International Journal of Electrical and Computer Engineering (IJECE)
Vol. 12, No. 3, June 2022, pp. 3072~3082
ISSN: 2088-8708, DOI: 10.11591/ijece.v12i3.pp3072-3082

Six skin diseases classification using deep convolutional neural network

Ramzi Saifan, Fahed Jubair
Department of Computer Engineering, University of Jordan, Amman, Jordan

Article Info

Article history:

Received Oct 17, 2020
Revised Dec 21, 2021
Accepted Jan 19, 2022

Keywords:

Convolutional neural networks
Deep learning
Machine learning
Medical image analysis
Skin diseases

ABSTRACT

Smart imaging-based medical classification systems help the human diagnose the diseases and make better decisions about patient health. Recently, computer-aided classification of skin diseases has been a popular research area due to its importance in the early detection of skin diseases. This paper presents at its core, a system that exploits convolutional neural networks to classify color images of skin lesions. It relies on a pre-trained deep convolutional neural network to classify between six skin diseases: acne, athlete's foot, chickenpox, eczema, skin cancer, and vitiligo. Additionally, we constructed a dataset of 3000 colored images from several online datasets and the Internet. Experimental results are encouraging, where the proposed model achieved an accuracy of 81.75%, which is higher than the state of the art researches in this field. This accuracy was calculated using the holdout method, where 90% of the images were used for training, and 10% of the images were used for out-of-sample accuracy testing.

Corresponding Author:

Ramzi Saifan
Department of Computer Engineering, University of Jordan
Queen Rania Street, Amman, Jordan
Email: r.saifan@ju.edu.jo

1. INTRODUCTION

Over the past years, the risk on life due to skin diseases has increased since some of them suddenly appear on the skin. Skin diseases are considered among the most spread diseases globally, affecting more than 900 million people worldwide. Additionally, every year, about 18% of the total population is affected by malignant growths on their skin. Skin diseases are ranked as the fourth most common cause of human illness. However, many affected people do not consult a physician [1]. Additionally, humans commonly assume that most skin diseases are not fatal, and therefore they apply some traditional treatments instead of consulting a certified dermatologist. However, if these treatments are not suitable for that skin problem, they make it worse. As an example, from the authors' local community, the Jordanian Medical Syndicate statistics show that the number of dermatologists in Jordan is small compared to the spread of skin diseases in Jordan, as shown in Table 1. The global situation is not much better, which has similar statistics.

Table 1. Physicians registered at the syndicate 2016 [2]

Specialization	Public Sector		Private Sector		Others		Total	
	Male	Female	Male	Female	Male	Female	Male	Female
Dermatologist Venereal	73	18	98	35	19	17	190	70
Total physicians including Other Specializations	5829	1025	4386	820	7223	2025	17438	3900

The number of dermatologists in Jordan, as shown in Table 1, is a small figure compared to other specializations, ranging from 1.3% for the public sector to 2.5% for the private sector, 0.4% for other sectors, and a total of 1.2% for all other specializations. Additionally, booking an appointment in a hospital with an accredited dermatologist can be disappointing. Many dermatologists are booked for weeks, some maybe even for months. An average American dermatology appointment wait is about 32 days [3]. Moreover, some hospitals require a referral from a primary care physician. In contrast, the Jordanian Ministry of Health referred to in its statistics about the School Health Services program the statistics in Table 2.

Table 2. School health services [2]

Particulars	2012/2013	2013/2014	2014/2015	2015/2016
Number of Tested Students	399278	422388	413646	410308
Number of Tested Schools	3366	3605	3863	3839
Number of Discovered Diseases	27393	28587	29244	31452
Skin Diseases Discovered	9409	8935	8323	7029

As shown, the percentage of students who have skin disease ranges between 34% in 2012/2013 and 31% in 2013/2014, while it decreased to 28% in 2014/2015 and 22% in 2015/2016. This is approximately one-third of the sick students. These statistics are for school students, and the number of skin patients is more significant for public people. According to the reasons that were described above, a smart computerized medical imaging-based diagnosis system for skin diseases will be helpful and a welcome system. Typically, skin diseases' traditional diagnosis includes long medical tests to determine the disease's type correctly [4]. Also, due to the difficulty and subjectivity of the human diagnosis and interpretation, and since there are various lesions and other factors encountered in practice, computerized analysis of medical images has become an important research area to support the diagnosis. In this paper, we are explicitly interested in a six-class classification problem: determine which skin disease the picture contains among acne, athlete's foot, chickenpox, eczema, skin cancer, or vitiligo.

This paper presents a novel method that uses a deep learning-based application to classify six globally widespread skin diseases. Additionally, the method is implemented and benchmarked against a self-built dataset collected from several online datasets, where high classification accuracy was achieved. The system will serve various society segments by early diagnosis of these diseases to decrease their spread, which will reduce effort and time. For example, the system can help medical students specializing in dermatology to compare their classification results of patients' skin diseases with the application's results from the deep learning model. Moreover, dermatologists can diagnose skin diseases using the proposed application. Due to many patients and time constraints, the proposed system helps them since they have no human bias. Also, patients can early detect their skin disease.

Skin is the largest organ in the human body. Adults have an average of 22 square feet of skin. Skin accounts for about 15% of the bodyweight. Accordingly, skin is subject to several diseases. The following is a brief description about each of the studied six diseases:

a) Acne: is one of the most common skin diseases in the US. It affects about 50 million people every year; it increases in adults and affects up to 15% of Females [5]. Some of Acne conditions: whiteheads, blackheads, small red and tender bumps, pimples, painful lumps beneath the surface of the skin, and cystic lesions [6]. It affects about 85% of people between 12 and 24 [5]. It usually appears on the face, back, shoulders, and forehead. It affects areas of the skin with a high number of oil glands.
b) Chickenpox is a traditional childhood disease; the highest propagation of it is in the 4 to 10-year-old. In 2013, there were 140 million patients of chickenpox around the world [7]. Some symptoms: body aches, feeling tired, feeling irritable, and headache. This disease results in a characteristic skin rash that forms itchy and small blisters. It starts on the face and the back, then it spreads to the rest of the body later on [8]. This disease is also known as varicella.
c) Athlete's foot: around 15% to 25% of people are affected by the athlete's foot. The athlete's foot is a common skin disease that appears on the feet. This disease can spread to other parts of the body and other people as well [9]. Some symptoms: itching, blisters, dry skin, raw skin, discolored, thick, and crumbly toenails. The same fungus may also affect the hands.
d) Eczema affects about 35 million people in the United States (US), 1% to 3% of adults, and 10% to 20% of children. About 60% of babies who have eczema have some symptoms of it in puberty [10]. Some symptoms: sensitive skin, inflamed skin, itching, dark-colored patches, scaly patches, and crusting. The appearance of skin affected by atopic dermatitis will depend on how much a person scratches and whether it is infected. Scratching increases make itchiness worse.

e) Vitiligo: it can be noticed more in dark-skinned people since it makes a loss in the skin tone. Some symptoms: white patches and a change in color of skin. The exact cause of vitiligo is unknown; patients lose pigment on many of their skin parts. After the patches appear, they may stay the same, but they might get more significant later. Patients may have cycles of pigment loss and stability [11].
f) Skin cancer is an abnormal growth of skin cells. it is the most common cancer in the US. More than one million people in America are living now with skin cancer. Every day there is about 9.5K people in the US are diagnosed with skin cancer [12]. Skin cancer occurs when errors occur in the deoxyribonucleic acid (DNA) of skin cells.

Using machine learning techniques in diagnosing, studying, and analyzing different medical diseases is a hot research area, especially as the accuracy of such techniques is increasing, and because it avoids human subjective bias. For example, Mahmood *et al.* [13] introduced a survey on neural network techniques to classify lymph, neck, head, and breast cancer. Also, Karim *et al.* [14] compared eight neural network training algorithms for the classification of heart disease data. Hijazi *et al.* [15] presents a deep ensemble learning for tuberculosis detection using chest x-ray and Canny edge detected images. Bakshi and Sathya [16] uses the adaptive cascading technique in the detection of acne skin disease. In [17], a disease diagnosis system is designed based on the internet of things (IoT).

There exist well and prepared datasets for skin cancer, such as the international skin imaging collaboration (ISIC) archive dataset [18] and the HAMl0000 dataset [19]. Much prior research is based on these datasets as shown in Table 3. For example, Nugroho *et al.* [20] used a convolutional neural network (CNN) for identification. CNN works through three stages: convolutional layer, pooling layer, and fully-connected layer. They use the HAM10000 skin cancer dataset. The accuracy of training and testing the skin cancer identification system is 80% and 78%, respectively [20]. Lopez *et al.* [21] uses the VGGNet CNN model and transfer learning on the ISIC dataset with 78.66% sensitivity value.

Table 3. Summary of related datasets

Dataset	Used Model	Accuracy
HAM l0000 [4]	CNN	78%
ISIC Archive [22]	FT VGG-16 CNN	70% (Sensitivity)
ISIC Archive [18]	VGGNet CNN	78.66%
HAM l0000 [4]	SVM	91%

Another paper was introduced by Kalouche [23]. It uses ISIC; they got 70% accuracy for classifying skin melanoma and a 78% using a fine-tuned VGG-16 CNN. The work in [24] paper uses an support vector machine (SVM) classifier to differentiate 172 Dermatoscopic images into two classes as “benign” and “malignant”, and they have 91% accuracy. This paper included 500 images of melanoma skin cancer as a category of our six classes from the ISIC archive dataset. In the paper [25], the authors trained the CNN architecture using 23K images from the Derm.Net dataset and tested it on both Derm.Net and OLE datasets. They get 73.1% Top-1 accuracy and 91% Top-5 accuracy for Derm.Net dataset testing, and Top-1 and Top-5 accuracies are 31.1% and 69.5% for OLE dataset testing.

In 2018, Hameed *et al.* [26] had used a hybrid approach, i.e., using deep convolution neural networks and error-correcting output coding (ECOC) SVM. They classify five categories: eczema, healthy, benign, acne, and malignant melanoma, using 9,144 images collected from different sources. The accuracy is 86.21%. Patnaik *et al.* [4] predict the several kinds of skin diseases using techniques of deep learning. The paper exploits three architectures of image recognition. The used models are can classify up to 1,000 classes of the images such as panda and parrot.

Most research on skin diseases classifications that use machine learning techniques focused on using one of the models: i) SVM [27], ii) trees [28], iii) K-nearest neighbor (KNN) [29], and iv) ensemble classifiers and CNN [30]. For example, in work [27] the authors obtained a set of features using CNN, and then they classified them into four classes using SVM classifier and using a dataset of 3753 images. The achieved accuracy is 94.2%. However, their work was for skin cancer classification only.

Esteva *et al.* [31] presented a study with a dataset of 129,450 images for skin cancer lesions. Moreover, they compared their classification with the diagnosis provided by twenty-one dermatologists. They achieved a good accuracy on a large dataset. But their system still works on only on the forms of skin cancer. De Guzman *et al*. [32] used a system of single-layer and multi-layer to detect eczema. The single-layer can do only binary classification (i.e., eczema or non-eczema). However, they used the multi-model to classify the images into three types: spotted, scattered, and dried eczema. Using artificial neural network (ANN), they reached accuracy 85.71% to 96.03% in the single-layer, and they reached accuracy 87.30% to

92.46% in the multi-layer. Moreover, their system can be used in multi-class classification, but for one skin lesion only.

Due to the limitations discussed, there is a need for an intelligent expert system that can perform the multi-class classification of a different range of skin diseases. In our system, we propose an automated Android mobile application that allows a live image capturing and do six skin diseases classification. The patient can take a photo and then get the result of the disease, which will be identified using the Keras CNN model with 81.75% accuracy. Our CNN model comprises four convolutional layers, four pooling layers, 4 dropout layers, and two fully connected layers. This paper is structured as follows: section 2 describes the proposed solution and associated methods and tools, section 3 presents the experiments' results and discusses their significance, and section 4 offers concluding remarks and directions for future work.

2. METHODS AND TOOLS

The suggested solution is a system that categorizes images of six skin diseases. It uses deep learning. In this section, the dataset and the proposed model will be discussed. Additionally, we will discuss briefly the designed mobile application and some implementation aspects.

2.1. Data preparation

Data preparation is the first step of the classification process. Data have different forms: images, data as "string" listed in a comma-separated values (CSV) file, JavaScript Object Notation (JSON) file, extensible markup language (XML) files into a tabular form, and more. Since we are going to deploy our model using Keras, we used images. The main steps for data preparation are:

a) Data collection: in our case, we have created our dataset by collecting about 3 K images for six of the prevalent skin diseases. The images were basically from Google images, Derma.Net, and ISIC [18] archive dataset. Derm.Net is one of the largest photo sources for skin diseases, which is available online. It has more than 23K images for skin diseases, while ISIC is for skin cancer data. Additionally, the data were accredited by a certified dermatologist. We have balanced the data categories, so each class has 500 images.
b) Profiling and exploration: once the data have been collected, we have to validate it. This can be done by a specialist who is typically a dermatologist, not an engineer or a programmer. Since the data is medical data, a medical field specialist must ensure its quality, convenience, and suitability. This specialist can eliminate inconsistent, inappropriate, missing, skewed, irrelevant data, or even data that suffer significant deviation. This step handles any issues that could give us an incorrect model's findings later on [22].
c) Formatting: the next step is going to ensure the data is formatted to fit the model. Anomalies will be discovered if the data is aggregated from different sources or if more than one stakeholder has manually updated it. A consistent data format takes away any errors, so the entire dataset uses the same input protocols [22].
d) Improving quality: this is the starting point to deal with erroneous data, missing and extreme values. It uses histograms to show the data distribution and examine the images outside the acceptance range. It does not delete all images with a missing value since many deletions can skew the dataset [22].
e) Feature engineering: this step compromises image pixels' transformation to features that represent a learning algorithm pattern. Segregating some data may provide the algorithm with more relevant information [22].
f) Splitting: the final step is to split the dataset into two sets, mainly; the first set for training the algorithm and the second one for validation. They have to be non-overlapping subsets of the primary dataset to ensure proper testing [22].

2.2. Model implementation

We used the Keras library to implement our CNN. Specifically, models in Keras can be implemented in two forms: sequential models or functional application programming interface (API) models. In this paper, we used the sequential model. In this model, the typical CNN Layers, as shown in Figure 1, are input layer ➔ convolution layer ➔ pooling layer ➔ flattening layer ➔ dense/output layer.

2.3. Mobile application (Skinvy)

To increase the proposed system's usability, we made it accessible through an easy to use Android mobile application (called Skinvy). The application allows anybody to capture an image of the skin infected by one of the six diseases. Then the application will classify the image to one of these six skin diseases. The interface of Skinvy is as shown in Figure 2. The user can choose to: i) take a picture of the disease and classify it and ii) to show the supported diseases list.

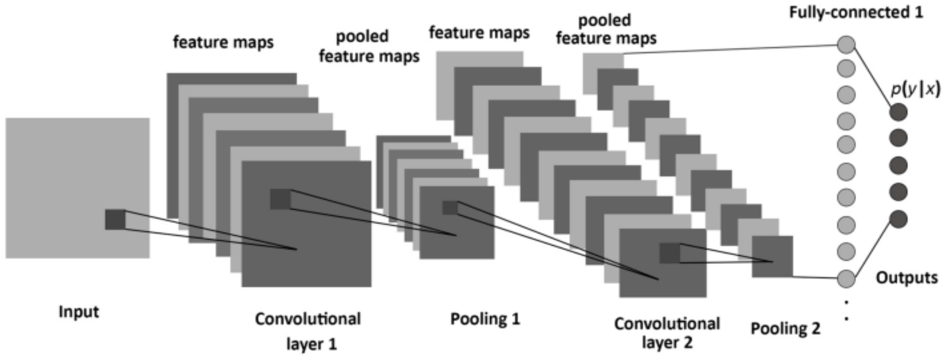

Figure 1. Typical block diagram of CNN [33]

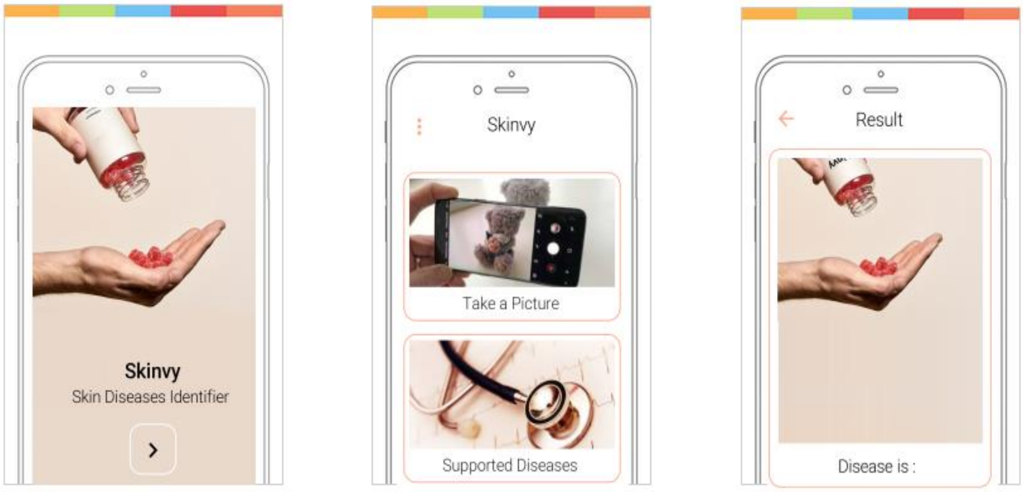

Figure 2. Skinvy app

We tested Skinvy application on many pictures of skin diseases by different people. It was straightforward to use and gave real-time results in less than a second. Such an application is supposed to give more access to the proposed system. Skinvy application allows a live image capturing. The patient can take a photo and then get the result of the disease, which will be identified using the Keras CNN model with 81.75% accuracy. The accuracy of the model is expected to increase with time as the application allows us to get more images. The advantages of having Skinvy application is that it allows us to collect more pictures of infected skin. This is because the application can be uploaded to the "Play store" of Android applications, and anybody around the world can get access to the application. Then, patients take pictures for their infected skin, and get an initial diagnosis. This way a bigger dataset can be built which may help in enhancing the accuracy. Also, it reduces the cost of going to the dermatologist and the patients may get an initial diagnosis faster.

2.4. Implementation aspects

Keras is a deep learning framework for Python, was utilized to implement the neural network architecture. Keras layers are the fundamental building blocks of any Keras model. Layers are created using various layer functions and are typically composed together by stacking calls to them.

a) Conv2D: it is a two-dimensional convolution layer; it creates a convolution kernel that is convolved with the input to supply a tensor of outputs.
b) MaxPooling2D: it reduces the size of data; it joins the outputs of neuron clusters at one layer into one neuron in the next layer. It combines small clusters, which are commonly 2×2. Pooling can compute a maximum or average. Max pooling uses the maximum value from each cluster of neurons at the previous layer [34].

c) Dropout: during model training, a specific percentage of neurons on a given layer will be deactivated. This is expected to improve the generalization process. It helps in reducing overfitting. At each training stage, individual neurons are either neglected from the net with a probability of 1-p or kept in the net with probability p. This p is small for input neurons because it is directly lost when input nodes are neglected [34].
d) Activation: it applies a specific activation function to the output. Rectified linear activation unit (ReLU) is an example of an activation function. It removes negative values from an activation map by setting them to zero. Other functions are also used to increase non-linearity, such as the saturating hyperbolic tangent and sigmoid. ReLU is often preferred because it trains the neural network several times faster [34].
e) Flatten: it removes all of the dimensions except one dimension, and it reshapes the tensor to have a shape that is equal to the number of elements contained in it. It is the same as making a one-dimensional array. It flattens the pooled feature map into a column since we need to insert this data into an artificial neural network later on. It ends up with a long vector of input data that will be processed further.
f) Dense: is a fully connected layer that connects all neurons in a layer to all neurons in another layer. The flattened matrix goes through this fully connected layer to specify the categories of the model. After several convolutional and max-pooling layers, the neural network's high-level reasoning is done via fully connected layers.

3. RESULTS AND DISCUSSION

This section discusses our model's results from tuning the hyperparameters on the same dataset to achieve our final model with 81.75% accuracy. We start by listing the set of parameters to be tuned in Keras. Then, we show how we choose the values for each parameter.

3.1. Keras model tuning

Tuning hyperparameters for the deep neural network is difficult and slow. Besides, there are many parameters to tune and configure, including:

a) Validation split: it determines the ratio of the validation set.
b) Learning rate: it has a small positive value; it ranges between 0.0 to 1.0. A too-large value can cause a too quick converge to the model, which leads to a sub-optimal solution. On the other hand, a too-small value can cause stuck.
c) The number of hidden layers and units (number of CNN layers): usually, it is good to add more layers until no improvements. The tradeoff is that it is computationally expensive to train the network. Having a small number of units may lead to underfitting while having more units are usually not harmful to appropriate regularization.
d) Output filter of the convolution: it determines the number of output filters in the convolution. The default filters used by Keras are 3x3 or 5x5.
e) Activation function: the AF of a node specifies the node's output given one input or set of inputs. It introduces non-linearity to the model. The alternatives of it can be ReLU or tanh.
f) Optimizer: it is one of two arguments required for compiling the Keras model. The alternatives of it can be stochastic gradient descent (SGD), Root Mean Square Propagation (RMSprop), Adagrad, Adadelta, Adamax, and Adam.
g) File size (image size): it determines the dimensions of the image.
h) Dropout rate value: it is a regularization technique to avoid overfitting in the deep neural networks. It merely drops out units in the neural network according to specific probability. A default value of 0.1 to 0.5 is useful to test with.
i) Batch size: mini batch is usually better in the learning process of the model. A range of 16 to 128 is useful to test with.
j) Max-Pooling size: it is an integer or two integers. If only one integer is determined, the same length will be used for the second dimension. It reduces the input's dimensions and allows for a supposition to be made about the features.
k) Kernel Size: it specifies the length of the 1-D convolution window.
l) Number of Epochs: it is the number of times the training set passes through the neural network. Usually, the process will increase the number of epochs until noticing a small gap between the test loss and the training loss. If the Early Stopping technique is used to overcome the overfitting, then the number of epochs will be assigned to a large number, and it automatically stops at the best epoch.
m) Padding: it can be either valid or the same.

3.2. Parameters tuning and achieved results

a) Validation split tuning: based on the results in Table 4 for different validation split ratios, the validation split ratio that we will use for the remaining results is 0.10 because it achieved the highest accuracy=0.7737. The value of validation split defines the probability for each picture to be chosen for training or testing. As the probability decreases, the accuracy increases. This means that using the given dataset and using the given parameters, the system prefers more images for training and less for testing. This may happen on small datasets. It is expected that if the dataset size increased, this value might increase up to 0.30.

Table 4. Results for validation split tuning

Validation Split	Test Accuracy
0.10	0.7737
0.15	0.7640
0.20	0.7591
0.25	0.7314

b) Learning rate tuning: this parameter controls the estimated error response each time the model weights are updated. Based on the results in Table 5, we will use the learning rate for the next tunings: 0.001. Learning rate is usually selected small (i.e., maximum is 1), which assures that the over-fitting and under-fitting in the results are well controlled.

Table 5. Results for learning rate tuning

Learning Rate	Test Accuracy
0.001	0.7810
0.0001	0.6971

c) Number of layers tuning: the number of convolutional layers determines the depth of the model. By increasing it, accuracy gets saturated. In our model, we get the best accuracy with four layers. For each convolution layer, there is an activation and max-pooling. Based on the results in Table 6, we will use four layers. Increasing number of hidden layers in the given model increases the test accuracy. That is expected in deep convolutional neural networks. This is also on the cost of delay in getting the results. However, there is usually a saturation level, after which the accuracy will not get improved by increasing the number of hidden layers. In our case, we arrived at this level by four hidden layers.

Table 6. Results for layers number tuning

Layers Number	Test Accuracy
1	0.7372
2	0.7664
3	0.7628
4	0.7847

d) Output filter: the output filter determines the number of output channels of a convolutional layer. Based on the results in Table 7, we will use the first line for the output filter.

Table 7. Results for output filter tuning

First Conv. Layer	Second Conv. Layer	Third Conv. Layer	Fourth Conv. Layer	Test Accuracy
32	64	128	512	0.7847
64	128	128	512	0.7774
128	256	512	1024	0.7299
64	128	256	1024	0.7591

e) Activation function tuning: as shown in Table 8, ReLU is the most commonly used activation function in neural networks, especially in CNNs.

Table 8. Results for activation function tuning

Activation Function	Test Accuracy
ReLU	0.7847
Tanh	0.7007

f) Optimizer tuning: based on the results in Table 9, Adam will be used.

Table 9. Results for optimizer tuning

Optimizer	Test Accuracy
Adam	0.7847
RMSprop	0.7664

g) Image size tuning: Image size is used for resizing the image before appending it to the training data and for reshaping the image for the feature extraction process. Based on the results in Table 10, the image size value that will be used is sixty.

Table 10. Results for image size tuning

Image size	Test Accuracy
40	0.7263
50	0.7737
60	0.7920
70	0.7372

h) Dropout rate tuning: Dropout is to minimize overfitting in the model to generalize the model. According to Table 11, we will use 0.20 between convolutional layers and 0.50 after denes.

Table 11. Results for dropout rate tuning

Dropout Rate between Conv.	Dropout Rate After Denes	Test Accuracy
0.20	0.50	0.7993
0.20	0.30	0.7847
0.25	0.40	0.7336
0.40	0.45	0.7628
0.50	0.50	0.6325

i) Batch size tuning: Batch size defines the number of inputs propagated through the neural network. We will use a batch size value of 64 based on the results in Table 12.

Table 12. Results for batch size tuning

Batch Size	Test Accuracy
128	0.7628
64	0.8175
32	0.7591
16	0.7628

In this paper, we used the following values for the hyperparameters of the deep neural network: i) validation split=0.10, ii) learning rate=0.001, iii) number of layers=4, iv) output filter=32, 64, 128, 512, v) activation function=ReLU, vi) optimizer=Adam, vii) image size=60, viii) dropout rate=0.50, ix) kernel size=3x3, x) max-pooling=2x2, xi) batch size=64, and xii) epochs number is set to 1000 since the early stopping technique is used.

Accordingly, tuning ended with a test accuracy of 81.75%. Figure 3 shows the training and validation leaning curves for both accuracy and loss. It is clear from both figures that there is no over fitting nor under fitting. This is because the validation accuracy and training accuracy are close to each other.

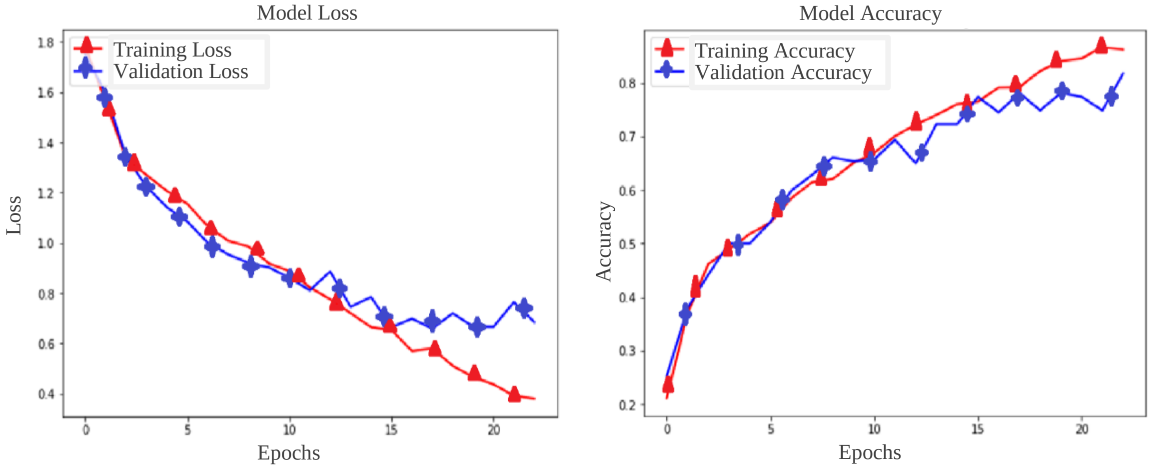

Figure 3. Model accuracy and model loss

3.3. Discussion

The results obtained using the values of the given hyperparameters are acceptable, and the accuracy is promising. Moreover, the small difference between training accuracy and the validation accuracy shows that neither overfits nor underfits happened. In the following text, we will discuss the results achieved above. We may choose a different value for each of the convolution hidden layers regarding the output filter tuning. By manipulating this value as the results show in Table 7, the values 32, 64, 128, and 512 were chosen for the first, second, third, and fourth convolution layers, respectively. On the other hand, and as expected, the best activation function for neural networks is ReLU. The results in Table 8 agrees with this trend. Additionally, Adam optimizer and image size of sixty were used as depicted in Tables 9 and 10, respectively. Choosing this optimizer and image size was after extensive experiments. Part of them is shown above. The image size usually is better to be bigger for other datasets.

As shown in Table 11, the dropout rate values were manipulated to get the best accuracy while avoiding overfitting. Table 11 only shows the accuracy result. However, we usually look also at the overfitting between the training and test results. We achieved the best accuracy on dropout 0.20 between convolutional layers and 0.50 after Denes based on the extensive experiments. On the other hand, the batch size selection is usually in the range of 16 to 128. There is no default best value. Selecting it is subject to experiments on the data we work on it. In our case, 64 batch sizes achieved the best results.

After the extensive experiments and trying different values for the hyperparameters, the best final result achieved is 81.75% accuracy. As shown in Figure 3, the difference between the training and the validation is low, indicating that the overfitting and underfitting are low. Typically, there is no way to prove that the achieved results are the best in any machine learning model. They could be enhanced in different ways. Nevertheless, 81.75% is considered good reasonable accuracy of such research. Especially, the dataset was self-collected, there is a need for a dermatologist to heavily participate in such research which is not easy to find due to their business, and we are working on six classes classification.

In this paper, we have designed and implemented a computer-based diagnosis system for six skin diseases. We also built an Android application that can take a picture for the skin and diagnose it. It can be concluded from the results that the suggested system can be capably used by patients and physicians to diagnose the skin diseases more accurately. Such an application significantly reduces the required time and cost for both the patient and the physician. The patient can get an initial diagnosis before going to the physician and without paying money and wasting long times in medical imaging systems. Similarly, the physician can reduce the effort by getting an initial diagnosis before seeing the patient, and so can serve more patients. This system is also useful for the rural areas where the dermatologists may not be available. Additionally, since the tool is supported by an Android application, images can be acquired in any conditions, and it can achieve the purpose of automatic diagnosis of skin diseases.

3.4. Datasets and challenges

It is easy to find images dataset of skin cancer. However, there are moderately not many datasets in the broad field of dermatology and much fewer datasets of skin disease pictures [21]. Besides, most of these datasets do not have enough pictures, and they are not freely accessible, which gives an extra hindrance to performing reproducible exploration in the territory. Instances of dermatology-related picture datasets in the late examination include: i) dermofit image library [35] is a dataset containing 1,300 pictures for ten classes and ii) dermnet contains more than 23,000 skin pictures isolated into 23 classes. In 2016, the international symposium on biomedical imaging (ISBI) delivered a test dataset for skin disease investigation towards melanoma discovery. Pictures in this dataset were obtained from the international skin imaging collaboration.

Another challenge is represented in finding a single dataset in one place that comprises several skin diseases. For example, it is easier to find a dataset that contains images for skin cancer only rather than finding a dataset that contains images for the six skin lesions. Therefore, in such research, the researcher should start building a self-dataset from different datasets and even from searching Google. Consequently, the images have different characteristics that need image processing to unify them and filter the noise. Additionally, since the images are collected from different sources, the images must be seen first by a dermatologist to label them and filter erroneous images. Such a step forms a big challenge because the dermatologists are very busy usually. One more challenge in this research area is to find images that represent the world population, if possible. Notably, the used dataset contains skin diseases from light-skinned people. Similarly, the images in the ISIC are mainly from US, Europe, and Australia. Therefore, the proposed system and other existing similar research may not give accurate classification for dark-skinned people. Hence, it is better to include dark-skinned people pictures during the CNN training.

4. CONCLUSION AND FUTURE WORK

With the daily increase of skin disease patients, the problem of classification becomes more challenging. The demand for automated classifiers is going to increase, especially after achieving good results in it. We propose a system for assisting dermatologists and patients during the diagnosis of skin diseases. Specifically, we designed and implemented a six-class classifier that takes as an input a picture of skin that is infected with one of six prevalent skin diseases, introduced a model on top of deep convolutional neural networks, and utilized this model to predict the type of skin disease in a given image. Additionally, we designed and implemented an Android application as an interface to our system. It takes a live picture from the patient, and then it categorizes it. The achieved accuracy is promising and up to 81.75%. However, some possibilities for accuracy improvements and as future work directions are summarized by: i) use more massive datasets, ii) work on binary classification for each type of the six diseases, iii) more intensive work on tuning the hyper parameters, which is a time-consuming operation, iv) cross dataset validation which is similar to cross-validation but using different datasets, v) feature engineering and feature selection, and iv) adding clinical data like age, race, skin type, or gender as inputs to the classifier.

REFERENCES

[1] L. Tizek, M. C. Schielein, F. Seifert, T. Biedermann, A. Böhner, and A. Zink, "Skin diseases are more common than we think: screening results of an unreferred population at the Munich Oktoberfest," *Journal of the European Academy of Dermatology and Venereology*, vol. 33, no. 7, pp. 1421–1428, Jul. 2019, doi: 10.1111/jdv.15494.

[2] Department of Statistics, "Jordan statistical yearbook 2016," *Department of Statistics (DOS)*, 2016.

[3] M. W. Tsang and J. S. Resneck, "Even patients with changing moles face long dermatology appointment wait-times: A study of simulated patient calls to dermatologists," *Journal of the American Academy of Dermatology*, vol. 55, no. 1, pp. 54–58, Jul. 2006, doi: 10.1016/j.jaad.2006.04.001.

[4] S. K. Patnaik, M. S. Sidhu, Y. Gehlot, B. Sharma, and P. Muthu, "Utomated skin disease identification using deep learning algorithm," *Biomedical and Pharmacology Journal*, vol. 11, no. 3, pp. 1429–1436, Sep. 2018, doi: 10.13005/bpj/1507.

[5] G. M. White, "Recent findings in the epidemiologic evidence, classification, and subtypes of acne vulgaris," *Journal of the American Academy of Dermatology*, vol. 39, no. 2, pp. S34–S37, Aug. 1998, doi: 10.1016/S0190-9622(98)70442-6.

[6] R. Ramli, A. S. Malik, A. F. M. Hani, and A. Jamil, "Acne analysis, grading and computational assessment methods: an overview," *Skin Research and Technology*, vol. 18, no. 1, pp. 1–14, Feb. 2012, doi: 10.1111/j.1600-0846.2011.00542.x.

[7] T. Vos *et al*., "Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013," *The Lancet*, vol. 386, no. 9995, pp. 743–800, Aug. 2015, doi: 10.1016/S0140-6736(15)60692-4.

[8] A. M. Arvin, "Chickenpox (Varicella)," in *Varicella-Zoster Virus*, Basel: KARGER, 1999, pp. 96–110.

[9] L. C. Schon and D. E. Baxter, "Neuropathies of the foot and ankle in athletes," *Clinics in Sports Medicine*, vol. 9, no. 2, pp. 489–509, Apr. 1990, doi: 10.1016/S0278-5919(20)30743-2.

[10] E. J. Teasdale, I. Muller, and M. Santer, "Carers' views of topical corticosteroid use in childhood eczema: a qualitative study of online discussion forums," *British Journal of Dermatology*, vol. 176, no. 6, pp. 1500–1507, Jun. 2017, doi: 10.1111/bjd.15130.

[11] N. Talsania, B. Lamb, and A. Bewley, "Vitiligo is more than skin deep: a survey of members of the Vitiligo Society," *Clinical and Experimental Dermatology*, vol. 35, no. 7, pp. 736–739, Oct. 2010, doi: 10.1111/j.1365-2230.2009.03765.x.

[12] H. W. Rogers, M. A. Weinstock, S. R. Feldman, and B. M. Coldiron, "Incidence estimate of nonmelanoma skin cancer (Keratinocyte Carcinomas) in the US population, 2012," *JAMA Dermatology*, vol. 151, no. 10, Oct. 2015, Art. no. 1081, doi: 10.1001/jamadermatol.2015.1187.

[13] M. Mahmood, B. Al-Khateeb, and W. Makki Alwash, "A review on neural networks approach on classifying cancers," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 9, no. 2, pp. 317–326, Jun. 2020, doi: 10.11591/ijai.v9.i2.pp317-326.

[14] H. Karim, S. R. Niakan, and R. Safdari, "Comparison of neural network training algorithms for classification of heart diseases," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 7, no. 4, pp. 185–189, 2018, doi: 10.11591/ijai.v7.i4.pp185-189.

[15] M. H. A. Hijazi, S. Kieu Tao Hwa, A. Bade, R. Yaakob, and M. Saffree Jeffree, "Ensemble deep learning for tuberculosis

detection using chest X-ray and canny edge detected images," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 8, no. 4, pp. 429–435, Dec. 2019, doi: 10.11591/ijai.v8.i4.pp429-435.
[16] S. Bakshi and Sathya A., "A performance study of the suitability of adaptive boosting in red acne detection," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 14, no. 3, pp. 1493–1498, Jun. 2019, doi: 10.11591/ijeecs.v14.i3.pp1493-1498.
[17] F. Abdali-Mohammadi, M. N. Meqdad, and S. Kadry, "Development of an IoT-based and cloud-based disease prediction and diagnosis system for healthcare using machine learning algorithms," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 9, no. 4, pp. 766–771, Dec. 2020, doi: 10.11591/ijai.v9.i4.pp766-771.
[18] "Melanoma project website," ISIC. https://www.isic-archive.com/#!/topWithHeader/wideContentTop/main (accessed Dec 2021).
[19] P. Tschandl, C. Rosendahl, and H. Kittler, "The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions," *Scientific Data*, vol. 5, no. 1, Art. no. 180161, Dec. 2018, doi: 10.1038/sdata.2018.161.
[20] A. A. Nugroho, I. Slamet, and Sugiyanto, "Skins cancer identification system of HAMl0000 skin cancer dataset using convolutional neural network," in *International Conference on Science and Applied Science (ICSAS) 2019*, 2019, Art. no. 020039, doi: 10.1063/1.5141652.
[21] A. Romero-Lopez, X. Giro-i-Nieto, J. Burdick, and O. Marques, "Skin lesion classification from dermoscopic images using deep learning techniques," 2017, doi: 10.2316/P.2017.852-053.
[22] J. Brownlee, *Master machine learning algorithms: discover how they work and implement them from scratch*. Machine Learning Mastery, 2016.
[23] S. Kalouche, "Vision-based classification of skin cancer using deep learning," *2015, conducted on Stanfords Machine Learning course (CS 229) taught*, 2016.
[24] E. Çevik and K. Zengin, "Classification of skin lesions in dermatoscopic images with deep convolution network," *European Journal of Science and Technology*, pp. 309–318, Oct. 2019, doi: 10.31590/ejosat.638247.
[25] H. Liao, "A deep learning approach to universal skin disease classification," University of Rochester Department of Computer Science, 2016.
[26] N. Hameed, A. M. Shabut, and M. A. Hossain, "Multi-class skin diseases classification using deep convolutional neural network and support vector machine," in *2018 12th International Conference on Software, Knowledge, Information Management and Applications (SKIMA)*, Dec. 2018, pp. 1–7, doi: 10.1109/SKIMA.2018.8631525.
[27] M. A. Taufiq, N. Hameed, A. Anjum, and F. Hameed, "M-skin doctor: a mobile enabled system for early Melanoma skin cancer detection using support vector machine," in *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*, Springer International Publishing, pp. 468–475, 2017.
[28] A. Victor and M. Ghalib, "Automatic detection and classification of skin cancer," *International Journal of Intelligent Engineering and Systems*, vol. 10, no. 3, pp. 444–451, Jun. 2017, doi: 10.22266/ijies2017.0630.50.
[29] H. Cataloluk and M. Kesler, "A diagnostic software tool for skin diseases with basic and weighted K-NN," in *2012 International Symposium on Innovations in Intelligent Systems and Applications*, Jul. 2012, pp. 1–4, doi: 10.1109/INISTA.2012.6246999.
[30] U.-O. Dorj, K.-K. Lee, J.-Y. Choi, and M. Lee, "The skin cancer classification using deep convolutional neural network," *Multimedia Tools and Applications*, vol. 77, no. 8, pp. 9909–9924, Apr. 2018, doi: 10.1007/s11042-018-5714-1.
[31] A. Esteva *et al.*, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, Feb. 2017, doi: 10.1038/nature21056.
[32] L. C. De Guzman, R. P. C. Maglaque, V. M. B. Torres, S. P. A. Zapido, and M. O. Cordel, "Design and evaluation of a multi-model, multi-level artificial neural network for eczema skin lesion detection," in *2015 3rd International Conference on Artificial Intelligence, Modelling and Simulation (AIMS)*, Dec. 2015, pp. 42–47, doi: 10.1109/AIMS.2015.17.
[33] S. Albelwi and A. Mahmood, "A framework for designing the architectures of deep convolutional neural networks," *Entropy*, vol. 19, no. 6, Art. no. 242, May 2017, doi: 10.3390/e19060242.
[34] A. Gulli and S. Pal, *Deep learning with Keras*. Packt Publishing, 2017.
[35] L. Ballerini, R. B. Fisher, B. Aldridge, and J. Rees, "A color and texture based hierarchical K-NN approach to the classification of non-melanoma skin lesions," in *Color Medical Image Analysis*, Springer Netherlands, pp. 63–86, 2013.

BIOGRAPHIES OF AUTHORS

Ramzi Saifan received the B.Sc. and M.S. degrees in computer engineering from Jordan University of Science and Technology, Irbid, Jordan, in 2003 and 2006 respectively, and received Ph.D. degree in computer engineering from Iowa State University in 2012. Since Jan 2013, he joined as an assistant professor in the Department of Computer Engineering at the University of Jordan. His current research interests include computer networks, computer and network security, cognitive radio networks, machine learning applications, and image processing. Saifan published several papers in peer reviewed journals and conferences. He can be contacted at email: r.saifan@ju.edu.jo.

Fahed Jubair graduated from Purdue University in 2014 with a Ph.D. degree in Electrical and Computer Engineering. He received his B.Sc. degree from the University of Jordan in 2007. Dr. Jubair is currently an assistant professor of Computer Engineering at the University of Jordan. His main research interests include optimizing compilers, parallel computing, heuristic algorithms, and machine learning. He can be contacted at email: f.jubair@ju.edu.jo.