DFU_QUTNet: Diabetic Foot Ulcer Classification Using Novel Deep Convolutional Neural Network

Laith Alzubaidi[1], Mohammed A. Fadhel[2], Sameer R. Oleiwi[3], Omran Al-Shamma[4], Jinglan Zhang[5]

Abstract

Diabetic Foot Ulcer (DFU) is the main complication of Diabetes, which, if not properly treated, may lead to amputation. One of the approaches of DFU treatment depends on the attentiveness of clinicians and patients. This treatment approach has drawbacks such as the high cost of the diagnosis as well as the length of treatment. Although this approach gives powerful results, the need for a remote, cost-effective, and convenient DFU diagnosis approach is urgent. In this paper, we introduce a new dataset of 754-foot images which contain healthy skin and skin with a diabetic ulcer from different patients. A novel Deep Convolutional Neural Network, DFU_QUTNet, is also proposed for the automatic classification of normal skin (healthy skin) class versus abnormal skin (DFU) class. Stacking more layers to a traditional Convolutional Neural Network to make it very deep does not lead to better performance, instead leading to worse performance due to the gradient. Therefore, our proposed DFU_QUTNet network is designed based on the idea of increasing the width of the network while keeping the depth compared to the state-of-the-art networks. Our network has been proven to be very beneficial for gradient propagation, as the error can be back-propagated through multiple paths. It also helps to combine different levels of features at each step of the network. Features extracted by DFU_QUTNet network are used to train Support Vector Machine (SVM) and K-Nearest Neighbors (KNN) classifiers. For the sake of comparison, we have fine-tuned then re-trained and tested three pre-trained deep learning networks (GoogleNet, VGG16, and AlexNet) for the same task. The proposed DFU_QUTNet network outperformed the state-of-the-art CNN networks by achieving the F1-score of 94.5%.

Keywords Diabetic foot ulcers. Convolutional Neural Networks. Deep learning. Classification. Transfer learning.

1 Introduction

Diabetes can lead to serious complications like lower limb amputation, blindness, kidney failure, and cardiovascular diseases. These complications are commonly preceded by Diabetic Foot Ulcers (DFUs) [1]. DFUs, generally known as sores, are open wounds that can develop on the feet of people with diabetes. .Based on global reporting, there were only 108 million people suffering from diabetes in 1980, while this number increased to more than 422 million people in 2014. For instance, looking at the adults over 18 years old, the overall prevalence increased from 4.7% to 8.5% during the period 1980-2014 [2]. Furthermore, the number of people living with diabetes is expected to increase up to 600 million by the end of 2035. It is also important to mention that about 80% of these people are from developing countries due to the lack of health care facilities and poor awareness [3]. Moreover, around 15% to 25% of diabetic patients will develop DFU at an advanced stage of the disease, which may lead to lower limb amputation if proper care is unavailable [4]. Due to failure to recognize and improper DFU treatment, over one million diabetic people with a 'high-risk foot' will lose part of their leg each year [5]. Hygienic personal care, continuous medication, and periodic checkup with doctors are required for people with diabetes to prevent additional consequences. Therefore, diabetes will cause a huge financial burden for people with this disease and their families, mainly in developing countries, since the treatment cost is about 5.7 times the annual income [6].

✉ Laith Alzubaidi
laith.alzubaidi@hdr.qut.edu.au

Faculty of Science & Engineering, Queensland University of Technology, Brisbane, Australia

Nowadays, the assessment of the early diagnosis of DFU in clinical systems involves several significant tasks, which represent continuous tracking of numerous lengthy actions enclosed in handling and managing of DFU for every individual case. These actions include; evaluation of patient's medical history, examination of the DFU carefully by a diabetic foot specialist, and extra tests for assessment in developing the treatment plan, such as x-ray, MRI, and CT scans. In general, DFU patients have an inflated leg problem which may be itchy and aching depending on each case. Additionally, the DFU usually has uncertain external boundaries and asymmetrical structure. Several factors affect the DFU surrounding skin and its visual appearance, which include; scaly skin, bleeding, slough, important tissues (e.g. Granulation), blisters, callus formation, and redness. Therefore, the exact evaluation of these visual symptoms like texture features and color descriptors is the base of generating powerful computer vision algorithms for ulcer assessment [7].

However, the task of classifying DFU has many challenges. Firstly, the collection and professional labeling of DFU images require large time. Secondly, the similarity between DFU and healthy skin has high inter-class and intra-class variations depending on the patient's ethnicity, lighting conditions, and the DFU classification [8]. Lastly, it is hard to differentiate between healthy skin and DFU in some cases, for instance, when there are small sizes of DFU, skin wrinkles, and images with a toe.

Our contributions reveal many interesting observations are summarized below:

(1) We have introduced a new dataset of total 754-foot images which consist of healthy skin and skin with a diabetic ulcer from different patients. The dataset has been manually labeled by a DFU expert.

(2) We have designed and implemented a novel and robust Deep Convolutional Neural Network, DFU_QUTNet, from scratch based on the idea of increasing in the network's width along with the global average pooling. Our network's structure has proven to be helpful since it has different levels of features at each step of the network. It has also proved to be very beneficial for gradient propagation, as the error can be back-propagated through multiple paths.

(3) Our model used to classify feet skin images into two classes: normal and abnormal (DFU). We improved the performance of DFU classification in term of the F1-score. Our model achieved a F1-score of 94.5% in classifying unseen images of our dataset, which outperforms the state-of-the-art CNN networks.

(4) We have trained the multi-classifiers of K-Nearest Neighbor and Support Vector Machine (SVM) classifiers with the features extracted by our proposed DFU_QUTNet network.

(5) We have fine-tuned and re-trained three pre-trained deep learning networks (GoogleNet, VGG16, and AlexNet) for DFU classification task to be compared with our model. DFU_QUTNet outperforms these three state-of-the-art CNN networks.

(6) We have re-implemented DFUNet which was designed by Goyal, et al. [59] and trained it with our dataset. As far as we know, this is the only paper that employed CNNs to classify normal skin against abnormal skin (DFU). Our proposed DFU_QUTNet network outperforms DFUNet network.

(7) As described in Section2, we have reviewed some of the state-of-the-art methods for deep learning and the traditional methods used to classify DFU.

The rest of the paper is organized as follows: Section 2 reviews the related work. Section 3 explains the proposed methodology. Section 4 presents the results. Lastly, section5 draws conclusions.

2 Related Work

Numerous medical-telecommunication systems, including diabetes monitoring systems, have been developed in recent years [9][10]. The objectives of these systems are: to offer automated solutions for the lack in the specialists, to enhance the access of medical facilities, to enhance the available health care systems, and to reduce the cost of medical facilities [11]. However, a small number of expert systems related to the evaluation of diabetic foot pathologies which can be classified as automated and non-automated systems have been developed.

2.1 DFU telemedicine systems

New technology devices, such as laptops smart-phones and Internet media, have been developed which offer the possibility of remote communications. With the sophisticated mobile Internet (e.g. 5G), these devices have the ability to capture and send very high-quality pictures, audio and video [12][13]. Medical-telecommunication systems of the non-automated type, which are frequently set-up in a distant location, offer facilities for patients' evaluation such as optical scanner [14], digital photography [15], 3D wound imaging [16], and video conferencing [17]. However, the availability of professional medical experts is still required to complete the patient's evaluation on the other side. Although these systems give powerful results, the need for automated DFU pathology detection is urgent.

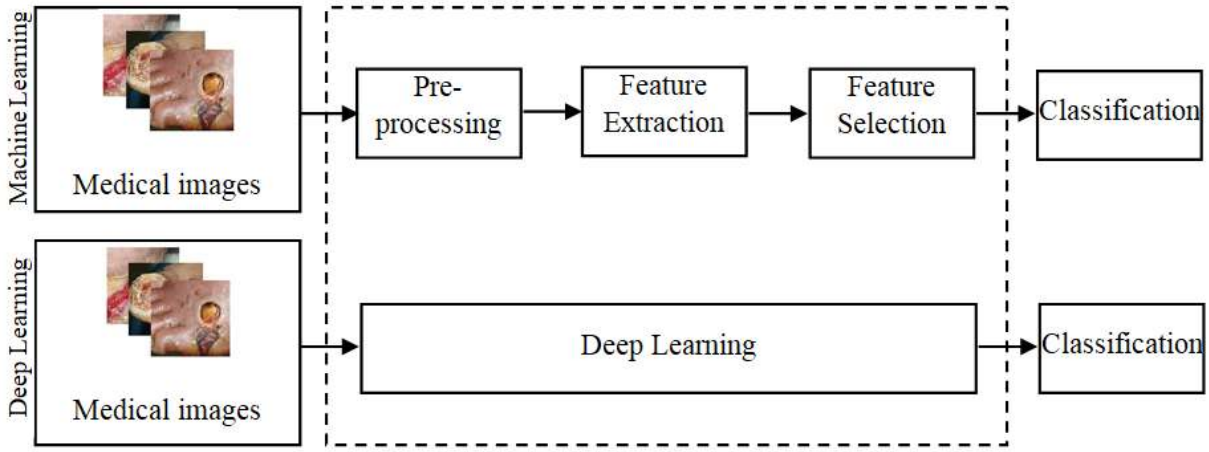

Fig.1 The difference between deep learning and machine learning

The employment of automated DFU detection systems is still in the earlier development stages and has extremely limited applications. Notably, in 2015, C. Liu et al. [4] and O. Jegede et al. [18] presented a smart system for DFU detection based on 3D surface reconstruction, infrared thermal images, and spectral imaging. However, the implementation of the system requires much expert training and expensive devices in order to work properly. Another system by Wang, et al. [19] is based on capturing an image using a capture box and determines the DFU area by utilizing the two-stage classification of Support Vector Machine.The first step of this system is segmentation which is executed by a super-pixel methodology. The extraction of features represents the second step which is used to implement two staged classification. In spite of the fact that the system presented powerful results, it has some drawbacks including its invalidity on a big dataset and the impracticality of the captured image box for data collection due to the need for patient's foot to be in contact with box surface which in turn does not allow by the health care setting due to worries concerning infection control. Other important research by Manu, et al. [20] has implemented the DFU segmentation on whole foot images.

Furthermore, computerized methodologies based on image processing or manually engineered-feature methods were applied for related skin cut segmentation (such as a wound) and tissue classification. In general, the conventional machine learning technique was applied through extracting several features like color descriptors and texture descriptors on small defined patches of injury images. Then, the classification task takes place to categorize skin patches into normal and abnormal by utilizing machine learning algorithms [21] [22] [23] [24]. The hand_crafting features methods are influenced by lighting conditions as in several computer vision systems, whereas skin color relying upon the patient ethnicity group. Generally, all the skin cuts connected to an ulcer and wound are labeled as a wound. From the medical viewpoint, the ulcer is due to an internal problem, while the wound is due to an external problem. There are more differences between ulcers and wounds in the skin cut such as the appearance of skin in case of wound or ulcer, pathology (disease processes), physiology (the method the body responds), and aetiology (the cause) [25,26]. Therefore, DFU takes into consideration the work of finding the difference between normal and abnormal skin at the same place of appearance.

2.2 Deep Learning

In recent years, deep learning has powered and solved many pattern recognition and computer vision tasks such as image classification [27], age classification [28], sketch recognition [29] and nuclei detection [30]. It is considered the main factor for such improvements in medical image analysis [31], [32]. It has been successfully employed in many biomedical image analysis challenges, such as malaria detection [33], skin biopsy histopathological image annotation [34] and retinal ailment detection [35]. Classical image analysis and Classical machine learning include a series of steps, which are preprocessing, feature extraction, and careful selection of features, learning, and classification. The performance of these methods is strongly dependent on the chosen features and the accuracy of the preceding steps while deep learning enables automated learning of the feature sets for particular problems instead of hand-crafting features [36]. Fig.1 shows the difference between deep learning and machine learning.

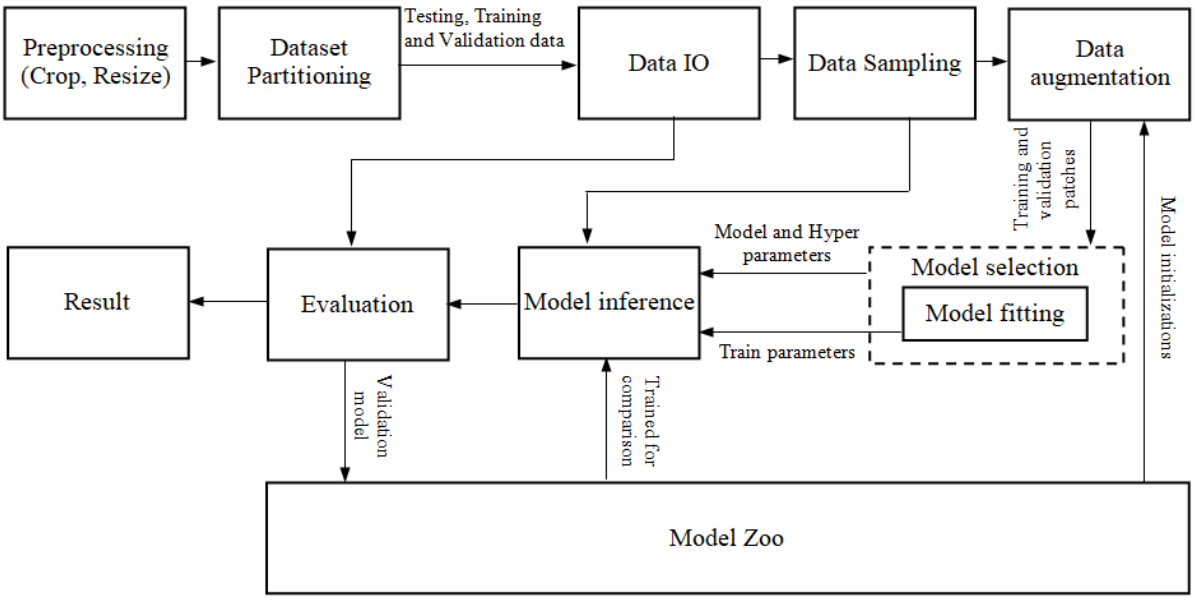

Fig.2 Workflow of deep learning tasks

Deep learning has been used in different tasks such as detection, counting, and classification. In spite of the fact that each of these tasks has its own specificities, there is essential overlap in pipeline implementation of these applications as shown in fig2. Classification is one of the most significant tasks in the medical field [31]. Several approaches of this task have been proposed in recent years, but still, remain challenging problems. There are numerous factors affect the performance of this task's algorithms such as noise, overlapping, complex shapes, as well as low-resolution of images, and there exists no single method that successfully handles all these factors. However, deep learning has overcame most of these challenges and has shown great improvements in the field of medical image classification as well as in different computer vision areas, especially towards the hard and important issues such as understanding images of different scopes such as spectral, marine organism classification, sickle cells anemia classification and hatching eggs classification [37], [38],[39], [40].

Machine learning algorithms and conventional computer vision have many limitations including the need of countless manual tuning for each image, the difficulties in data representation with multi-level abstraction, and the processing ability of a large image data. The solution is a newly developed machine-learning algorithm, termed as Deep Convolutional Neural Networks [36]. This solution has the ability to obtain multi-level representation methods using non-linear simple modules first. Then, the simple feature representations are transformed into extra advanced abstract representations by these modules. In other words, the deep convolutional networks employ images as input for learning features, like locations from the pixel array values and edges at certain directions. Then, a combination of these edges has performed at a higher level in learning additional significant abstract features like desirable object components. The final stage is the connection of these components together to form the final objects [41].

Convolutional Neural Networks (CNNs) are one of the reasons that make deep learning popular today [42]. They have shown record-breaking performance in different image classification tasks [27]. CNN was first proposed for image classification tasks [43]. CNN is a feed-forward neural network consisting of one or more convolutional layers followed by pooling layers as shown in Fig3. Additionally, the high performance and classification accuracy of CNN has made it popular for almost all classification tasks specifically for image classification. Convolutional Neural Networks (CNNs) play the core role in medical image classification [31].

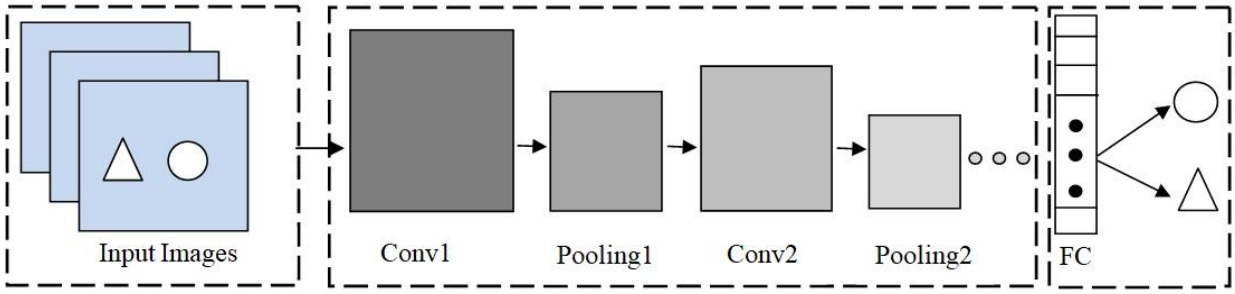

Fig.3 basic structure of CNN. Conv refers to Convolution layer, Pooling refers to Pooling layer, FC refers to a fully connected layer

One of these CNNs is AlexNet. It was the first significant development in 2012 [43]. These significant developments have included several improvements. Firstly, it introduced better non-linearity in the network with ReLU activation. ReLU helps to avoid demising gradient problems. Secondly, it presents the dropout concept identical to regularization. In the dropout concept, the neurons are arbitrarily activated and deactivated in all layers for avoiding overfitting problem. Accordingly, the data is compelled to come across different routes. Hence, the network is capable of better generalization. Thirdly, introducing data augmentation indicates that the images are displayed with an arbitrary crop, rotation, and translation when supplied to the network. Consequently, it enforces the network since it has extra awareness of the image attributes, instead of the image itself. Last, but not least, for improving the classification accuracy, extra convolutional layers are stacked in advance the pooling layers. Afterward, the VGGNet network was introduced [44]. It inserted additional layers of convolutions and pooling for boosting the accuracy [44]. After VGGNet, GoogleNet was introduced [41], where convolution layers with different filter sizes are processed on the same input and then concatenated together. This type of structure assists in having multi-level feature extraction at each step. The ResNet network was introduced after the GoogleNet network. ResNet presented the concept of residual connection which is every two layers; there is an identity mapping via an element-wise addition [45]. An extra complicated structure was developed by the DenseNet network by connecting the entire blocks of layers to each other [46]. These networks have been fine-tuned to classify different tasks of medical image classification [59], [60], [61]. Stacking extra layers to traditional CNN will not always improve the performance, but it may be the inverse because of the gradient. Therefore, we present a convolutional neural network model based on the Directed acyclic graph (DAG) concept where we increased the width of the network. Our proposed model will help to have different levels of features at each step of the network and will be very beneficial for gradient propagation.

3 Methodology

This section consists of five parts: (i) our dataset with examples of diabetic foot ulcer for different patients (ii) labeling process of or dataset as normal and abnormal patched skin (iii) pre-processing of Training patches (data augmentation) (iv) fine-tuned CNNs architectures of pre-trained models (ALEXNET [43], GoogleNet [41], and VGG16 [44]) (v) our proposed model (DFU_QUTNet) which is our own CNN architecture, DFU_QUTNet, to enhance the DFU classification performance.

3.1 Our Dataset

We first collected a standardized dataset of color images of diabetic foot ulceration from different patients. Our comprehensive dataset consists of 754 images of patient's feet with DFU and healthy skin from the Nasiriyah Hospital's diabetic center which is located in the south of Iraq, with ethical approval and written consent from all relevant persons and patients. These images are captured by Samsung galaxy note 8 and iPad with different brightness and angles. All collected images have been de-identified and will be managed following related policies. The collected images were then pre-processed to create patches with standard sizes in order to be used to train and test the proposed model and pre-trained deep-learning models (GoogleNet, VGG16, and AlexNet) for DFU classification.

3.2 Expert Labeling of Images

We first cropped the Region of Interest (ROI) with a size of 224×224. This region is a significant region around the ulcer which includes important tissues of both skin classes (normal and abnormal), then the specialist labeled the cropped patches. The ground-truth labels are marked by a medical specialist in two forms of normal and abnormal skin patches. We collected a total of 1609 skin patches with 542 normal and 1067 abnormal (DFU). Finally, we divided the data-set into 80% of patches for training,

and 20% of patches for a testing set of images. Fig 4 shows samples of normal and abnormal patches before the cropping process. Fig 5 shows some samples from our dataset.

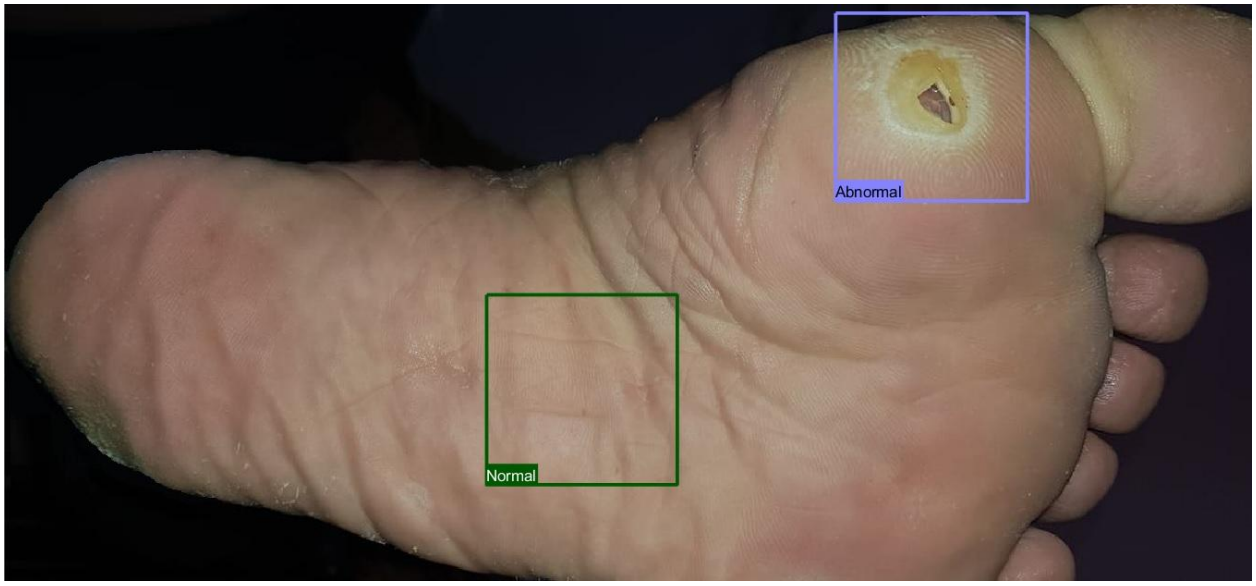

Fig.4 Samples of Normal and Abnormal patches

Fig.5 Samples of patches with labels in our dataset

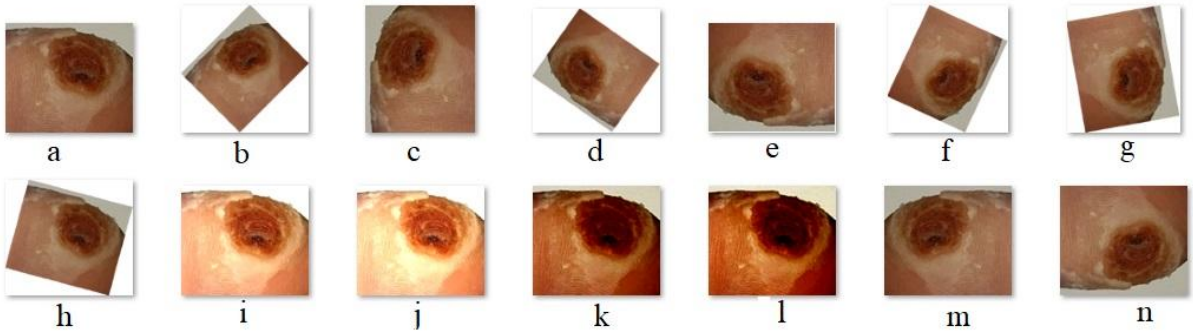

Fig.6 Data augmentation samples a) original image b) rotation of 45 c) rotation of 90 d) rotation of 145 e) rotation of 180 f) rotation of 245 g) rotation of 280 h) rotation of 345 i) brightness of 70 j) brightness of 90 k) contrast of 70 l) contrast of 90 m) horizontal flip n) vertical flip.

3.3 Augmentation of Training patches

CNN requires a large volume of the labeled training set to perform well. If the training set is small, the parameters of CNN will not be well tuned and results in considerable overfitting. The data augmentation process generally improves the performance of different tasks with deep learning [47][48]. Furthermore, collecting a large amount of medical data is costly and difficult. Thus, we applied data augmentation techniques to enhance the performance of the deep learning models and avoiding the overfitting problem. We utilized a set of different image processing methods like flipping, rotation, contrast improvement, using various color models, and random scaling to complete the data augmentation. The rotation is achieved by rotating the image by an angle of 45, 90,145,180,245,280,345. Then, two types of flipping (horizontal flip, vertical flip) with contrast and brightness of 70 and 90 degrees were applied to the original patches. The size of the patches is 224×224. With data augmentation techniques, these patches are duplicated 13 times for training. Fig 6 illustrates all types of image processing methods.

3.4 Pre-trained CNN Architectures

The state-of-the-art CNNs networks have been trained and tested on enormous image datasets such as ImageNet dataset which is approximately 1.28 million natural images of various domains for several classification tasks [49]. Pre-trained CNNs can be fine-tuned on medical image data sets, enabling the large networks for learning specific features of the interesting task. A large number of studies have shown that transfer learning of pre-trained models is effective and efficient for medical image classification [50][51][52]. Using the transfer learning approach of pre-trained models enhances performance. We have used three CNNs models which are (GoogleNet[41], AlexNet[43], and VGG16[44]). These CNNs models have achieved high accuracy in different domains and we fine-tuned them for classification of abnormal (DFU) and normal (healthy skin) classes. Fig 7 explains the steps of our process of transfer learning:

(1) VGGNet 16: This network was proposed by Simonyan, et al, [44] and was one of the winners of ILSVRC 2014. VGG-16 has 13 convolutional layers, 5 layers of max-pooling, and 3 fully connected layers [44]. The inputs image size is 224x224x3 with 3×3 filters size for convolutional layers. VGGNet was used for the classification of different 1000 classes.

(2) AlexNet: Alex Krizhevsky, et al, [43], introduced the AlexNet and won the ILSVRC 2012. It is comprised of five convolutional layers with max-pooling layers which follows some of them plus three full-connected layers with an ending of softmax layer. The first convolutional layer has 96 kernels of 11x11x3 pixel pace size for filtering the input image of size 227x227x3 while the second convolutional layer has 256 kernels of size 5x5x48 for filtering the first convolutional layer output. The pooled/normalized outputs of the second layer are connected to the third convolutional layer, which consists of 384 kernels of size 3x3x256. The next convolutional layer (fourth) contains 384 kernels of size 3x3x192 while the last convolutional layer (the fifth) comprises 256 kernels of size 3x3x192. Note that the last three convolutional layers (third, fourth, and fifth) do not have any normalization or pooling layers connected between each other. Finally, there are 4096 neurons in the fully-connected layers. AlexNet was used to classify 1000 classes of images. We resized the images of our dataset to fit with the input size of AlexNet.

(3) GoogLeNet: This model was originally trained on ImageNet dataset and introduced by Szegedi, et al, [41] at 2015. It is composed of 22 convolutional layers involving nine Inception modules. Each Inception module has three dissimilar kernel sizes, which are 1x1, 3x3, and 5x5 for convolutional, as well as, 3x3 for pooling. The input size is 224x224x3 of the color image. In a similar manner to the other CNNs, the training stage of the convolutional filter entries is based on the stochastic gradient descent algorithm (SGD). GoogLeNet structure helps to have multi-level feature extraction at each step.

All these pre-trained Networks have been fine-tuned to classify feet's skin into two classes namely normal and abnormal (DFU) and re-trained with our dataset. GoogLeNet, AlexNet, and VGG16 Networks are commonly employed for classification with transfer learning. GoogLeNet utilizes multi-size convolution kernels, in addition to pooling, inside a single layer. AlexNet is considered the first improvement of CNNs. Finally, the VGG network follows the common arrangement of the basic CNN, i.e., a sequence of convolutional, max pooling, activation, and full-connected layers.

3.5 Proposed Approach-DFU_QUTNet Model

An innovative architecture of a deep convolutional neural network called DFU_QUTNet is proposed for improving the extraction of major features related to the classification of DFU. It is designed based on the idea of the Directed Acyclic Graph (DAG) concept. Two significant issues are considered for using this type of networks. The first issue is that increasing the number of convolutional layers to a traditional CNN model in order to enhance the accuracy is good at a limited number of layers then increasing more layers which may lead to a decline in performance (ii) A network with a small number of layers and simple structure is sufficient for some applications. However, DFU Classification requires a network with a more complicated structure to extract more features in order to discriminate between normal and abnormal classes. Our proposed model has an excellent benefit of increasing its width without significantly increasing its computing cost. This is also helping not only to boost the details possible to learn but also its accuracy. The general workflow of our classification illustrated in Fig8.

The DFU_QUTNet structure consists of several layers, which are:

(1) Input layer: it consists of three channels. Each channel has a size of 224×224 pixels. Initially, the input image is cropped into 224×224 patches of two classes of normal class and abnormal class as shown in Fig 5, then these patches are used to train the DFU_QUTNet network.

(2) Convolutional layer: This layer convolves the outcome of the previous layer through learnable filter set [53] since the weights identify the convolution filter. All filters are slid throughout the height and width of the input volume to produce two-dimensional activation maps of the corresponding filters. Note that each filter has an identical depth as the input [54]. In addition, the output size can be manipulated through three hyper-parameters, which are the zero-padding (i.e., padding zeros around the input borders for preserving its size), stride (i.e., number of pixels that the filter skips whereas sliding across the image), and depth (i.e., it represents the number of filters operated to the input image. These filters identify structures, for example, blobs, corners, edges, etc.). In this work, the model consists of 17 convolutional layers with filters of the size of 3×3. Each convolutional layer is followed by layers of batch normalization and the Rectified Linear Unit.

(3) Batch normalization (BN) layer: This layer normalizes each input channel through a mini-batch. It is utilized for speeding up CNN's training process, as well as, diminishing the sensitivity of the network initialization [55]. In this work, the batch normalization layer is placed between the convolution and ReLU layers. There are 17 batch normalization layers used in this work. The batch normalization layer mechanism starts by normalizing the activations of each channel via subtracting the mini-batch average and dividing by the standard deviation of the mini-batch. Then, the input is moved the BN layer with a learnable offset β and scaled it via a learnable scale factor γ.

(4) Rectified Linear Unit (ReLU): This layer performs data filtering by employing the function max (0, x) [56], where x is the input to the neuron.

(5) Addition Layer: This layer adds inputs from two or multiple neural network layers. In order to use this layer, all inputs to this layer should have the same dimension.

(6) Average pooling layer: This layer reduces its input size by partitioning it into rectangular pooling areas with different sizes such as 2×2, 3×3, etc., then computing the average values of small spatial block [57]. Average pooling gives normalized feature information from a block which may contain significant as well as less significant pixel information. Even a block may contain only significant pixel information which may denote sharpening features or the block which may contain only pixels having less significant characteristics other than sharpening features. Max polling derives sharpening features which may include sharp edges, corners, and lines. In order to avoid losing some features which can be eliminated by the max pooling layer, we have applied global average pooling in the last part of the network instead of max pooling. All these features both significant and less significant are important to discriminate between classes.

(7) Dropout Layer-: This layer is used to avoid overfitting and improve performance [58]. In this layer, neurons are randomly turned on and off in each layer to prevent overfitting problem. One dropout layer has been employed in the proposed method between the fully connected layers with probability p=0.5.

(8) Fully Connected (FC) layer: this layer connects to all the neurons of the previous layer [57]. This layer combines the features for classifying the patches of the feet's skin into two categories: normal, and abnormal. Two FC layers are utilized in our proposed DFU_QUTNet network.

As shown in Fig. 10, the output layer is located on the top of the last fully connected layer. In this layer, the Softmax function is employed for classification as shown in Fig 9. The overall number of DFU_QUTNet layers is 58 as described in Table 1. The output extracted features by DFU_QUTNet network are utilized for training the KNN classifier (DFU_QUTNet+KNN) and the SVM (DFU_QUTNet+SVM) as shown in Fig 9. Note that the KNN classifier categorizes the objects according to the nearest training samples in the space of feature. An object is categorized via it's neighbor's majority. Support vector machine (SVM) is a margin-based classifier. The idea of this algorithm is to discover the optimal linear partition between two classes so that objects are in the maximum distance of that line. Different types of kernels which are linear, polynomial, and radial basis function (RBF) that can be employed in SVM. The DFU_QUTNet structure is new which helps to have multi-level feature extraction at each step.

With each convolution layer, there are more discriminative features. Healthy skin has lower activations because of the lack of skin abnormalities as shown in Fig 11. On the other hand, skin with an ulcer has higher activations because of skin abnormality as shown in Fig 12 and Fig 13. Our model and pre-trained models have been trained on our dataset for 100 epochs until the learning stopped.

By using same training parameters that were used to train our network and pre-trained networks (GoogleNet[41], AlexNet[43], and VGG-16[44]), we trained DFUNet described in [59].DFUNet is the state-of-the-art network in the task of DFU classification. We wrote the code for the DFUNet network structure from scratch. The permission to use the DFUNet network has been approved by the author of the paper.

Lastly, the processor properties used in this experiment are Intel (R) Core TM i7-5829K CPU @ 3.30 GHz, the RAM was 16 GB and the GPU was 8 GB. Matlab2018 was used to implement the experiments.

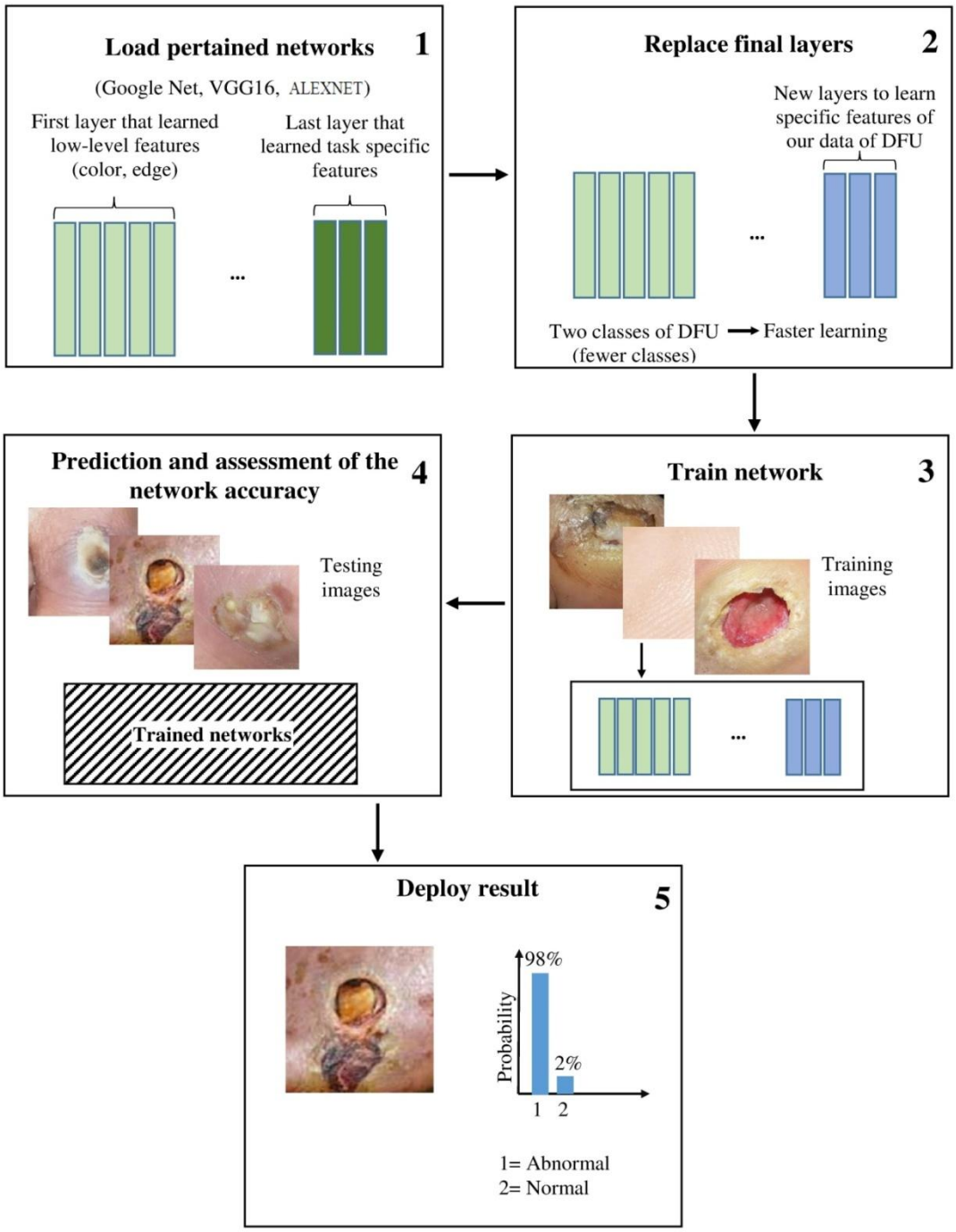

Fig.7 Transfer Learning Steps

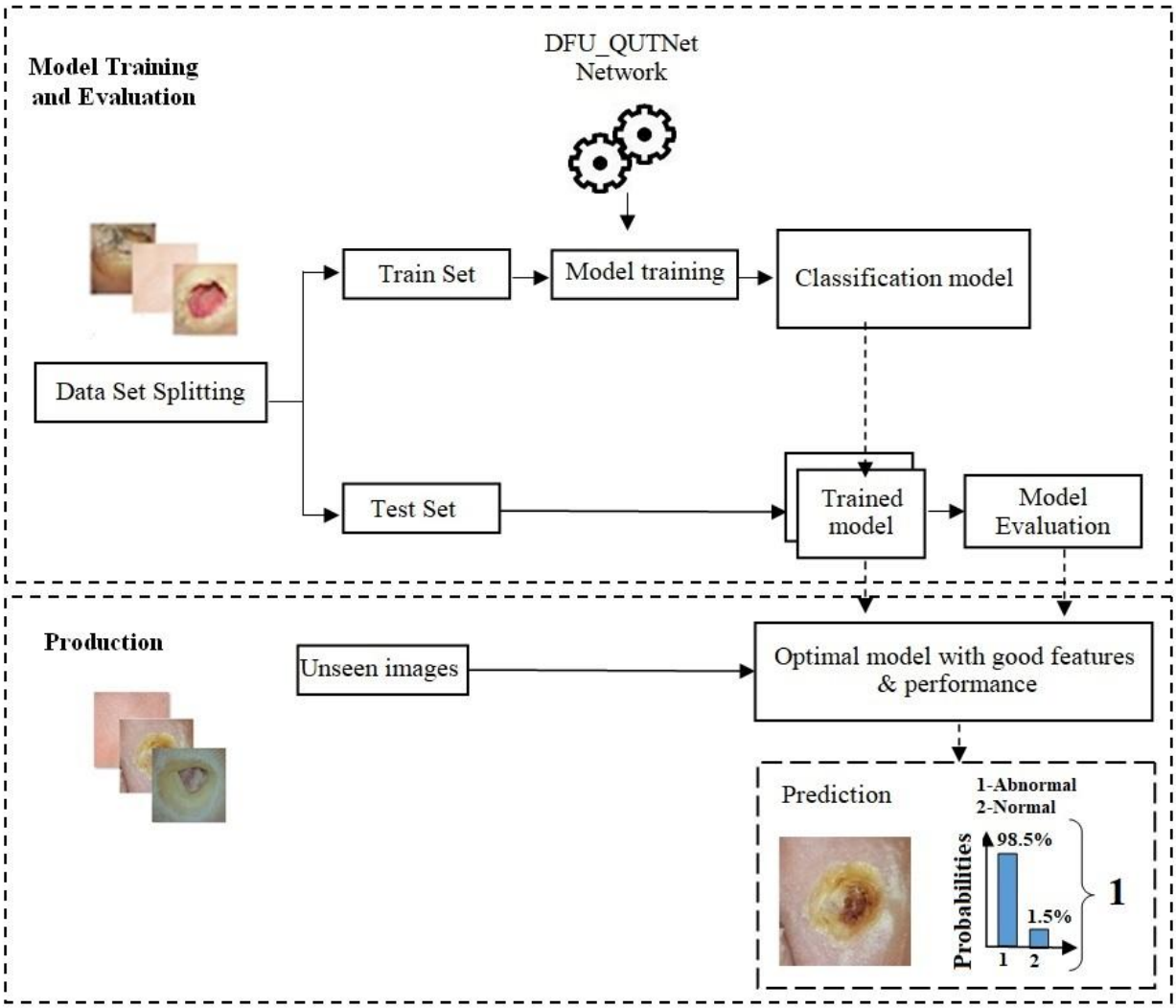

Fig.8 Illustration of our classification pipeline

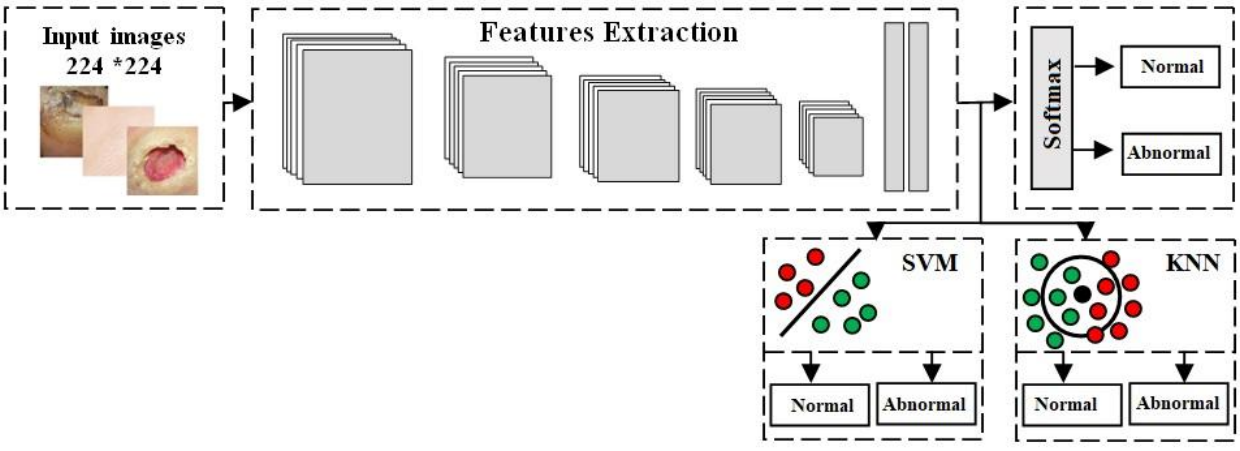

Fig.9 Overview of the network training steps

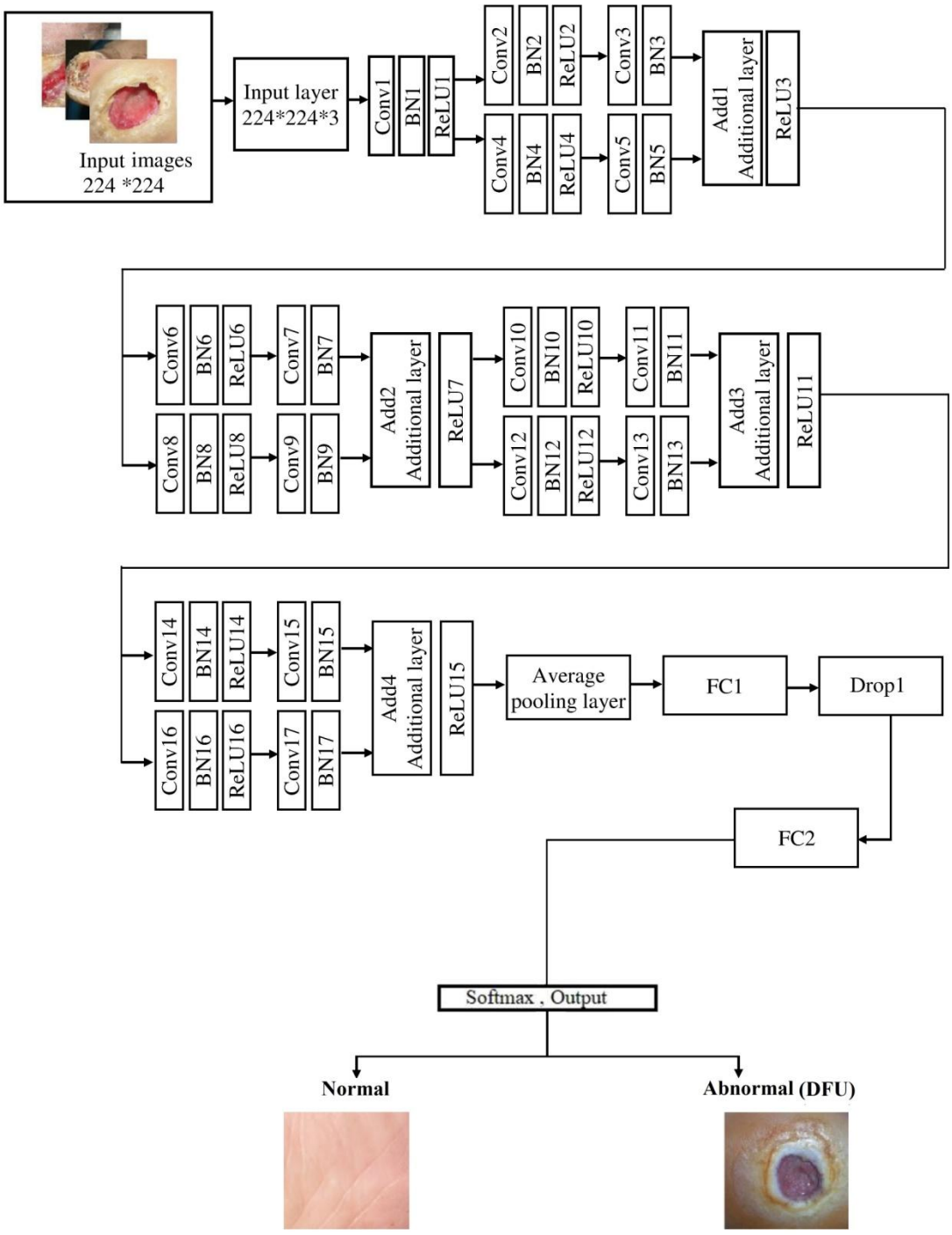

Fig.10 DFU_QUTNet architecture

Table1 DFU_QUTNet architecture, Conv refers to convolutional layers, BN refers Batch Normalization layer, ReLU refers Rectified Linear Unit, drop refers dropout layer, FC refers fully connected layer.

Name of layer	Kernel size and stride	Activations
Input layer	-	224×224×3
Conv1, BN1, ReLU1	Conv1: Kernel size=3×3, stride=1	224×224×32
Conv2, BN2, ReLU2	Conv2: Kernel size=3×3, stride=2	112×112×32
Conv3, BN3	Conv3: Kernel size=3×3, stride=1	112×112×32
Conv4, BN4, ReLU4	Conv4: Kernel size=3×3, stride=2	112×112×32
Conv5, BN5	Conv5: Kernel size=3×3, stride=1	112×112×32
Add1	Addition of two inputs	112×112×32
ReLU3	Activation Function	112×112×32
Conv6, BN6, ReLU6	Conv6: Kernel size=3×3, stride=2	56×56×64
Conv7, BN7	Conv7: Kernel size=3×3, stride=1	56×56×64
Conv8, BN8, ReLU8	Conv8: Kernel size=3×3, stride=2	56×56×64
Conv9, BN9	Conv9: Kernel size=3×3, stride=1	56×56×64
Add2	Addition of two inputs	56×56×64
ReLU7	Activation Function	56×56×64
Conv10, BN10, ReLU10	Conv10: Kernel size=3×3, stride=2	28×28×128
Conv11, BN11	Conv11: Kernel size=3×3, stride=1	28×28×128
Conv12, BN12, ReLU12	Conv12: Kernel size=3×3, stride=2	28×28×128
Conv13, BN13	Conv13: Kernel size=3×3, stride=1	28×28×128
Add3	Addition of two inputs	28×28×128
ReLU11	Activation Function	28×28×128
Conv14, BN14, ReLU14	Conv14: Kernel size=3×3, stride=2	14×14×256
Conv15, BN15	Conv15: Kernel size=3×3, stride=1	14×14×256
Conv16, BN16, ReLU16	Conv16: Kernel size=3×3, stride=2	14×14×256
Conv17, BN17	Conv17: Kernel size=3×3, stride=1	14×14×256
Add4	Addition of two inputs	14×14×256
ReLU15	Activation Function	14×14×256
Average Pooling Layer	Kernel size=8×8, stride=1,	7×7×256
FC1	100 fully connected	1×1×100
Drop1	Dropout layer with learning rate:0.5	1×1×100
FC2	2 fully connected	1×1×2
Softmax layer	0=Normal, 1=Abnormal(DFU)	1×1×2

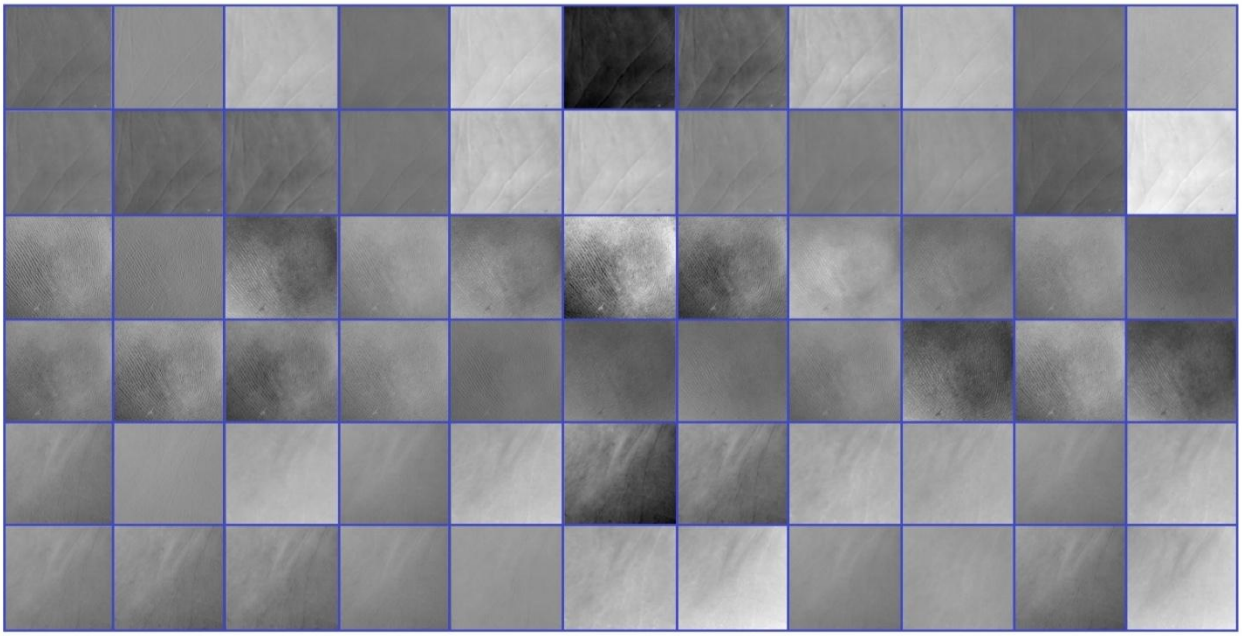

Fig. 11 Some learnable filters from the first convolutional layer for normal skin

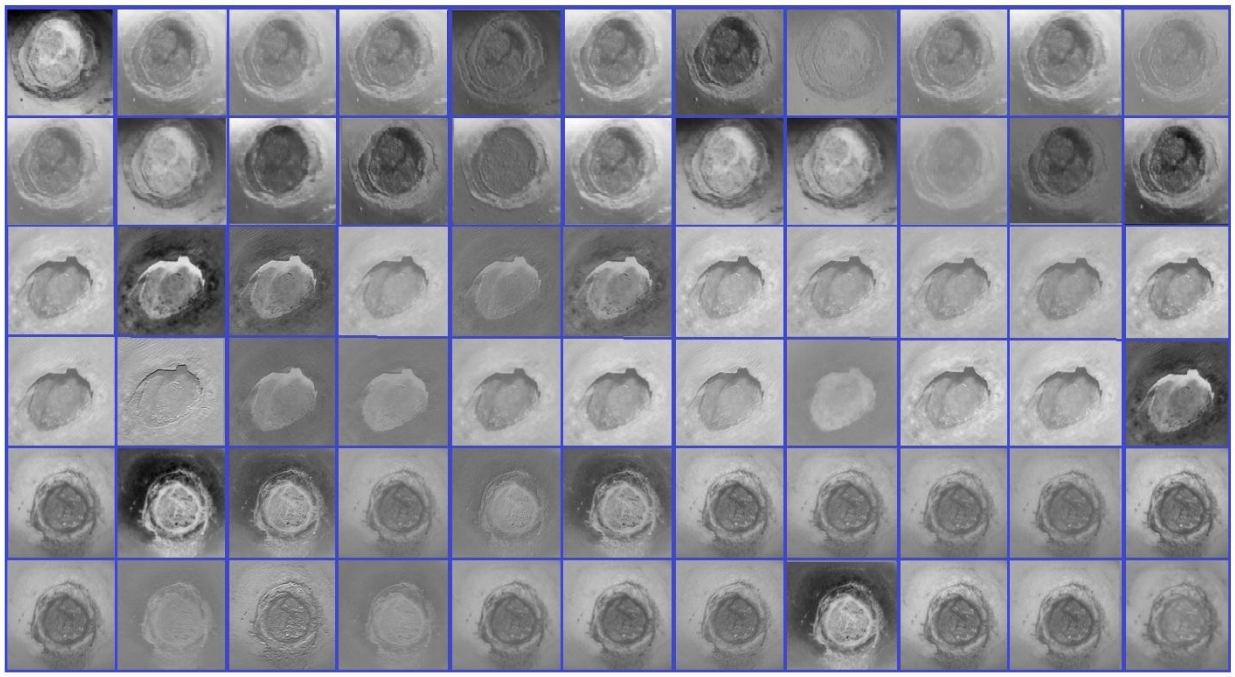

Fig. 12 Some learnable filters from the first convolutional layer for abnormal skin

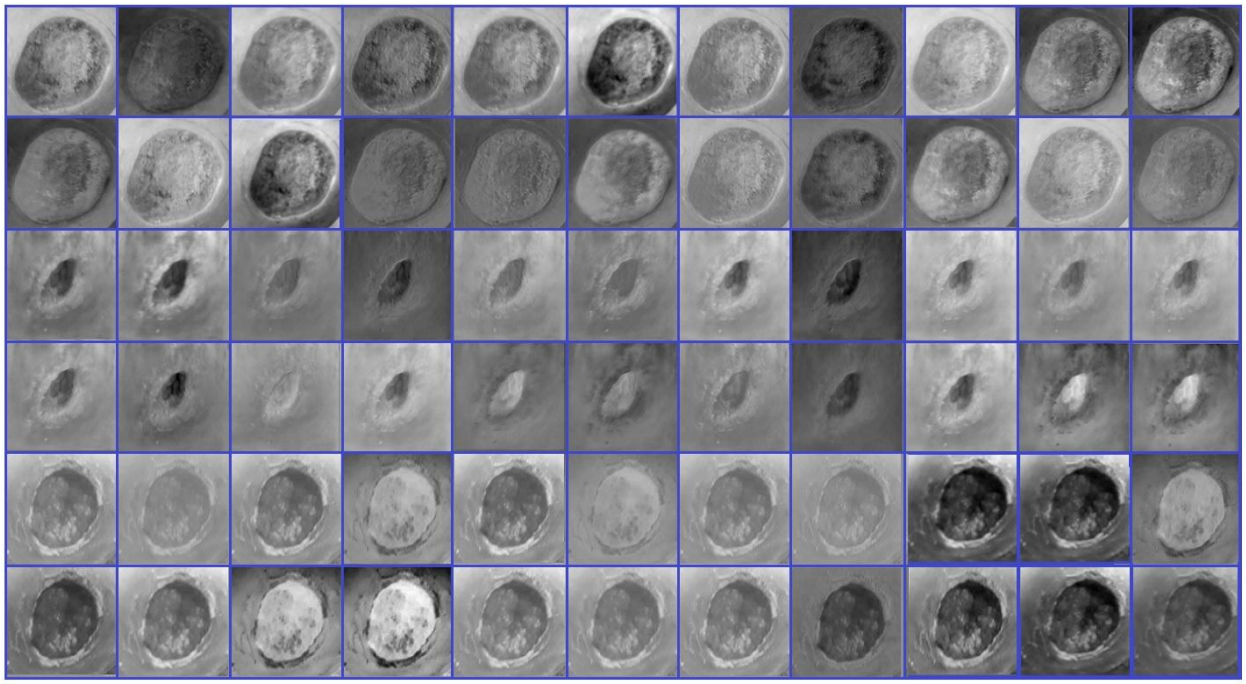

Fig. 13 Some learnable filters from the second convolutional layer for abnormal skin

4 Experimental Results

As mentioned above, we divided our dataset to the training phase and test phase. We implemented several experiments on our challenging dataset to evaluate the classification performance of our network and fine-tuned networks that we have used. Our model and fine-tuned models effectiveness are measured by a F1-score. F1- scores convey the balance between the precision (P) and the recall(R) (Eq3). Recall and Precision are two basic parameters for the assessment of the proposed approach, which are calculated (Eq 1,2):

$$Precision = \frac{TP}{TP+FP} \quad (1)$$

$$Recall = \frac{TP}{TP+FN} \quad (2)$$

$$F1-Score = 2 \times \frac{Precision \times Recall}{Precision+Recall} \quad (3)$$

Where TP (True Positive) is the number of images that the Network correctly identifies as relevant.TN (True Negative) is the number of images the Network correctly identifies as irrelevant. FP (False Positive) is the number of images the Network falsely identifies as relevant.FN (False Negative) is the number of relevant images that the Network fails to identify.

Table 2 DFU_QUTNet with different classifiers

classifier	Precision (%)	Recall (%)	F1-Score (%)
DFU_QUTNet	94.2	92.6	93.4
DFU_QUTNet+KNN	93.8	92.7	93.2
DFU_QUTNet+SVM	**95.4**	**93.6**	**94.5**

Table 2 reported the measurements of our proposed DFU_QUTNet with different classifiers. DFU_QUTNet with SVM has the highest measures in precision, recall, and F1-score with 95.4%, 93.6%, and 94.5% respectively. DFU_QUTNet with KNN achieved the lowest scores of 93.8% for precision and 93.2% for the F1-score while the recall score is 92.7% which is higher than the recall score of DFU_QUTNet using softmax. DFU_QUTNet using softmax achieved 94.2%, 92.6%, 93.4% for precision, recall, and F1-score respectively.

Table 3 Comparison of DFU_QUTNet with state-of-the-art Networks for classification

Network	Precision (%)	Recall (%)	F1-Score (%)
AlexNet[43]	91.1	87.2	89.1
VGG16[44]	92.3	89.7	90.9
GoogleNet[41]	**95.6**	90.5	92.9
Proposed DFU_QUTNet+SVM	95.4	**93.6**	**94.5**

For the sake of comparison, we have trained and tested three state-of-the-art networks on our dataset and compared them to our proposed DFU_QUTNet as reported in Table3. AlexNet achieved the lowest scores with 91.1% for precision, 87.2% for recall and 89.1 for F1-score. Our proposed DFU_QUTNet with SVM has the highest evaluation performance in the recall, with a score of 93.6% and F1-score with 94.5%, whereas, GoogleNet has the highest score in precision with 95.6%. We also compared our proposed model to DFUNet as reported in Table4. Our proposed model outperformed DFUNet network in all evaluation measurements. DFU_QUTNet without any other classifiers is still superior to the state-of-the-art Networks.

Due to the type of structure that our model is designed for, it has the ability to find more subtle changes and discriminate between two classes of normal and abnormal. It also handled many hard cases such as small sizes of DFU, skin wrinkles and patches with a toe.

Table4: Competitive results of DFU_QUTNet+SVM classification with DFUNet Network

Network	Precision (%)	Recall (%)	F1-Score (%)
DFUNet [59]	93.8	92.5	93.1
Proposed DFU_QUTNet+SVM	**95.4**	**93.6**	**94.5**

5 Conclusions and Future Work

A novel CNN model, DFU_QUTNet, for automated classification of DFU into two classes: normal (healthy skin) and abnormal (DFU) is proposed in this paper. The architecture of DFU_QUTNet is designed based on the idea of increasing the width. It has the great advantage of boosting its width without drastically boosting its computational cost. We collected a large dataset of total 754-feet of patients with DFU and healthy skin. The dataset has been labeled into two classes normal and abnormal by an expert. Our proposed DFU_QUTNet model has been trained on our dataset. Features extracted by our proposed model were used to train the SVM and KNN classifiers. We compared the performance of our network with state-of-the-art CNN networks (GoogleNet, AlexNet, and VGG16). These CNN networks were trained on a large dataset of the ImageNet dataset. After the fine-tuning process, we re-trained them on our dataset with transferring the previous learning of these deep learning networks. The comparison shows that the proposed framework, DFU_QUTNet with SVM, has higher F1-Score of 94.5% than the state-of-the-practice methods. The proposed DFU_QUTNet network is not limited to the DFU classification task only. For instance, we aim to fine-tune DFU_QUTNet model to classify seven classes of skin cancer. Since DFU_QUTNet network extracted significant features to differentiate between classes of normal and abnormal, we also aim to employ DFU_QUTNet network in the task of DFU detection.

References

1. S.Wild, G. Roglic, A. Green, R. Sicree, and H. King (2004) Global prevalence of diabetes estimates for the year 2000 and projections for 2030. Diabetes care, vol. 27, no. 5, pp. 1047–1053.
2. World Health Organization (2016) Global report on diabetes. http://www.who.int/diabetes/global-report/en Accessed 6 Feb. 2009.
3. A. J. Boulton, L. Vileikyte, G. Ragnarson-Tennvall, and J. Apelqvist (2005) The global burden of diabetic foot disease. The Lancet, vol. 366, no. 9498, pp. 1719–1724.
4. C. Liu, J.J. van Netten, J.G. van Baal, S.A. Bus, and F. van der Heijden (2015) Automatic detection of diabetic foot complications with infrared thermography by asymmetric analysis. J. Biomed. Opt., vol. 20, no. 2, p. 26003.
5. D. G. Armstrong, L. A. Lavery, and L. B. Harkless (1998) Validation of a diabetic wound classification system: the contribution of depth, infection, and ischemia to risk of amputation. Diabetes care, vol. 21, no. 5, pp. 855 859.
6. P. Cavanagh, C. Attinger, Z. Abbas, A. Bal, N. Rojas, and Z.-R. Xu (2012) Cost of treating diabetic foot ulcers in five different countries. Diabetes / metabolism research and reviews, vol. 28, no. S1, pp. 107– 111.
7. C.E. Hazenberg, J.J. van Netten, S.G. van Baal, S.A. Bus, and S. Bus (2014) Assessment of signs of foot infection in diabetes patients using photographic foot imaging and infrared thermography. Diabetes Technol. Ther., vol. 16, no. 6, pp. 370-377.
8. L. A. Lavery, D. G. Armstrong, and L. B. Harkless (1996) Classification of diabetic foot wounds. Journal of foot and Ankle Surgery, vol. 35, no. 6, pp. 528–531.
9. O. El-Gayar, P. Timsina, N. Nawar, and W. Eid (2013) A systematic review of it for diabetes self-management: are we there yet?. International journal of medical informatics, vol. 82, no. 8, pp. 637–652.
10. L. Vilcahuaman, R. Harba, R. Canals, M. Zequera, C. Wilches, M. Arista, L. Torres, and H. Arbanil (2015) Automatic analysis of plantar foot thermal images in at-risk Type II diabetes by using an Infrared camera . IFMBE Proc., vol. 51, pp. 228-231.
11. S. Franc, A. Daoudi, S. Mounier, B. Bou cherie, H. Laroye, C. Peschard, D. Dardari, O. Juy, E. Requeda, L. Canipel et al. (2011) Telemedicine: what more is needed for its integration in everyday life?. Diabetes & metabolism, vol. 37, pp. S71–S77.
12. L. Fraiwan, M. AlKhodari, J. Ninan, B. Mustafa, A. Saleh, and M. Ghazal (2017) Diabetic foot ulcer mobile detection system using smart phone thermal camera: A feasibility study. Biomed. Eng. Online, vol. 16, no. 1, p. 117.
13. Fraiwan, L., Ninan, J., & Al-Khodari, M. (2018). Mobile application for ulcer detection. The open biomedical engineering journal, 12, 16.

14. P. Foltynski, J. M. Wojcicki, P. Ladyzynski, K. Migalska-Musial, G. Rosinski, J. Krzymien, and W. Karnafel (2011) Monitoring of diabetic foot syndrome treatment: some new perspectives. Artificial organs, vol. 35, no. 2, pp. 176–182.
15. Hazenberg, C. E., van Netten, J. J., van Baal, S. G., & Bus, S. A. (2014). Assessment of signs of foot infection in diabetes patients using photographic foot imaging and infrared thermography. Diabetes technology & therapeutics, 16(6), 370-377.
16. F. L. Bowling, L. King, J. A. Paterson, J. Hu, B. A. Lipsky, D. R. Matthews, and A. J. Boulton (2011) Remote assessment of diabetic foot ulcers 11 using a novel wound imaging system. Wound Repair and Regeneration, vol. 19, no. 1, pp. 25–30.
17. J. Clemensen, S. B. Larsen, M. Kirkevold, and N. Ejskjaer (2008) Treatment of diabetic foot ulcers in the home: video consultations as an alternative to outpatient hospital care. International journal of telemedicine and applications, vol. 2008, p. 1.
18. Jegede, O., Ferens, K., Griffith, B., & Podaima, B. (2015, July). A smart shoe to prevent and manage diabetic foot diseases. In Int'l Conf. Health Informatics and Medical Systems (pp. 47-54).
19. Wang, L., Pedersen, P. C., Agu, E., Strong, D. M., & Tulu, B. (2017). Area determination of diabetic foot ulcer images using a cascaded two-stage SVM-based classification. IEEE Transactions on Biomedical Engineering, 64(9), 2098-2109.
20. Goyal, M., Yap, M. H., Reeves, N. D., Rajbhandari, S., & Spragg, J. (2017, October). Fully convolutional networks for diabetic foot ulcer segmentation. In 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC) (pp. 618-623). IEEE.
21. Kolesnik, M., & Fexa, A. (2006, June). How robust is the SVM wound segmentation? In Proceedings of the 7th Nordic Signal Processing Symposium-NORSIG 2006 (pp. 50-53). IEEE.
22. F. Veredas, H. Mesa, and L. Morente (2010) Binary tissue classification on wound images with neural networks and bayesian classifiers. IEEE transactions on medical imaging, vol. 29, no. 2, pp. 410–427.
23. Veredas, Francisco J., Rafael M. Luque-Baena, Francisco J. Martín-Santos, Juan C. Morilla-Herrera, and Laura Morente (2015) Wound image evaluation with machine learning. Neurocomputing, vol. 164, pp.1–11.
24. Y. Deng and B. S. Manjunath (2001) Unsupervised segmentation of colortexture regions in images and video. IEEE Trans. Pattern Anal. Mach. Intell., vol. 23, no. 8, pp. 800–810.
25. Ram Avrahami, M. D., Jonathan Rosenblum, D. P. M., Michael Gazes, D. P. M., & Sean Rosenblum, D. P. M. (2015). The effect of combined ultrasound and electric field stimulation on wound healing in chronic ulcerations. Wounds, 27(7), 199-208.
26. M. H. Hermans (2010) Wounds and ulcers: back to the old nomenclature. Wounds, vol. 22, no. 11, pp. 289–93.
27. Zhu, Xiaobin, Zhuangzi Li, Xiao-Yu Zhang, Peng Li, Ziyu Xue, and Lei Wang (2018) Deep convolutional representations and kernel extreme learning machines for image classification. Multimedia Tools and Applications 1-20.
28. Huang, Jin, Bin Li, Jia Zhu, and Jian Chen (2017) Age classification with deep learning face representation. Multimedia Tools and Applications 20231-20247.
29. Sert, M. & Boyacı (2019) Sketch recognition using transfer learning. Multimedia Tools and Applications. https://doi.org/10.1007/s11042-018-7067-1
30. Al-Zubaidi, L. (2016). Deep learning based nuclei detection for quantitative histopathology image analysis (Doctoral dissertation, University of Missouri--Columbia).
31. Litjens, G.; Kooi, T.; Bejnordi, B. E.; Setio, A. A. A.; Ciompi, F.; Ghafoorian, M. and Sánchez, C. I (2017) A survey on deep learning in medical image analysis. Medical image analysis, 42, 60-88.
32. Bakator, M., and Radosav, D (2018) Deep Learning and Medical Diagnosis: A Review of Literature. Multimodal Technologies and Interaction, 2(3), 47.
33. Vijayalakshmi A & Rajesh Kanna B (2019) Deep learning approach to detect malaria from microscopic images. Multimedia Tools Application, 1-21.
34. Zhang, Gang, Ching-Hsien Robert Hsu, Huadong Lai, and Xianghan Zheng (2018) Deep learning based feature representation for automated skin histopathological image annotation. Multimedia Tools and Applications 77.8, 9849-9869.
35. Nair, Lekha R. (2019) RetoNet: a deep learning architecture for automated retinal ailment detection. Multimedia Tools and Applications 1-10.

36. LeCun, Y.; Bengio, Y. and Hinton, G (2015) Deep learning. Nature 521(7553), 436.
37. Zeiler, M. D. and Fergus, R (2014) Visualizing and understanding convolutional networks. In European conference on computer vision, (pp. 818-833), Springer.
38. Lu, H., Li, Y., Uemura, T., Ge, Z., Xu, X., He, L., & Kim, H. (2018). FDCNet: filtering deep convolutional network for marine organism classification. Multimedia tools and applications, 77(17), 21847-21860.
39. Alzubaidi L., Al-Shamma O., Fadhel M.A., Farhan L., Zhang J. (2020) Classification of Red Blood Cells in Sickle Cell Anemia Using Deep Convolutional Neural Network. In: Abraham A., Cherukuri A., Melin P., Gandhi N. (eds) Intelligent Systems Design and Applications. ISDA 2018 2018. Advances in Intelligent Systems and Computing, vol 940. Springer, Cham.
40. Geng, Lei, Tingyu Yan, Zhitao Xiao, Jiangtao Xi, and Yuelong Li. (2017) Hatching eggs classification based on deep learning. Multimedia Tools and Applications 1-12.
41. 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).
42. Sainath, T. N., Mohamed, A. R., Kingsbury, B., & Ramabhadran, B. (2013, May). Deep convolutional neural networks for LVCSR. In 2013 IEEE international conference on acoustics, speech and signal processing (pp. 8614-8618). IEEE.
43. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems (pp. 1097-1105).
44. Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.
45. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).
46. Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4700-4708).
47. Salamon, Justin, and Juan Pablo Bello (2017) Deep convolutional neural networks and data augmentation for environmental sound classification. IEEE Signal Processing Letters 24.3: 279-283.
48. Wang, J., & Perez, L. (2017). The effectiveness of data augmentation in image classification using deep learning. Convolutional Neural Networks Vis. Recognit.
49. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S. & Berg, A. C. (2015). Imagenet large scale visual recognition challenge. International journal of computer vision, 115(3), 211-252.
50. Ravishankar, H., Sudhakar, P., Venkataramani, R., Thiruvenkadam, S., Annangi, P., Babu, N., & Vaidya, V. (2016). Understanding the mechanisms of deep transfer learning for medical images. In Deep Learning and Data Labeling for Medical Applications (pp. 188-196). Springer, Cham.
51. Yu, Y., Lin, H., Meng, J., Wei, X., Guo, H., & Zhao, Z. (2017). Deep transfer learning for modality classification of medical images. Information, 8(3), 91.
52. Cheng, Phillip M., and Harshawn S. Malhi. (2017) Transfer learning with convolutional neural networks for classification of abdominal ultrasound images. Journal of digital imaging 234-243.
53. Vedaldi, A., & Lenc, K. (2015, October). Matconvnet: Convolutional neural networks for matlab. In Proceedings of the 23rd ACM international conference on Multimedia (pp. 689-692). ACM.
54. Gibiansky,Andrew (2018) Convolutional neural networks. http://andrew.gibiansky.com/blog/machine-learning / convolutional- neural-networks/. Accessed 26 December 2018.
55. Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. arXiv preprint arXiv:1502.03167.
56. Dahl, G. E., Sainath, T. N., & Hinton, G. E. (2013, May). Improving deep neural networks for LVCSR using rectified linear units and dropout. In 2013 IEEE international conference on acoustics, speech and signal processing (pp. 8609-8613). IEEE.
57. Koushik, J. (2016). Understanding convolutional neural networks. arXiv preprint arXiv:1605.09081.
58. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. The Journal of Machine Learning Research, 15(1), 1929-1958.
59. 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.

60. Vesal, S., Ravikumar, N., Davari, A., Ellmann, S., & Maier, A. (2018, June). Classification of breast cancer histology images using transfer learning. In International Conference Image Analysis and Recognition (pp. 812-819). Springer, Cham.
61. Nawaz, W., Ahmed, S., Tahir, A., & Khan, H. A. (2018, June). Classification Of Breast Cancer Histology Images Using ALEXNET. In International Conference Image Analysis and Recognition (pp. 869-876). Springer, Cham.