Journal of Engineering Science and Technology
Vol. 16, No. 3 (2021) 2001 - 2017

COMPARISON OF HYBRID CONVOLUTIONAL NEURAL NETWORKS MODELS FOR DIABETIC FOOT ULCER CLASSIFICATION

LAITH ALZUBAIDI[1,2,*], ALAA AHMED ABBOOD[2], MOHAMMED A. FADHEL[3], OMRAN AL-SHAMMA[2], JINGLAN ZHANG[1]

[1]School of Computer Science, Queensland University of Technology, Brisbane, QLD 4000, Australia
[2]AlNidhal Campus, University of Information Technology & Communications, Baghdad 10001, Iraq
[3]College of Computer Science and Information Technology, University of Sumer, Thi Qar 64005, Iraq
*Corresponding Author: laith.alzubaidi@hdr.qut.edu.au

Abstract

In this paper, we present a comparison of four proposed hybrid deep convolutional neural network models for diabetic foot ulcer (DFU) classification to discriminate between abnormal (DFU) and normal (healthy skin) classes. Increasing the depth in single branch deep convolutional neural networks does not always significantly contribute to their overall performance. It may actually lead to a drop in performance due to gradient vanishing issue. Therefore, our proposed models were designed based on the concept of multiple branches network. Traditional convolutional layers and multi-branch parallel convolutional layers were combined to design four deep aggregated models. All four models have six blocks of parallel convolutional layers, but the number of branches of parallel convolutional layers ranges from two to five. Parallel convolutional layers have been employed using different filter sizes on the same input and then concatenated for better feature extraction. To overcome the issues of overfitting and a small amount of training data, we applied several data augmentation techniques. The proposed models were trained with original images first, then with original images plus augmented images, which improved the performance. We empirically prove that a model with four branches outperforms models with two, three, or five branches of parallel convolutions in the task of DFU classification. This model also outperformed the latest DFU classification methods by achieving an F1 score of 95.8% on the DUF dataset.

Keywords: Classification, Deep convolutional neural network (DCNN), Deep learning, Diabetic foot ulcer, Multi-branch network.

1. Introduction

One of the major diabetic complications that influences the lower extremities is diabetic foot ulcers (DFUs). DFUs can be very dangerous, and diabetics can develop unknown cuts or wounds on the bottom of their feet if they lack feeling in them. This can result in open sores, which may lead to infections, possible amputation of the foot or leg, and even death [1]. Every year, 9.1 to 26.1 million diabetes patients worldwide develop DFUs [2]. This statistic is based on the 2015 overall occurrence data from the International Diabetes Federation [2]. It was found that diabetic people had a 15-25% lifespan probability of getting a DFU, with approximately 85% of them receiving lower limb amputations if they did not get DFU treatment [3].

Several economical solutions for distant detection and avoidance of DFUs have been released due to the propagation of information communication technology. One of these solutions involves the use of automatic intelligent telemedicine systems, which alongside the available healthcare services, can provide high quality, efficient, and cost-effective DFU treatment. The last few years have shown a huge progression in the field of computer vision, mainly in terms of the critical and complicated issue of recognizing images from numerous domains, like human motion [4]. More specifically, medical image analysis of the different modalities has significantly utilized deep learning (DL) and computer vision techniques, which include ultrasound, dermoscopy, X-rays, CT scans, and MRI [5]. Recently, computer vision algorithms have expanded to include the evaluation of several skin conditions, such as DFUs and skin cancer [6, 7].

Many of the contributions associated with computer vision techniques for DFU classification have been considered by various researchers. In general, it is possible to derive four classes from these contributions: development of algorithms constructed based on traditional machine learning techniques and the fundamentals of image processing, development of algorithms constructed based on DL techniques, research constructed using several image modalities, and smartphone applications for DFU classification.

Several researchers have proposed computer vision techniques constructed using the fundamentals of image processing and methods of supervised traditional machine learning for wound/DFU classification and detection. More specifically, these researchers have implemented the segmentation task via extracting colour and texture descriptors on small pieces of DFU/wound images, then applying traditional machine learning algorithms to recognize the normality of the skin pieces into DFU/healthy skin classes [8]. Handcrafted features in conventional machine learning are generally influenced by image resolution, illumination, and skin shades. In addition, these techniques have struggled with the segmentation of irregular wound or ulcer contours.

In contrast, unsupervised approaches depend on clustering algorithms, morphological operations, and image processing techniques that utilize several colour spaces to segment wounds from images [9]. One of the techniques for capturing image data involves using a capture box and was applied by Wang et al. [10]. They also used support vector machine-based classification with a cascaded two-stage process to determine the DFU area and presented the super-pixel method for the segmentation and extraction of DFU features for classification. Although they demonstrated encouraging outcomes, however, this system did not perform

well on a large dataset. Unfortunately, the patient needs to place their bare foot in a straight line with the display of the image capture box, which means the box cannot be used to collect data. Note that this would not be allowed in healthcare facilities because of worries related to the control of infection.

Most of these techniques include a manual parameter-tuning process based on various multi-stage processes as well as input images. This makes the implementation of these techniques difficult within clinical settings. These methods have been used on small datasets that ranged from 10-180 images. Generally, the current novel approaches depend on traditional machine learning techniques and the fundamentals of image processing, which are not robust because of their reliance on certain rules and specific assumptions.

In contrast, DL approaches do not need powerful assumptions and have validated power in DFU segmentation and object localization. By adopting DL approaches, the completely automatic powerful DFU detection, classification, and segmentation have been achieved [6, 11, 12]. Numerous contributions have been made by several researchers on DFU segmentation and classification tasks. DFUNet is a novel DL model proposed by Goyal et al. [11] for classifying skin wounds on the foot area into two categories: healthy (normal skin) and DFU (abnormal skin). Another novel architecture of DL was developed by Wang et al. [13] to measure wound healing progress. It is based on the encoder-decoder architecture for executing segmentation and examination of the wound. After that, in unrelated computer vision techniques research, van Netten et al. [14] presented a diverse modality known as infrared thermal imaging for DFU detection. They found that there is considerable variation in temperature between the healthy skin of the foot and the DFU area and used a heatmap to detect the latter.

For the task of DFU classification, most of the methods used are traditional machine learning methodologies that are sensitive to different sizes, colours, and complex shapes. DFU images are very complex and require effective methods to classify them. DL techniques have become an alternative solution for diagnosis and overcome the problems of traditional machine learning methodologies. As DL techniques have so far been underemployed in the classification of DFU images, we were motivated to design effective and accurate DL models for DFU classification.

The contributions of this study are outlined below:

- Four hybrid deep convolutional neural network models are designed that aggregate the traditional convolutional layers with six blocks of parallel convolutional layers along with the global average pooling layer.
- It is empirically proven that the model with four branches of parallel convolutional layers is superior to the models with five, three, or two branches of parallel convolutions.
- The performance of DFU classification is enhanced. Our four-branch model achieved the highest F1 score of 95.8% with augmented data, which surpasses the state-of-the-art methods used for DFU classification.
- Several data augmentation techniques are employed to address the issues associated with small datasets.
- A concise review of the state-of-the-art DL methods and the classical techniques used to classify, detect, and segment DFUs is provided in the introduction section. Moreover, we review some of the scientific research

papers that describe the benefits of increasing the width of DL models in the related work section.

The rest of the paper is organized as follows: Section 2 presents the Related Work. Section 3 explains the paper methodology. Section 4 reports the results. Lastly, section 5 concludes the paper.

2.Related Work

DL is a machine learning technique that learns features from data that can involve text, images, or sounds. In the machine learning discipline, DL represents a recent and rapidly developing field. It aims to model abstraction from big data by utilizing multiple-layer deep neural models (DNN), which creates a data sense like texts, sounds, and images [15]. In general, two properties characterize DL: multiple layers of nonlinear processing units and unsupervised or supervised learning of feature presentations within each layer [15].

The first DL development was established in an artificial neural network (ANN) [15]. ANNs are inspired by biological neural networks, which can be described as networks of simple processing elements, i.e., neurons. These neurons work together to find a solution for scientific issues, such as image recognition and object segmentation. ANNs are considered the basis for DL [16]. The ANN simulated the object perceptions through connecting the artificial neurons in layers in such a way that it can extract the object features. In general, ANN improvements have declined due to their shallow architectures and the restricted computational ability of computers [17]. The actual influence of DL became obvious in 2006 [18], and since then, DL has been involved in several disciplines, such as bioinformatics and image recognition [19, 20]. Another type of ANN is the recurrent neural network (RNN), which has artificial neurons and behaves in a dynamic manner [21]. RNNs became an essential tool for processing sequential data [22]. Convolutional neural networks (CNNs) were then proposed, which are the reason that behind DL is popular today [23].

CNNs have worked well in several tasks, such as image classification and object detection systems [24, 25]. They also led to huge progress in the field of medical images [5]. CNNs have three main layers [23], which are presented in Fig. 1.

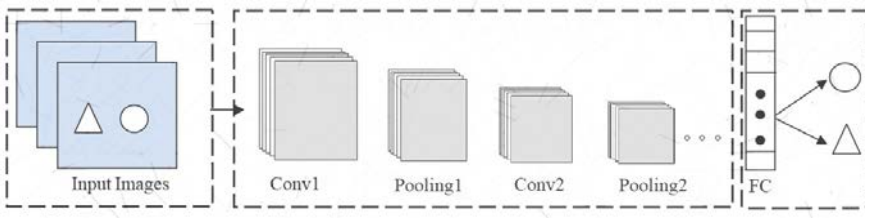

Fig. 1. A CNN's main layers. Conv = a convolutional layer, Pooling = a pooling layer, FC = a fully connected layer.

A CNN starts by searching the low-level features (like curves, lines, and edges) and then constructs further abstract features. The utilization of CNN techniques has become very attractive and valuable in applications where the extraction of features is necessary. Before entering the input image into the neural

network, it passes through a sequence of filters. The filters are trained on input data rather than having constant numbers in kernels. As the CNN is trained, the filters' learning becomes better and better for the related information, so the parameters of the model are optimized to minimize the loss function. There is frequently a straight correlation between model performance and the required amount of data. If the training set is larger, there is a greater variety of data, and the achieved generalization is better. In the last ten years, several CNN architectures have been proposed [21, 23].

In 2012, the AlexNet network was presented by Krizhevsky et al. [26]. This was one of the first networks to perform well on a challenging ImageNet dataset and significantly outperform previous methods. Three filter sizes (FS) of 11×11, 5×5, and 3×3 was used in the convolutional layers of AlexNet along with max pooling, and it was considered to be one of the first significant improvements to CNNs. Some of the key improvements were the ReLU activation function to avoid the vanishing gradient problem and the dropout layer to prevent overfitting. Furthermore, data augmentation was introduced, which meant that the images fed to the network were shown with random translation, rotation, and crop. This forced the network to be more aware of the attributes of the images rather than the images themselves. Lastly, more convolutional layers were stacked before the pooling layers, which improved the classification accuracy.

In 2013, The ZFNet network was proposed, which had a similar architecture to AlexNet but with some small changes [27]. For example, the size of the first convolutional layer filter was changed to 7×7 with a decreased stride (S) value rather than the 11×11 kernel used in the initial layer of the AlexNet. Also, ZFNet used 1.3 million images for training, whereas AlexNet used 15 million images. Then, in 2014, the VGGNet network was introduced [28], which added layers to improve accuracy. The VGGNet group applied filters that were only 3×3 in size, smaller than AlexNet's 11×11 first-layer filters and ZFNets 7×7 filters.

In 2015, the GoogleNet network was introduced [29], which proposed the use of parallel convolutions. This network used a filter size of 1×1 in the first layers, which risked the loss of some of the large features and the creation of a bottleneck. In 2016, the ResNet network introduced the idea of residual links, which meant the outcome of every two layers concatenated with the outcome of the previous two layers, and so on [30]. The ResNets group proposed three versions of ResNet: ResNet18, ResNet50, and ResNet101. Then, in 2017, the DenseNet network was proposed, which has entire blocks of layers connected to one another, leading to a more complex structure [31].

The models mentioned above were trained on the ImageNet dataset, which consists of nature images rather than medical images. Employing pre-trained models of ImageNet dataset for medical imaging tasks could not improve the performance of these tasks due to different learned features domains. It has been shown that a model trained from scratch can be as good as these models at medical imaging tasks [32] and that a different domain of transfer learning can slightly improve performance [20, 33]. In terms of model design, we have studied the advantages of all the models mentioned above and strived to employ these advantages (e.g., a dropout layer and parallel convolutions) in our proposed design.

3. Methodology

This section has three parts looking at the dataset, the data augmentation, and the proposed models.

3.1. Dataset

We utilized a dataset that has 754 images of patients' feet [12]. These images were classified into abnormal (DFU) and normal (healthy skin) classes. The dataset was collected from the Nasiriyah Hospital's diabetic center in Iraq, and ethical approval and written consent was obtained from all the relevant persons and patients. The images were cropped to a size of 224×224, which represents the two classes of normal and abnormal skin patches as shown in Fig. 2. We split the dataset into 80% for training and 20% for testing.

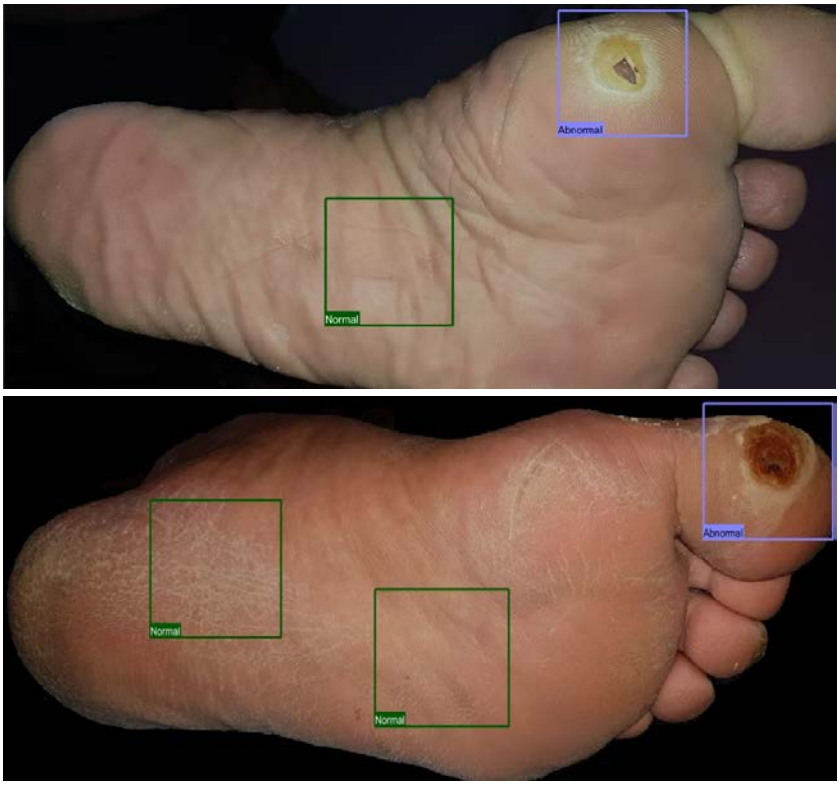

Fig. 2. Samples from the dataset. The blue square is the abnormal (DFU) class and the green squares are the normal class.

3.2. Data augmentation

The technique of data augmentation was used to overcome the data limitation and avoid overfitting. DL models require many training images to perform effectively. If there are a small number of training images, the parameters of the CNN will not learn well, which could lead to overfitting. Moreover, collecting large sets of medical images is costly and hard. We applied several image processing

techniques: rotating the training images by angles of 45, 90, 135, 180, 225, 270, and 315 degrees; horizontal and vertical flips; contrast and brightness in degree of 70 and 90 ; zooming in and out of the images; and colour space by isolating a single colour channel, such as R, G, or B. By applying the augmentation techniques, there were 18 times more training patches.

3.3. The proposed models

A hybrid CNN model was proposed to enhance significant feature extraction in DFU classification. It combined the key aspects of CNNs, including parallel and traditional convolutional layers. At the start of the model, there are sequences of traditional convolutional layers, which represent the first mode and utilize three filter sizes (3×3, 5×5, and 7×7). The second mode is parallel convolutional layers, which utilize multi-convolutional layers with different filter sizes. Parallel convolutions were employed for multiple levels of feature extraction and because they are useful for gradient propagation as the error can be backpropagated through multiple paths.

The proposed model was designed to have better extracted distinctive features for learning. Six parallel convolution blocks with different numbers of branches in the parallel convolutional layers were applied in the design. The first model has two parallel convolutions with two filter sizes of 1×1 and 3×3 at each block. The second model has three parallel convolutions with three filter sizes of 1×1, 3×3, and 5×5. The third model has four convolutional layers with four different filter sizes of 1×1, 3×3, 5×5, and 7×7. The fourth model has five convolutional layers with five different filter sizes of 1×1, 3×3, 5×5, 7×7, and 11×11.

Overall, we designed four hybrid deep convolutional neural models aggregating the traditional convolutional layers with six blocks of parallel convolutional layers and a global average pooling layer.

For the pooling part, we employed a global average pooling layer to reduce the spatial dimensions of a three-dimensional tensor to a one-dimensional tensor. Average, minimum, and maximum pooling layers use a sliding window (such as a size of 2×2 or 3×3) to reduce the size, however, the global average pooling layer performs a more extreme kind of dimensionality reduction [34], as illustrated in Fig. 3. This layer is more robust to spatial translations and helps to avoid overfitting.

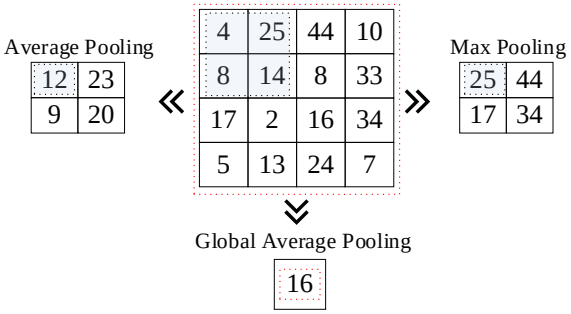

Fig. 3. Example of the average, maximum, and global average pooling layers.

The proposed models all have three major parts: the initialization layers; the parallel convolutional layers, which distinguish the DFU more effectively; and the global average pooling layer and the fully connected (FC) layers (plus the dropout layer between the fully connected layers to prevent overfitting). The last layer is the output classifier, which uses the softmax function. This function is employed to map the non-normalized output of the model to a probability distribution of the predicted output classes. It is also more robust to multi-class tasks, which is beneficial as although we only have two classes, we plan to utilize our model for different tasks to classify multiple classes.

We chose the proposed model of four branches to get in deep with its details and other models are changing according to the number of branches. Figure 4 presents the general structure of the proposed model of four branches, which has 27 convolutional layers and a total number of 100 layers (as described in Table 1 and Fig. 5). The number of convolutional layers in the other models is based on the number of branches, as shown in Fig. 5.

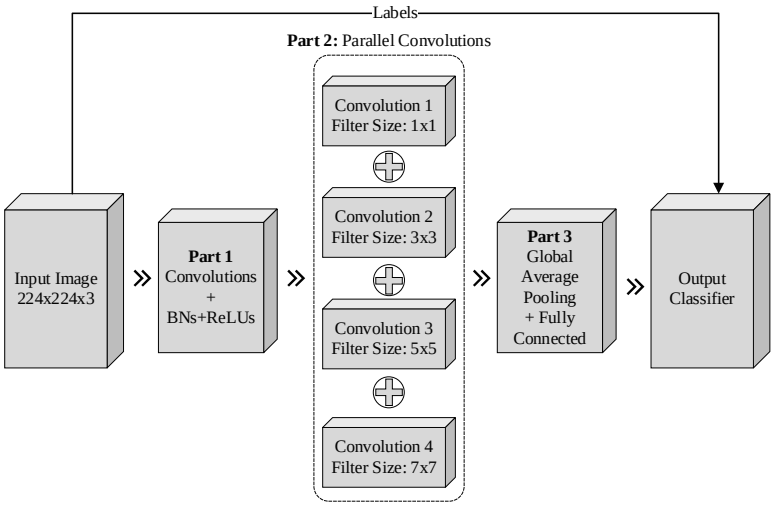

Fig. 4. The general structure of the proposed model with four branches. BN = batch normalization layer, ReLU = rectified linear unit.

- **Part 1** (see Fig. 5): This part is the same for all four proposed models. The size of the input patches was 224×224, and these were taken from regions of healthy skin and feet with DFU, as shown in Fig. 2. These patches form the images of the training dataset. This part consists of three convolutional layers with three different filter sizes of 3×3, 5×5, and 7×7. Each convolution layer is followed by a batch normalization layer and a rectified linear unit (ReLU) layer. This part is very important for ensuring that any large crude input images are decreased in dimensionality before the next part begins.
- **Part 2**: This part presents the parallel convolutions. A parallel convolutional filter is mainly a chain of multi-input convolutional filters that permit the extraction of multi-level features and envelop extra-wide clusters from similar input. In addition, the convolutions' design was weighted to generate differentiated features to highlight every DFU in the images. The six blocks of parallel convolutions in each model work in parallel, and the outputs are

then concatenated by the concatenation layer. The first model has two parallel convolution branches with two filter sizes of 1×1 and 3×3 at each block, as shown in Fig. 5C. The second model has three parallel convolutions with three filter sizes of 1×1, 3×3, and 5×5 at each block, as shown in Fig. 5B. The third model has four convolutional layers with four different filter sizes of 1×1, 3×3, 5×5, and 7×7 at each block, as shown in Fig. 5A.

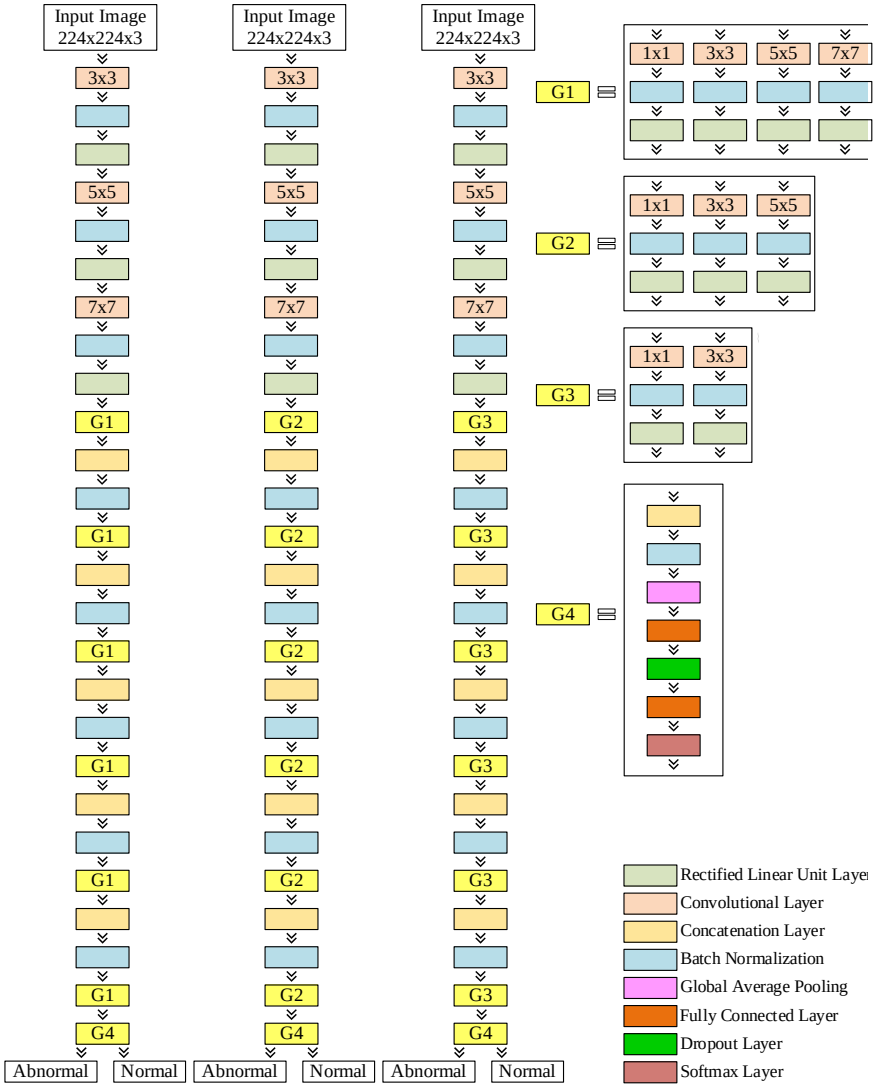

Fig. 5. The proposed model with (A) four branches, (B) three branches, and (C) two branches.

Table 1. Proposed model architecture with four branches. C = convolutional layer, PC = parallel convolutional layer, BN = batch normalization layer, ReLU = rectified linear unit, Drop = dropout layer, FC = fully connected layer

Name of layer	(FS) and (S)	Activations
Input layer	-	224 × 224 × 3
C1, BN1, ReLU1	FS=3×3, S = 1	224 × 224 × 16
C2, BN2, ReLU2	FS=5×5, S = 2	112 × 112 × 16
C3, BN3, ReLU3	FS=7×7, S = 2	56 × 56 × 16
PC1, BN4, ReLU4	FS=1×1, S = 1	56 × 56 × 16
PC2, BN5, ReLU5	FS=3×3, S = 1	56 × 56 × 16
PC3, BN6, ReLU6	FS=5×5, S = 1	56 × 56 × 16
PC4, BN7, ReLU7	FS=7×7, S = 1	56 × 56 × 16
Concatenation Layer1	Four inputs	56 × 56 × 64
CBN1	Batch Normalization	56 × 56 × 64
PC5, BN8, ReLU8	FS=1×1, S = 2	28 × 28 × 32
PC6, BN9, ReLU9	FS=3×3, S = 2	28 × 28 × 32
PC7, BN10, ReLU10	FS=5×5, S = 2	28 × 28 × 32
PC8, BN11, ReLU11	FS=7×7, S = 2	28 × 28 × 32
Concatenation Layer2	Four inputs	28 × 28 × 128
CBN2	Batch Normalization	28 × 28 × 128
PC9, BN12, ReLU12	FS=1×1, S = 1	28 × 28 × 32
PC10, BN13, ReLU13	FS=3×3, S = 1	28 × 28 × 32
PC11, BN14, ReLU14	FS=5×5, S = 1	28 × 28 × 32
PC12, BN15, ReLU15	FS=7×7, S = 1	28 × 28 × 32
Concatenation Layer3	Four inputs	28 × 28 × 128
CBN3	Batch Normalization	28 × 28 × 128
PC13, BN16, ReLU16	FS=1×1, S = 2	14 × 14 × 64
PC14, BN17, ReLU17	FS=3×3, S = 2	14 × 14 × 64
PC15, BN18, ReLU18	FS=5×5, S = 2	14 × 14 × 64
PC16, BN19, ReLU19	FS =7×7, S = 2	14 × 14 × 64
Concatenation Layer4	Four inputs	14 × 14 × 256
CBN4	Batch Normalization	14 × 14 × 256
PC17, BN20, ReLU20	FS=1×1, S = 1	14 × 14 × 128
PC18, BN21, ReLU21	FS=3×3, S = 1	14 × 14 × 128
PC19, BN22, ReLU22	FS=5×5, S = 1	14 × 14 × 128
PC20, BN23, ReLU23	FS=7×7, S = 1	14 × 14 × 128
Concatenation Layer5	Four inputs	14 × 14 × 512
CBN5	Batch Normalization	14 × 14 × 512
PC21, BN24, ReLU24	FS=1×1, S = 2	7 × 7 × 256
PC22, BN25, ReLU25	FS=3×3, S = 2	7 × 7 × 256
PC23, BN26, ReLU26	FS=5×5, S = 2	7 × 7 × 256
PC24, BN27, ReLU27	FS=7×7, S = 2	7 × 7 × 256
Concatenation Layer6	Four inputs	7 × 7 × 1024
CBN6	Batch Normalization	7 × 7 × 1024
Global average Pooling	-	1 × 1 × 1024
FC1	100 FC	1 × 1 × 100
Drop1	Dropout layer with learning rate:0.5	1 × 1 × 100
FC2	2 FC	1 × 1 × 2
Softmax	Abnormal, Normal	1 × 1 × 2

The question here is: "Does adding more branches improve the performance?" We added a fifth branch of the parallel convolutional layer. The fourth model has five branches of parallel convolutions with five different filter sizes of 1×1, 3×3, 5×5, 7×7, and 11×11 at each block, as shown in Fig. 6. After each concatenation layer, there is a batch normalization layer, which is very helpful for normalizing these activations behind any concatenation layer. It is also useful for preventing overfitting.

- **Part 3**: This part is the same for all four proposed models. It presents the global average pooling layer and the two FC layers, plus the output classifier based on softmax. The filter size of the global average pooling layer is 7×7. This layer is

followed by the two FC layers with 100 units for the first one and two units for the second one. A dropout layer was inserted between the two FC layers to prevent overfitting and to improve performance. We employed one dropout layer with a probability of p = 0:5. The last part is the softmax output of the class probabilities, which represents how close the training and proof data labels are to the parameters. The two classes represent the DFU outputs: normal (healthy) or abnormal (DFU).

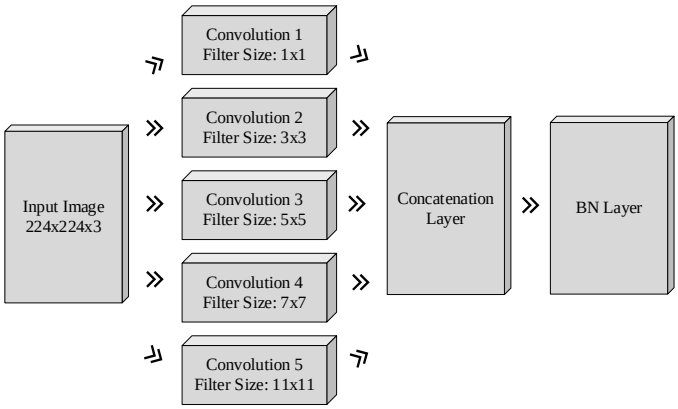

Fig. 6. The second part of the proposed model of five branches.

The training option used was stochastic gradient descent with momentum set to 0.9, mini-batch size set to 64, MaxEpochs set to 100, and a learning rate set to 0.001. The proposed models were trained with original images and original images plus augmented images. Figure 7 shows some filters from the first convolutional layer that learned abnormal skin features from the proposed model with four branches. Lastly, we implemented our experiments using MATLAB 2018 as software and an Intel (R) Core TM i7-5829K processor. The CPU was 3.30 GHz, the RAM was 32 GB, and the GPU was 8 GB.

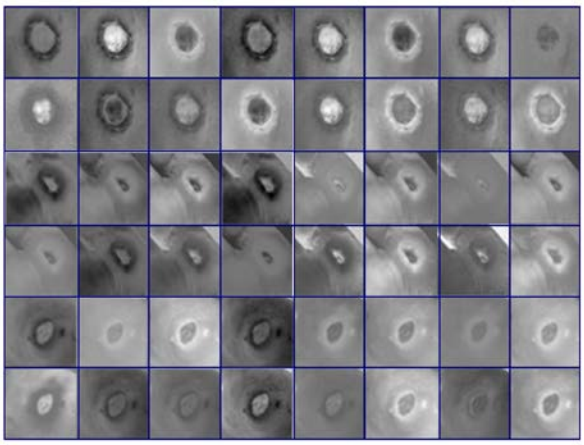

Fig. 7. Some learned filters from the first convolutional layer, which learned abnormal skin features from the proposed model with four branches.

4. Experimental Results

4.1. Evaluation metrics

The performance of the proposed models was evaluated in terms of recall, Eq.(1), precision, Eq. (2), and F1 score, Eq. (3):

$$\text{Recall} = TP/(TP+FN) \quad (1)$$

$$\text{Precision} = TP/(TP+FP) \quad (2)$$

$$F1_{score} = 2\times ((\text{Precision} \times \text{Recall})/(\text{Precision} + \text{Recall})) \quad (3)$$

Where TP represents true positives, FP represents false positives, and FN represents false negatives.

4.2. Results of training the models with original images

We first evaluated the proposed models trained with only original images, as reported in Fig. 8. The model with four branches of parallel convolutions achieved the highest evaluation measurements, scoring 90.8%, 87.9%, and 89.3% for precision, recall, and the F1 score, respectively. The model with five branches of parallel convolutions scored the second-highest evaluation measurements by achieving 89.5% for precision, 86.8% for recall, and 88.1% for the F1 score. The model with three branches of parallel convolutions came in third place by achieving 87.9% for precision, 84.8% for recall, and 86.2% for the F1 score. Lastly, the model with two branches of parallel convolutions achieved the lowest evaluation measurements by achieving 86.1% for precision, 84.5% for recall, and 85.2% for the F1 score.

In the case of DFU classification, the model with five branches did not show improved performance. Instead, the model with four branches achieved the highest performance measurements. The proposed model with five branches required more training images to achieve high performance, and in some cases, extracting more features led to model confusion when classifying images.

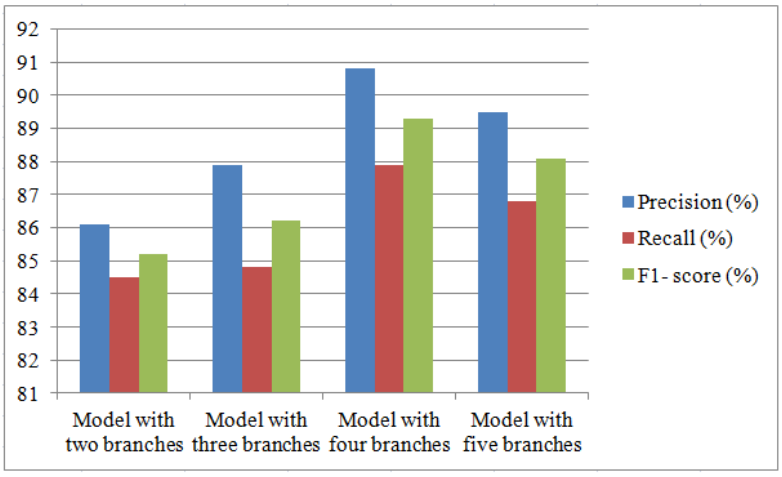

Fig. 8. Evaluation results for the proposed four models trained only with original images.

4.3. Results of training the models with original images plus augmented images

We then evaluated the proposed models trained with original images plus augmented images, as reported in Fig. 9. The model with four branches of parallel convolutions achieved the highest evaluation measurements by scoring 97.3%, 94.5%, and 95.8% for precision, recall, and the F1 score, respectively. The model with five branches of parallel convolutions scored the second-highest evaluation measurements by achieving 96.5% for precision, 94.2% for recall, and 95.3% for the F1 score. The model with three branches of parallel convolutions came in third by achieving 94.7% for precision, 92.9% for recall, and 93.7% for the F1 score. Lastly, the model with two branches of parallel convolutions achieved the lowest evaluation measurements by achieving 93.6% for precision, 90.7% for recall, and 92.1% for the F1 score. It is clear that the augmented images significantly enhanced performance.

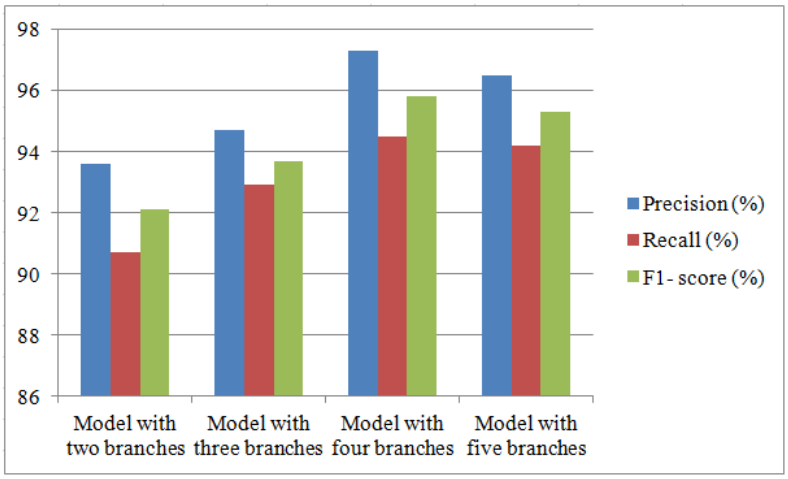

Fig. 9. Evaluation results for the proposed models trained with original images plus augmented images.

4.4. Performance comparison between our model and the latest state-of-the-art DFU classification methods

As the model with four branches of parallel convolutions achieved the highest results, we compared it to the latest state-of-the-art DFU classification methods, which are DFUNet [11] and DFU_QUTNet [12]. Our proposed models with four branches outperformed these models, as reported in Fig. 10. The predictions made by the four-branch model for some test images are shown in Fig. 11.

The proposed model with four branches has the sufficient number of layers and filters among other models, which assisted in having sufficient features to differentiate between classes. The proposed models with two and three branches required more extracted features to differentiate between classes, while the proposed model with five branches extracted more unbidden features that may have

confused the network. Overall, all models achieved high accuracy compared to prior methods.

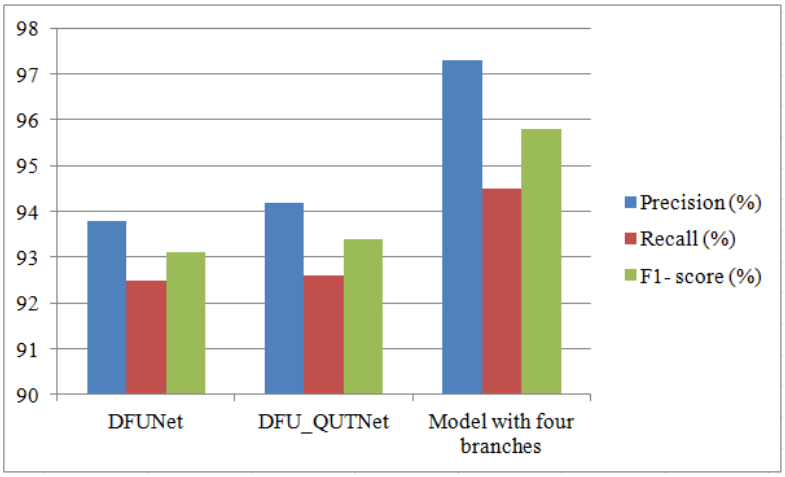

Fig. 10. Comparison of the proposed model with four branches to the latest CNN methods.

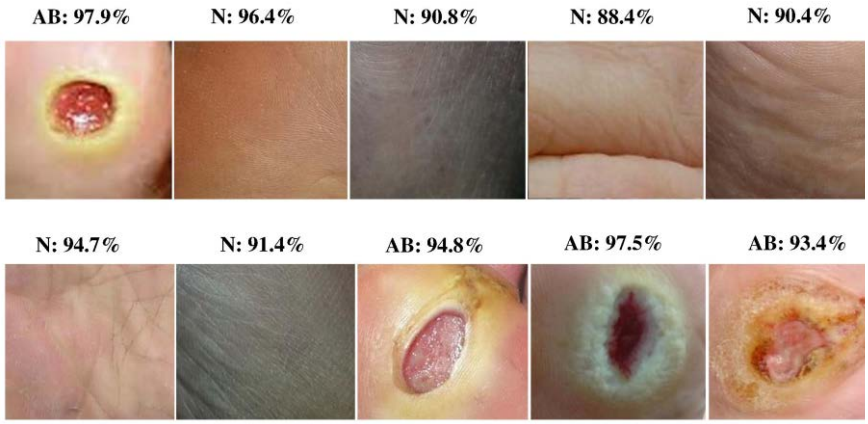

Fig. 11. Predictions made by the four-branch model for some test images. N = normal class, AB = abnormal class (DFU).

5. Conclusion

The comparison of four hybrid CNN models has been presented for automated classification of DFU into two classes: normal (healthy skin) and abnormal (DFU). The architecture of the proposed models was designed by integrating traditional convolutional layers with multi-branch parallel convolutional layers ranging from two to five branches. This type of structure has the ability to extract different features to differentiate between classes. We trained the proposed models with original images first, then trained them with original images plus augmented images. We showed that the model with four branches of parallel convolutional layers performed better than the models with five, three, or two branches in terms

of DFU classification. This indicates that increasing the width of the network does not always improve performance. The proposed model with four branches was effective and outperformed the latest DFU classification methods by achieving an F1 score of 95.8%. We aim to fine-tune the proposed model with four branches to classify five classes of DFU levels. We also plan to utilize the knowledge that the proposed model learned from the DFU dataset as transfer learning to classify several wound types.

References

1. Lefrancois, T.; Mehta, K.; Sullivan, V.; Lin, S.; and Glazebrook, M. (2017). An evidence-based review of literature on detriments to the healing of diabetic foot ulcers. *Foot and Ankle Surgery*, 23(4), 215-224.
2. Ogurtsova, K.; da Rocha Fernandes, J.D. ; Huang, Y., Linnenkamp, U. ;Guariguata, L.; Cho, N.H.; Cavan, D.; Shaw, J.E.; and Makaroff, L.E. (2017). IDF Diabetes Atlas: Global estimates for the prevalence of diabetes for 2015 and 2040. *Diabetes research and clinical practice*, 128, 40-50.
3. Ramsey, S.D.; Newton, K.; Blough, D.; McCulloch, D.K.; Sandhu, N.; Reiber, G.E.; and Wagner, E.H. (1999). Incidence, outcomes, and cost of foot ulcers in patients with diabetes. *Diabetes care*, 22(3), 382-387.
4. Leightley, D.; McPhee, J.S.; and Yap, M.H. (2016). Automated analysis and quantification of human mobility using a depth sensor. *IEEE journal of biomedical and health informatics*, 21(4), 939-948.
5. Litjens, G.; Kooi, T.; Bejnordi, B.E.; Setio, A.A.A.; Ciompi, F.; Ghafoorian, M.; Van Der Laak, J.A.; Van Ginneken, B.; and Sánchez, C.I. (2017). A survey on deep learning in medical image analysis. *Medical image analysis*, 42, 60-88.
6. Goyal, M.; Reeves, N.D.; Rajbhandari, S.; and Yap, M.H. (2018). Robust methods for real-time diabetic foot ulcer detection and localization on mobile devices. *IEEE journal of biomedical and health informatics*, 23(4), 1730-1741.
7. Yu, Z.; Jiang, X.; Zhou, F.; Qin, J.; Ni, D.; Chen, S.; Lei, B.; and Wang, T. (2018). Melanoma recognition in dermoscopy images via aggregated deep convolutional features. *IEEE Transactions on Biomedical Engineering*, 66(4), 1006-1016.
8. Dargaville, T.R.; Farrugia, B.L.; Broadbent, J.A.; Pace, S.; Upton, Z.; and Voelcker, N.H. (2013). Sensors and imaging for wound healing: a review. *Biosensors and Bioelectronics*, 41, 30-42.
9. Babu, K.S.; Subudhi, A.; and Sabut, S. (2018). Segmentation of diabetic wound by multidimensional clustering for quantitative assessment of healing process. *Current Medical Imaging*, 14(1), 71-76.
10. Wang, L.; Pedersen, P.C.; Agu, E.; Strong, D.M.; and Tulu, B. (2016). Area determination of diabetic foot ulcer images using a cascaded two-stage SVM-based classification. *IEEE Transactions on Biomedical Engineering*, 64(9), 2098-2109.
11. Goyal, M.; Reeves, N.D.; Davison, A.K.; Rajbhandari, S.; Spragg, J.; and Yap, M.H. (2018). DFUNet: Convolutional neural networks for diabetic foot ulcer classification. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 4(5), 728 - 739.

12. Alzubaidi, L.; Fadhel, M. A.; Oleiwi; S. R.; Al-Shamma, O.; and Zhang, J. (2020). DFU_QUTNet: diabetic foot ulcer classification using novel deep convolutional neural network. Multimedia Tools Applications, 79, 15655-15677.
13. Wang, C.; Yan, X.; Smith, M.; Kochhar, K.; Rubin, M.; Warren, S.M.; Wrobel, J.; and Lee, H. (2015). A unified framework for automatic wound segmentation and analysis with deep convolutional neural networks. *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC.* Milano, Italy, 2415-2418.
14. van Netten, J. J.; van Baal, J. G.; Liu, C.; van Der Heijden, F.; and Bus, S.A. (2013). Infrared thermal imaging for automated detection of diabetic foot complications. *Journal of diabetes science and technology*. 7(5), 1122-1129.
15. Alzubaidi, L.; Zhang, J.; Humaidi, A.J.; Al-Dujaili, A.; Duan, Y.; Al-Shamma, O.; Santamaría, J.; Fadhel, M.A.; Al-Amidie, M.; Farhan, L. (2021). Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *Journal of big Data*, 8, Article number: 53 (2021) .
16. Cao, C.; Liu, F.; Tan, H.; Song, D.; Shu, W.; Li, W.; Zhou, Y.; Bo, X.; and Xie, Z. (2018). Deep learning and its applications in biomedicine. *Genomics, proteomics and bioinformatics*, 16(1), 17-32.
17. Minsky, M. L.; and Papert, S.A. (2017). *Perceptrons: An introduction to computational geometry*. MIT press.
18. Hinton, G.E.; Osindero, S.; and Teh, Y.W. (2006). A fast learning algorithm for deep belief nets. *Neural Computation* (2006). 18(7), 1527-1554.
19. Alzubaidi, L.; Fadhel, M. A.; Al-Shamma, O.; Zhang, J.; Santamaría, J.; and Duan, Y. (2021). Robust application of new deep learning tools: An experimental study in medical imaging. *Multimedia Tools and Applications*, 1-29.
20. Alzubaidi, L.; Al-Shamma, O.; Fadhel, M.A.; Farhan, L.; Zhang, J.; and Duan, Y. (2020). Optimizing the Performance of Breast Cancer Classification by Employing the Same Domain Transfer Learning from Hybrid Deep Convolutional Neural Network Model. *Electronics*, 9(3), 445.
21. Alom, M.Z.; Taha, T.M.; Yakopcic, C.; Westberg, S.; Sidike, P.; Nasrin, M.S.; Hasan, M.; Van Essen, B.C.; Awwal, A.A.; and Asari, V.K. (2019). A state-of-the-art survey on deep learning theory and architectures. *Electronics*, 8(3), 292.
22. Oord, A.V.D.; Kalchbrenner, N.; and Kavukcuoglu, K. (2016). Pixel recurrent neural networks. *Proceedings of Machine Learning Research*. 48, 1747-1756.
23. Alzubaidi, L.; Al-Amidie, M.; Al-Asadi, A.; Humaidi, A.J.; Al-Shamma, O.;Fadhel, M.A.; Zhang, J.; Santamaría, J.; and Duan, Y. (2021). Novel transfer learning approach for medical imaging with limited labeled data. *Cancers*, 13(7), 1590.
24. Alzubaidi, L.; Fadhel, M.A.; Al-Shamma, O.; Zhang, J.; and Duan, Y. (2020). Deep Learning Models for Classification of Red Blood Cells in Microscopy Images to Aid in Sickle Cell Anemia Diagnosis. *Electronics*, 9(3), 427.
25. Hasan, R.I.; Yusuf, S.M.; and Alzubaidi, L. (2020). Review of the State of the Art of Deep Learning for Plant Diseases: A Broad Analysis and Discussion. *Plants*, 9(10), 1302.

26. Krizhevsky, A.; Sutskever, I.; and Hinton, G.E. (2012). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6), 84-90.
27. Zeiler, M.D.; and Fergus, R. (2014). Visualizing and understanding convolutional networks. *13th European Conference on computer vision*. Zurich, Switzerland, 818-833.
28. Simonyan, K.; and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. eprint arXiv, 1409-1556.
29. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; and Rabinovich, A. (2015). Going deeper with convolutions. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Boston, MA, USA, 1-9.
30. He, K.; Zhang, X.; Ren, S.; and Sun, J. (2016). Deep residual learning for image recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas, NV, USA, 770-778.
31. Huang, G.; Liu, Z.; Van Der Maaten, L.; and Weinberger, K.Q. (2017). Densely connected convolutional networks. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Honolulu, HI, USA, 2261-2269.
32. Raghu, M.; Zhang, C.; Kleinberg, J.; and Bengio, S. (2019). Transfusion: Understanding transfer learning for medical imaging. *33rd Conference on Neural Information Processing Systems (NeurIPS 2019)*. Vancouver, Canada, 1-11. .
33. Alzubaidi, L.; Fadhel, M.A.; Al-Shamma, O.; Zhang, J.; Santamaría, J. M.; Duan, Y.; and Oleiwi, S. R. (2020). Towards a Better Understanding of Transfer Learning for Medical Imaging: A Case Study. *Applied Sciences*, 10 (13), 4523.
34. Lin, M.; Chen, Q.; and Yan, S. (2013). Network in network. eprint arXiv, 1312-4400.