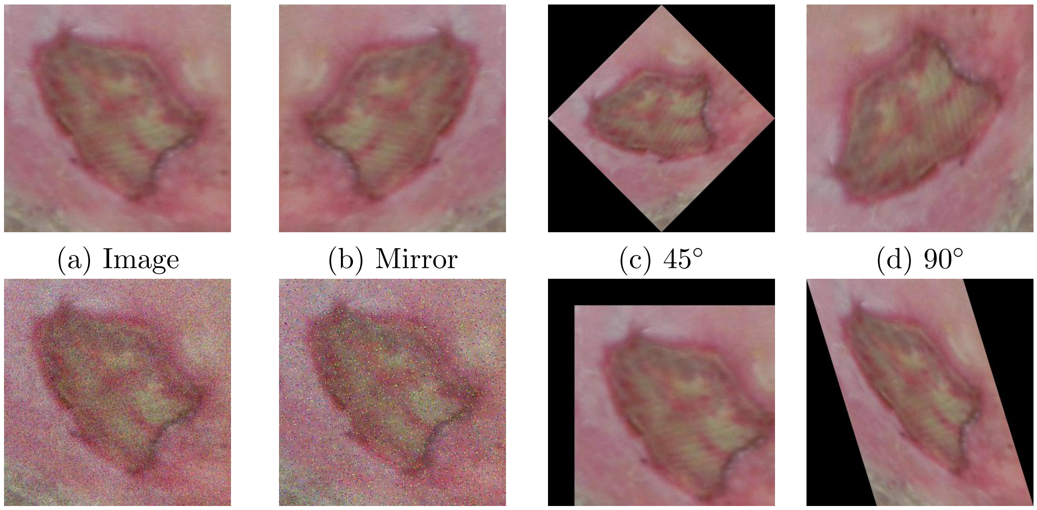


(a) Image (b) Ist MAG (c) 2nd MAG (d) 3rd MAG

Figure 3: Natural Data-augmentation produced from the original image with different magnifications (three magnifications in this experiment). MAG refers to magnification

in assisting the machine algorithms to pinpoint ROI of DFU on foot images and focus on finding the strong features that exists in this area. We used the deep learning-based localization method, Faster-RCNN with Inception-ResNetV2, to get ROI of the DFU on foot images [24, 25]. Depending upon the size of DFU and image, the natural data-augmentation on the DFU dataset with different magnification is demonstrated in Fig. 3. Flexible parameters can be used to choose the number of magnification factors (3 in this classification), as well as magnification distance, which can be adjusted from a single DFU image by natural augmentation. After magnification, further, data-augmentation is achieved with the help of angles, mirror, gaussian noise, contrast, sharpen, translation, shearing using our proposed methods as shown in Fig. 4.

As shown in Table 1, the number of DFU patches generated by cropping multiple DFU on foot images and augmented patches are generated



(a) Image (b) Mirror (c) 45° (d) 90°

(e) Gaussian Noise (f) Salt and pepper (g)Translate (h)Shear

Figure 4: After magnification, different types of data-augmentation is achieved by the proposed Natural Data-augmentation

by natural data-augmentation (Fig. 3) and different data augmentations (Fig. 4). The total number of cases for ischaemia and non-ischaemia in this DFU dataset is imbalanced (1249 cases vs 210 cases) whereas infection (628 cases) and non-infection (831 cases) are fairly balanced as shown in Table 1. We performed binary classification of ischaemia and infection with machine learning algorithms because for multi-class classification, this DFU dataset is imbalanced especially for cases (Ischaemia and No Infection) as shown in 5.

3.2. Handcrafted Superpixel Color Descriptors

We investigated the use of human design features with traditional machine learning on the binary classification of infection and ischaemia. Our first attempt was experimenting with texture descriptors (Local Binary Patterns and Histogram of Gradient) and color descriptors as used in related works [19, 21]. However, we achieved very poor results for these binary classification problems. Hence, we propose a novel Superpixel Color Descriptors (SPCD) to extract the colors region of interest from DFU images that could be the important visual cues for the identification of ischaemia and infection in DFU. In the first step, we used a SLIC superpixels technique to produce superpixel over-segmentation of DFU patches based on pixel color and intensity values

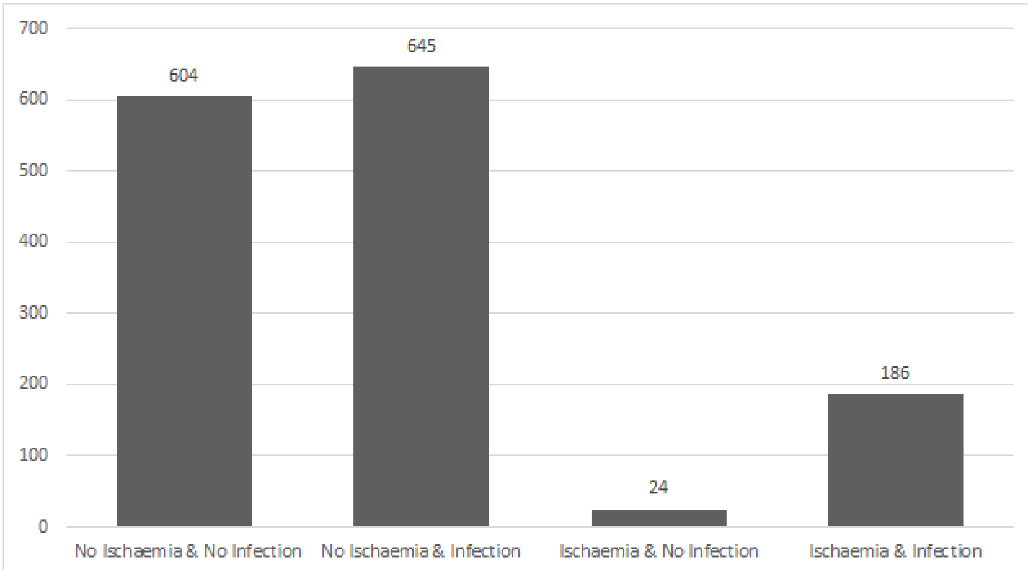


Figure 5: Distribution of ischaemia and infection cases as multi-class classification problem.

Table 1: The number of Infection and ischaemia cases, number of DFU patches and augmented patches using Natural Data-augmentation in DFU Dataset

Category	Definition	Cases	DFU patches	Augmented patches
Ischaemia	Absent	1249	1431	4935
	Present	210	235	4935
Total images		1459	1666	9870
Bacterial infection	None	628	684	2946
	Present	831	982	2946
Total images		1459	1666	5892

[26]. SLIC superpixels technique performs a localized k-means optimization in the 5-D CIELAB color and image space to cluster pixels as described by equations 1 - 4:

$$S = \sqrt{\frac{N}{k}} \tag{1}$$

$$D_s = d_{lab} + \frac{m}{S} d_{xy} \tag{2}$$

$$d_{lab} = \sqrt{(l_k - l_i)^2 + (a_k - a_i)^2 + (b_k - b_i)^2} \tag{3}$$

$$d_{xy} = \sqrt{(x_k - x_i)^2 + (y_k - y_i)^2} \tag{4}$$

where in eq. 1, S is the approximate size of a superpixel, N is the number of pixels and k is the number of superpixels; in eq. 2, D_s is the sum of the lab distance ($\mathrm{d_{lab}}$)and the xy plane distance ($\mathrm{d_{xy}}$); in eq. 3, l, a and b represent the lab colorspace; and in eq. 4, x and y represent the pixel positions.

In the second step, the mean RGB color value of each superpixel is computed and applied to each superpixel (S) denoted by:

$$S_i = mean(P(R, G, B)), i = 1, \ldots, k \tag{5}$$

where in eq. 5, *P(R,G,B)* is the pixel values of R,G,B channel in each *ith* position of S and k is total number of superpixels in the image.

Finally, with a different number of superpixels and threshold values from each color channel, we extracted regions of two particular colors of interest that are red and black from the DFU patches. For these classification tasks, we used the number of superpixels (k=200) and threshold values (T1: 0.40,0.45,0.50,.055,0.60; T2: 0.15,0.20,0.25,0.30,0.35) to extract the color features from DFU patches of 256×256. The threshold values are used to restrict the intensities of red and black pixels to be utilized as handcrafted features. Hence, we utilised a feature vector of 10 with SPCD algorithm along with texture descriptors (LBP, HOG) and color features (RGB, CIELAB) to train traditional machine learning approaches. The pseudocode for the SPCD algorithm is explained in Algorithm 1. The example of extracting color features using our novel SPCD algorithm is shown in Fig. 6.

For these classification problems, we experimented with a number of classifiers with standard hyper-parameters on these color features. BayesNet,

Algorithm 1 Pseudocode for the Superpixel Color Descriptors Extraction

Over-segmentation of DFU patch with SLIC superpixel is performed;
Mean RGB value of each superpixel is calculated and applied;
Initialize variable S_Red & S_Black to 0
procedure RedAndBlackRegion
  for *each* $Superpixel(S_{\mathrm{i}})$ **do**
    if $S_{\mathrm{i}}(R) > T_1 * (S_{\mathrm{i}}(R) + S_{\mathrm{i}}(G) + S_{\mathrm{i}}(B))$ **then return** S_Red= S_Red + 1
    if $S_{\mathrm{i}}(R) < T_2$ & $S_{\mathrm{i}}(G) < T_2$ & $S_{\mathrm{i}}(B) < T_2$ **then return** S_Black= S_Black + 1
    $RedColorFeature = S\_Red \div n$
    $BlackColorFeature = S\_Black \div n$

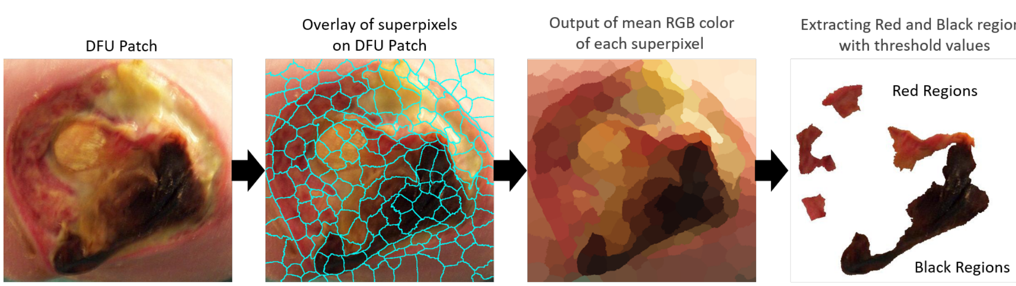


Figure 6: Example of extracting red and black regions from DFU patch with proposed Superpixel Color Descriptor algorithm which was then used to inform identification of ischaemia and infection. The k value of 200 for superpixel algorithm effectively overseg-mented the DFU patches.

Random Forest, and Multilayer Perceptron were selected and achieved the highest accuracy among other machine learning classifiers.

3.3. Deep Learning Approaches

For comparison with the traditional features, deep learning algorithms are used to perform binary classification to classify (1) infection and non-infection; and (2) ischaemia and non-ischaemia classes in DFU patches. For this work, we fine-tune (transfer learning from pre-trained models) the CNN models, i.e. Inception-V3, ResNet50, and InceptionResNetV2 [27, 28, 29]. To train the CNN networks, we froze the weights of the first few layers of the pre-trained networks for common features, such as edges and curves. Subsequently, layers of networks are unfrozen to focus on learning dataset-specific features.

Additionally, we utilized the Ensemble CNN method, which is a very effective CNN approach to obtain very good accuracy on difficult datasets. The Ensemble CNN model combines the bottleneck features from multiple CNN models (Inception-V3, ResNet50, and InceptionResNetV2), and use SVM classifier to produce predictions, as shown in Fig. 7.

4. Results and Discussion

Both infection and ischaemia datasets were split into 70% training, 10% validation and 20% testing sets and we adopted the 5-fold cross-validation technique. We utilized the natural data-augmentation technique for training and validation sets in both traditional machine learning and deep learning approaches. Hence, in this ischaemia dataset, we used approximately 11,564 patches, 1,652 patches, and 3,304 patches in training, validation, and testing sets respectively whereas, in the infection dataset, we used 7,136 patches (training), 1,019 patches (validation), and 2,038 patches (testing) from the 2611 original foot images. As mentioned previously, we used both hand-crafted traditional machine learning (henceforth TML) models and CNN models to perform the classification task and utilized 256×256 RGB images as input for TML and InceptionV3, AlexNet, and ResNet50. For Inception-ResNetV2, we resized the dataset to 299×299. For this experiment, TensorFlow is used for deep learning and Matlab is used for traditional machine learning approaches.

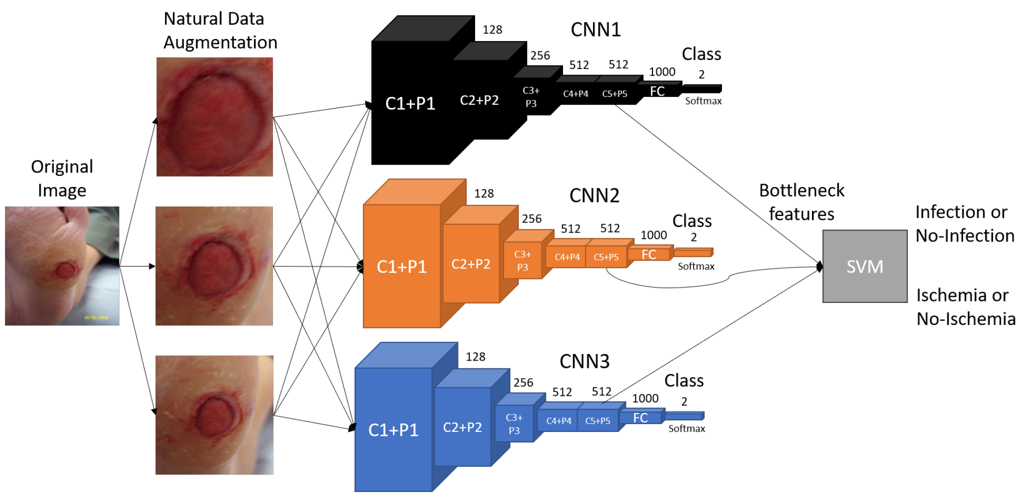


Figure 7: Extracting bottleneck features from CNNs and fed into SVM classifier to perform binary classification of ischaemia and infection, where C1-C5 are convolutional layers, P1-P5 are pooling layers and FC is fully connected layer. **Note: The CNNs in this figure are just representations of general CNNs architecture and do not represent the original CNN architectures of Inception-V3, ResNet50, and InceptionResNetV2.**

Table 2: The performance measures of binary classification of ischaemia by our proposed handcrafted traditional machine learning and CNN approaches.

	Accuracy	*Sensitivity*	*Precision*	*Specificity*	*F-Measure*	*MCC Score*	*AUC Score*
BayesNet	0.785±0.022	0.774±0.034	0.809±0.034	0.800±0.027	0.790±0.020	0.572±0.044	0.783
Random Forest	0.780±0.041	0.739±0.049	0.872±0.029	0.842±0.034	0.799±0.033	0.571±0.078	0.780
Multilayer Perceptron	0.804±0.022	0.817±0.040	0.787±0.046	0.795±0.031	0.800±0.023	0.610±0.045	0.804
InceptionV3 (CNN)	0.841±0.017	0.784±0.045	0.886±0.018	0.898±0.022	0.831±0.021	0.688±0.031	0.840
ResNet50 (CNN)	0.862±0.018	0.797±0.043	0.917±0.015	0.927±0.017	0.852±0.022	0.732±0.032	0.865
InceptionResNetV2 (CNN)	0.853±0.021	0.789±0.054	0.906±0.017	0.917±0.019	0.842±0.027	0.714±0.039	0.851
Ensemble (CNN)	**0.903±0.012**	**0.886±0.035**	**0.918±0.019**	**0.921±0.021**	**0.902±0.014**	**0.807±0.022**	**0.904**

Table 3: The performance measures of binary classification of Infection by our proposed handcrafted traditional machine learning and CNN approaches.

	Accuracy	*Sensitivity*	*Precision*	*Specificity*	*F-Measure*	*MCC Score*	*AUC Score*
BayesNet	0.639±0.036	0.619±0.018	0.653±0.039	0.660±0.015	0.622±0.079	0.290±0.070	0.643
Random Forest	0.605±0.025	0.608±0.025	0.607±0.037	0.601±0.069	0.606±0.012	0.211±0.051	0.601
Multilayer Perceptron	0.621±0.026	0.680±0.023	0.622±0.057	0.570±0.023	0.627±0.074	0.281±0.055	0.619
InceptionV3 (CNN)	0.662±0.014	0.693±0.038	0.653±0.015	0.631±0.034	0.672±0.019	0.325±0.029	0.662
ResNet50 (CNN)	0.673±0.013	0.692±0.051	0.668±0.023	0.654±0.051	0.679±0.019	0.348±0.028	0.673
InceptionResNetV2 (CNN)	0.676±0.015	0.688±0.052	0.672±0.015	0.664±0.039	0.680±0.024	0.352±0.031	0.678
Ensemble (CNN)	**0.727±0.025**	**0.709±0.044**	**0.735±0.036**	**0.744±0.050**	**0.722±0.028**	**0.454±0.052**	**0.731**

In Table 2 and 3, we report *Accuracy*, *Sensitivity*, *Precision*, *Specificity*, *F-Measure*, *Matthew Correlation Coefficient (MCC)* and *Area under the ROC curve (AUC)* as our evaluation metrics.

When comparing the performance of the computerized methods and our proposed techniques, CNNs performed better in the binary classification of ischaemia than infection despite more imbalanced data in the ischaemia dataset, due to more cases of non-ischaemia in the dataset. The average performance of all the models in terms of accuracy in the ischaemia dataset was 83.3% which is notably better than the average accuracy of 65.8% in infection dataset. Similarly, *MCC Score* and *AUC Score* are considered to be viable performance measures to compare the classification results. We obtained an average *MCC Score* and *AUC Score* for ischaemia classification of 67.1% and

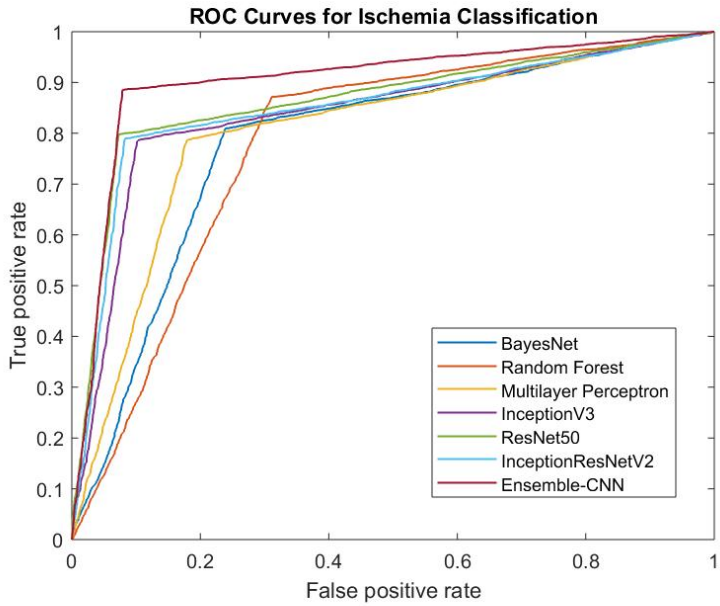


Figure 8: ROC curve for all TML and CNN methods for ischaemia classification.

83.2% respectively, as compared to the infection classification of 32.3% and 65.8% respectively. The ROC curves for all the algorithms, including TML and CNNs for binary classification of ischaemia and infection, are shown in Fig. 8 and 9. When comparing the performances in ischaemia classification of TML and CNNs, CNNs (86.5%) performed better than the TML models (79%). Similarly, in infection classification, the accuracy of CNNs (68.4%) performed better than TML (62.1%) with a margin of 6.3%. Notably, Ensemble CNN method achieved the highest score in all performance measures in both ischaemia and infection classification.

Sensitivity and *Specificity* are considered important performance measures in medical imaging. The ensemble method yielded high *Sensitivity* for the ischaemia dataset with a margin of 6.9% from the second best performing algorithm multilayer perceptron. Interestingly, a multilayer perceptron performed worst in the *Specificity* with a score of 79.5%. For *Specificity* in the ischaemia dataset, the ensemble method again obtained the highest score of 92.9% which is marginally better than ResNet50 (92.7%).

In infection classification, both TML and CNN methods received moderate scores in the performance measures. Again, CNN methods performed

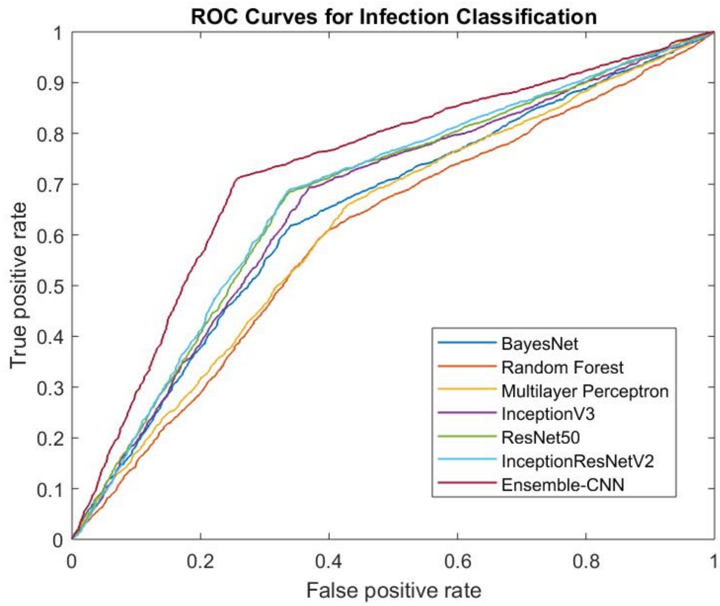
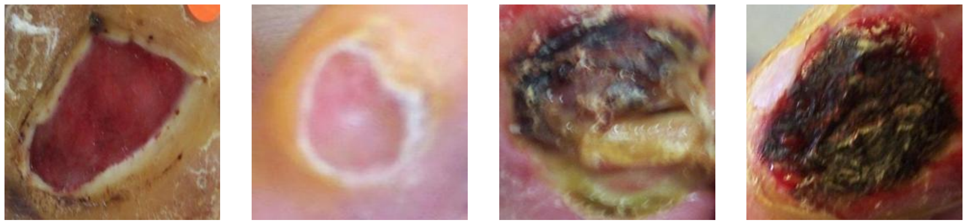


Figure 9: ROC curve for all TML and CNN methods for Infection classification.

better than TML methods achieving the highest score in all performance measures. The Ensemble CNN method performed better than other CNN classifiers especially for *Specificity* with a score of 74.4% in infection classification with a notable margin of 8% than the second-best performing algorithm InceptionResNetV2(66.4%). For *Sensitivity*, all the CNNs performed marginally well with Ensemble method achieving the highest score of 70.9%. When comparing the performance of TML methods, Multilayer Perceptron (68.0%) performed well in *Sensitivity*, whereas BayesNet (66%) better in *Specificity*.

4.1. Experimental Analysis and Discussion

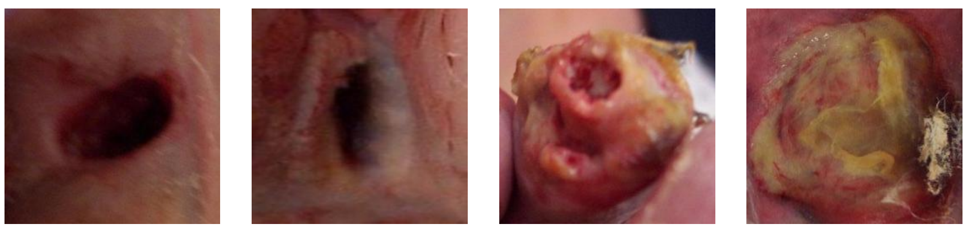
Assessment of DFU with computerized methods is very important for supporting global healthcare systems through improving triage and monitoring procedures and reducing hospital time for patients and clinicians. This preliminary experiment is focused on automatically identifying the important conditions of ischaemia and infection of DFU. The main aim of this experiment was to identify ischaemia and infection from images of the feet using machine learning. We have illustrated examples of correctly and incorrectly classified cases in both binary classifications of ischaemia (Fig. 10 and 11)



(a) (b) (c) (d)

Accurate non-ischaemia cases Accurate ischaemia cases

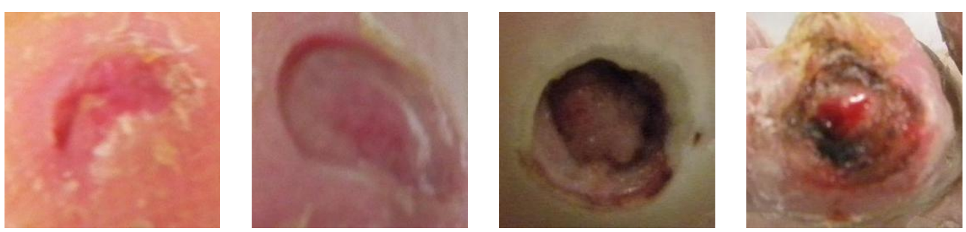
Figure 10: Examples of correctly classified cases by Ensemble-CNN on ischaemia dataset. (a) and (b) represent non-ischaemia cases. (c) and (d) represent ischaemia cases.



(a) (b) (c) (d)

Misclassified non-ischaemia cases Misclassified ischaemia cases

Figure 11: Examples of misclassified cases by Ensemble-CNN on ischaemia dataset. (a) and (b) represents non-ischaemia cases. (c) and (d) represents ischaemia cases.

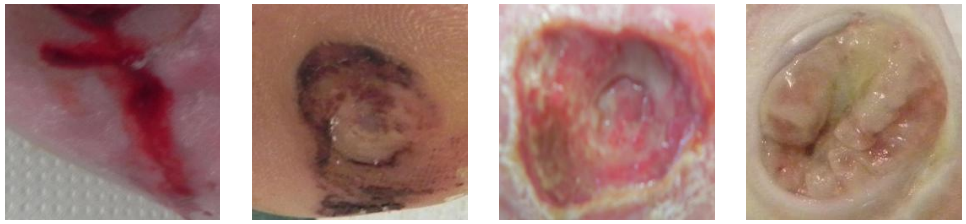


(a) (b) (c) (d)

Accurate non-infection cases Accurate infection cases

Figure 12: Examples of correctly classified cases by Ensemble-CNN on Infection dataset. (a) and (b) represents non-infection cases. (c) and (d) represents infection cases.

and infection (Fig. 12 and 13). As for the misclassified cases, there are huge intra-class dissimilarities and inter-class similarities between (1) infection and non-infection; (2) ischaemia and non-ischaemia cases in the DFU that make classifiers difficult to predict the correct class. Additionally, there are other



(a) (b) (c) (d)

Misclassified non-infection Misclassified infection cases

Figure 13: Examples of misclassified cases by Ensemble-CNN on Infection dataset. (a) and (b) represents non-infection cases. (c) and (d) represents infection cases.

influencing factors in the classification of these conditions such as lighting conditions, marks and skin tone. In misclassified cases of non-ischaemia as shown in Fig. 11, the cases (a) and (b) are hindered by the lighting condition (shadow) respectively, whereas in the (c) and (d) misclassified ischaemia cases, the ischaemia features may be too subtle to be recognised from the images by the algorithm. Alternatively it is likely we needed a more sensitive objective measure of the ground truth from vascular assessments. We found that shadows are particularly problematic because machine learning algorithms can be deceived by shadows especially in determining the important conditions such as ischaemia. In Fig. 13, misclassified cases of non-infection, the presence of blood in the case (a), whilst case (b) belongs to one of the rare cases with the presence of ischaemia and non-infection. In misclassified infection cases, the visual indicators of infection were likely too subtle, or we needed more sensitive objective ground truth provided through blood analysis.

In this work, we used the proposed natural data-augmentation with the help of DFU localisation to create DFU patches from full-size foot images. These patches are useful to focus more on finding the visual indicators for important factors of DFU such as infection and ischaemia. Then, we investigated the use of both TML and CNNs to determine these conditions as binary classification. In this experiment, we received very good performance in terms of correctly classifying ischaemia despite the imbalanced cases in the DFU dataset. However, in the case of infection, the classifiers did not perform as well, since the condition of infection is hard to recognise from the foot images even by experienced medical experts specialized in DFU and therefore likely requires ground truth determined using objective blood tests to identify bacterial infection.

Current research focuses on ischaemia and infection recognition in medical classification systems, which requiring the guidance of medical experts specialized in DFU. To develop a computer-aided tool for medical experts in remote foot analysis, i.e. a remote DFU diagnosis system, the following are challenges need to be addressed:

1. Recognition of the ischaemia and infection with machine learning algorithms as an important proof-of-concept study for foot pathologies classification. Further analysis of each pathology on foot images is required according to the medical classification systems, such as the University of Texas Classification of DFU [8] and SINBAD Classification System [9]. This requires close collaboration with medical experts specialized in DFU.
2. Deep learning algorithms need substantial datasets to obtain very good accuracy, especially for medical imaging. This experiment included an imbalanced DFU dataset (1459 foot images) for both ischaemia and infection conditions. In the future, if these algorithms were to train with a larger number of a more balanced dataset, it can possibly improve the recognition of ischaemia and infection.
3. A study of the performance of algorithms on different types of capturing devices is an important aspect of future work. This experiment evaluates the performance of machine learning algorithms on the DFU dataset collected with different cameras (heterogeneous sources of data). This leads to more variability of image characteristics. Since the algorithms have to deal with more heterogeneous patterns and characteristics that are not intrinsic to the pathology itself. In this experiment, we know that three types of devices were used, we do not have the information on the association of images and the type of devices.
4. The current ground truth is based on visual inspection by experts only and not supported by the medical notes or clinical tests (vascular assessment for ischaemia and blood tests to identify the presence of any bacterial infection). Furthermore, DFU images were debrided before these images were captured. Hence, the debridement of DFU removes important visual indicators of infection such as colored exudate. Therefore, the sensitivity and specificity of these algorithms could be further improved in the future, by feeding in ground truth from clinical tests such as vascular assessments (ischaemia) and blood tests (to identify

the presence of any bacterial infection).

5. Current clinical practice obtains the foot photo using different camera models, poses and illumination. It is a great challenge for a computer algorithm to predict the depth and the size of the wound based on non-standardized images. Standardized dataset, such as the data collection method proposed by Yap et al. [30] will help to increase the accuracy of the DFU diagnosis system.
6. Dataset annotation is a laborious process, particularly for medical experts to label the foot pathologies into 16 classes according to the University of Texas classification system. To reduce the burden upon medical experts in the delineation and annotation of the dataset, there is an urgent need to focus on developing unsupervised or self-supervised machine learning techniques.
7. Collecting the time-line dataset is crucial for early detection of key pathologies. This will enable monitoring of foot health and changes longitudinally, where medical experts and computer algorithms can learn the early signs of DFU. In the longer-term, the DFU diagnosis system will be able to predict the healing process of ulcers and prevent DFU before it happens.
8. A smart-phone app could be developed for remote triage and monitoring of DFU. To scale-up the DFU diagnosis system, the application should run on multiple devices, irrespective of the platform and/or the operating system.

5. Conclusion

In this work, we trained various classifiers based on traditional machine learning algorithms and CNNs to discriminate the conditions of: (1) ischaemia and non-ischaemia; and (2) infection and non-infection related to a given DFU. We found high-performance measures in the binary classification of ischaemia, compared to moderate performance by classifiers in the classification of infection. It is vital to understand the features of both conditions in relation to the DFU (ischaemia and infection) from a computer vision perspective. Determining these conditions especially infection from the non-standard foot images is very challenging due to: (1) high visual intra-class dissimilarities and inter-class similarities between classes; (2) the visual indicators of infection and ischaemia potentially being too subtle in DFU; (3) objective medical tests for vascular supply and bacterial infection are needed to provide more objective ground truth and further improve the classification of these conditions; and (4) other factors such as lighting conditions, marks and skin tone are important to incorporate into the prediction.

With a more balanced dataset and improved data capturing of DFU, the performance of these methods could be improved in the future. Further optimization in hyper-parameters of both deep learning and traditional machine learning methods could improve the performance of algorithms on this dataset. Ground truths enhanced by clinical tests for the ischaemia and infection may provide further insight and further improvement of algorithms even where there is no apparent visual indicator by eye. In the case of infection even after debridement, ground truth informed by blood tests for infection may yield improvements to sensitivity and specificity even in the absence of overtly obvious visual indicators. This work has the potential for technology that may transform the recognition and treatment of diabetic foot ulcers and lead to a paradigm shift in the clinical care of the diabetic foot.

Acknowledgements

The authors express their gratitude to Lancashire Teaching Hospitals and the clinical experts for their extensive support and contribution in carrying out this research. We would like to thank Kim's English Corner (https://kimsenglishcorner.com) for proofreading.

Reference

[1] J. D. Santilli, S. M. Santilli, Chronic critical limb ischemia: diagnosis, treatment and prognosis., American family physician 59 (7) (1999) 1899–1908.

[2] M. Albers, A. C. Fratezi, N. De Luccia, Assessment of quality of life of patients with severe ischemia as a result of infrainguinal arterial occlusive disease, Journal of vascular surgery 16 (1) (1992) 54–59.

[3] L. Prompers, M. Huijberts, J. Apelqvist, E. Jude, A. Piaggesi, K. Bakker, et al., High prevalence of ischaemia, infection and serious comorbidity in patients with diabetic foot disease in europe. baseline results from the eurodiale study, Diabetologia 50 (1) (2007) 18–25.

[4] B. A. Lipsky, A. R. Berendt, P. B. Cornia, J. C. Pile, E. J. Peters, D. G. Armstrong, et al., 2012 infectious diseases society of america clinical practice guideline for the diagnosis and treatment of diabetic foot infections, Clinical infectious diseases 54 (12) (2012) e132–e173.

[5] L. A. Lavery, D. G. Armstrong, R. P. Wunderlich, J. Tredwell, A. J. Boulton, Diabetic foot syndrome: evaluating the prevalence and incidence of foot pathology in mexican americans and non-hispanic whites from a diabetes disease management cohort, Diabetes care 26 (5) (2003) 1435–1438.

[6] G. H. Skrepnek, J. L. Mills, L. A. Lavery, D. G. Armstrong, Health care service and outcomes among an estimated 6.7 million ambulatory care diabetic foot cases in the us, Diabetes Care 40 (7) (2017) 936–942.

[7] F. W. Wagner, The diabetic foot, Orthopedics 10 (1) (1987) 163–172.

[8] L. A. Lavery, D. G. Armstrong, L. B. Harkless, Classification of diabetic foot wounds, The Journal of Foot and Ankle Surgery 35 (6) (1996) 528–531.

[9] P. Ince, Z. G. Abbas, J. K. Lutale, A. Basit, S. M. Ali, F. Chohan, S. Morbach, et al., Use of the sinbad classification system and score in comparing outcome of foot ulcer management on three continents, Diabetes care 31 (5) (2008) 964–967.

[10] J. J. van Netten, D. Clark, P. A. Lazzarini, M. Janda, L. F. Reed, The validity and reliability of remote diabetic foot ulcer assessment using mobile phone images, Scientific Reports 7 (1) (2017) 9480.

[11] V. Gulshan, L. Peng, M. Coram, M. C. Stumpe, D. Wu, A. Narayanaswamy, et al., Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs, Jama 316 (22) (2016) 2402–2410.

[12] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, et al., Dermatologist-level classification of skin cancer with deep neural networks, Nature 542 (7639) (2017) 115–118.

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

[14] F. Veredas, H. Mesa, L. Morente, Binary tissue classification on wound images with neural networks and bayesian classifiers, IEEE transactions on medical imaging 29 (2) (2009) 410–427.

[15] H. Wannous, Y. Lucas, S. Treuillet, Enhanced assessment of the wound-healing process by accurate multiview tissue classification, IEEE transactions on Medical Imaging 30 (2) (2010) 315–326.

[16] L. Wang, P. Pedersen, E. Agu, D. Strong, B. Tulu, Area determination of diabetic foot ulcer images using a cascaded two-stage svm based classification, IEEE Transactions on Biomedical Engineering (2016).

[17] M. H. Yap, M. Goyal, F. M. Osman, R. Martí, E. Denton, A. Juette, et al., Breast ultrasound lesions recognition: end-to-end deep learning approaches, Journal of Medical Imaging 6 (1) (2018) 011007.

[18] E. Ahmad, M. Goyal, J. S. McPhee, H. Degens, M. H. Yap, Semantic segmentation of human thigh quadriceps muscle in magnetic resonance images, arXiv preprint arXiv:1801.00415 (2018).

[19] M. Goyal, N. D. Reeves, A. K. Davison, S. Rajbhandari, J. Spragg, M. H. Yap, Dfunet: convolutional neural networks for diabetic foot ulcer classification, IEEE Transactions on Emerging Topics in Computational Intelligence (2018) 1–12doi:10.1109/TETCI.2018.2866254.

[20] M. Goyal, M. H. Yap, N. D. Reeves, S. Rajbhandari, J. Spragg, Fully convolutional networks for diabetic foot ulcer segmentation, in: 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC), 2017, pp. 618–623. doi:10.1109/SMC.2017.8122675.

[21] M. Goyal, N. D. Reeves, S. Rajbhandari, M. H. Yap, Robust methods for real-time diabetic foot ulcer detection and localization on mobile devices, IEEE Journal of Biomedical and Health Informatics 23 (4) (2019) 1730–1741. doi:10.1109/JBHI.2018.2868656.

[22] C. Wang, X. Yan, M. Smith, K. Kochhar, M. Rubin, S. M. Warren, et al., A unified framework for automatic wound segmentation and analysis

with deep convolutional neural networks, in: Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE, IEEE, 2015, pp. 2415–2418.

[23] J. L. Mills Sr, M. S. Conte, D. G. Armstrong, F. B. Pomposelli, A. Schanzer, A. N. Sidawy, et al., The society for vascular surgery lower extremity threatened limb classification system: risk stratification based on wound, ischemia, and foot infection (wifi), Journal of vascular surgery 59 (1) (2014) 220–234.

[24] M. Goyal, M. H. Yap, Region of interest detection in dermoscopic images for natural data-augmentation, arXiv preprint arXiv:1807.10711 (2018).

[25] J. Huang, V. Rathod, C. Sun, M. Zhu, A. Korattikara, A. Fathi, et al., Speed/accuracy trade-offs for modern convolutional object detectors, arXiv preprint arXiv:1611.10012 (2016).

[26] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, S. Süsstrunk, Slic superpixels, Tech. rep. (2010).

[27] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, Rethinking the inception architecture for computer vision, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 2818–2826.

[28] C. Szegedy, S. Ioffe, V. Vanhoucke, Inception-v4, inception-resnet and the impact of residual connections on learning, CoRR abs/1602.07261 (2016). `arXiv:1602.07261`.
URL `http://arxiv.org/abs/1602.07261`

[29] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.

[30] M. H. Yap, K. E. Chatwin, C.-C. Ng, C. A. Abbott, F. L. Bowling, S. Rajbhandari, et al., A new mobile application for standardizing diabetic foot images, Journal of diabetes science and technology 12 (1) (2018) 169–173.