Technoscience Journal for Community Development in Africa 2:1 (2021) 57–68
ISSN 2659-0573

Research Article

Convolutional neural network based approach for dermatological disease prediction

Afiz Adeniyi Adeyemo[*,1], Sulaimon A. Bashir[1], Olawale S. Adebayo[1], Shefiu O. Ganiyu[1,2]

[1] School of Information and Communication Technology, Federal University of Technology Minna, Nigeria
[2] School of Mathematics and Computing, Kampala International University, Kampala, Uganda

Abstract: The fast growing application of Artificial Intelligence in the field of medicine has led to an improvement in the detection and treatment of many kinds of ailments including skin diseases. However, there exists deficiency in the performances of the existing image pixel scaling techniques when applied to skin diseases detection. Pixel scaling is a major preprocessing tool in the classification of images. To improve the performance of skin disease detection and classification model, this paper proposes a new pixel scaling technique called *Mean Pixel Division*. At the preprocessing stage, each pixel value in the skin diseases images is divided by the mean of the entire channel pixel values. This reduces the range of the pixel values to a manageable level. A synthetic minority oversampling technique (SMOTE) was used to overcome the challenge of unbalanced classes' distribution in the dataset. Then, a CNN architecture was designed and trained with the earlier pre-processed images. The evaluation of the proposed approach compared with some existing scaling techniques shows that our approach outperformed the existing techniques, having recorded 99.62%, 98.66% and 99.78% in terms of performance accuracy, sensitivity and specificity respectively..

Keywords: Dermatological diseases; image pixel scaling; convolution neural network; data balancing

I. Introduction

The adoption of Artificial Intelligence (AI) in medicine has simplified the activities of medical experts in health management. The medical personnel have been able to utilize digitized medical images in overcoming problems of time consumption and misclassification that arise from the traditional method of diagnosis. It also eases some processes such as patients monitoring, drug administration, medical data collection and in selecting appropriate treatment approach (Mendeley Careers, 2018). The utilization of automated system in providing clinical services has made all these services deliverable in the administration of various illnesses.

The disorder of the skin has been proven to be among the most common and dangerous ailments (ACS, 2019, Randolph, 2002 and Salah et al., 2021) but is curable if it is promptly and accurately detected and treated accurately. However, the procedure involved in the traditional method of detecting skin diseases involves a number of complex stages (Wenhai et al., 2004), this mostly leads to waste of time, efforts and resources.

Traditionally, detecting and classifying skin diseases undergo a three-step procedure. The first is that the pathologists, who are the experts in managing skin infections, use any of the available approaches in carrying out prognosis examination on the infectious part of the skin (Mayo Clinic, 1998). The prognosis is followed by biopsies where the cell sample is removed from the suspected infectious part of the skin to

* Corresponding author:
Email: abdulhafeez_adeyemo@yahoo.com

KWASU PRESS AND PUBLISHING

undergo laboratory analysis (NBCF, 2019) to confirm the presence of infection. If it is established that the skin portion is infected, it undergoes another process known as staging. Staging is carried out to grade the degree of spread of the disease in the skin (NCI, 2020) on the skin sample. The classification of the diagnosed tumour is carried out by histopathologists who perform autopsies examination (AnaPath, 2019). This process consumes times and the results many times lack accuracy, which invariably leads to wrong management of the skin lesion which is more calamitous.

The acceptance of machine learning techniques in automating the diagnosis process have proffered simplified solutions to the problems of time consumption, high cost of resources involved and that of the misclassification (Yanming et al., 2016, Shin et al., 2006). Hence mismanagement of skin related issues is increasingly being overcome. The machine learning algorithm most suitable for image classification is the Convolutional Neural Network (CNN) because of its ability to group images by matching and also to identify substances. It extracts and learns special features (or representations) of the image and maps them from the upper layer of the algorithm to the lower layer, down to the output point where the detection and classification are performed.

Features contained in an image make a great impact on the performance accuracy of any classifying algorithm (Wang et al., 2019). In the past, efforts were made by researchers to establish the influences of some image attributes on the performance of image classifiers in some existing research works. For instance, Samuel and Lina (2016) worked on the effect of image quality distortions, Mathieu et al. (2017) explored the effect of compression of image, Sanchez et al. (2016) studied the impact of illumination feature. Similarly, the attention of Chavalier et al., (2015) was on the image resolution, Suresh and Gaurav (2018) examined the effect of image spatial resolution in the study.

However, the amount of data and the structure of datasets have a noticeable impact as well on the classification performance accuracy of a classifier (Foody et al., 1995 and Soleyman, 2018). The performance of a classifier, where some classes dominate others by having more instances, will be biased towards the classes with higher samples while little or no attention will be paid to the classes with few instances. Medical images are generally scarce, and sometimes irregularity in the number of instances in the various classes of the dataset (Belarouci & Chikh, 2017). These are some of the challenges faced when classifying medical data and the ones related to dermatology is not an exception. Hence, this obvious challenge requires data resampling to improve the performance of the classifier.

Interestingly, several research works have come up with different scaling techniques to suggest solutions to the problems of diagnosis and classification of skin infections. However, low performances recorded by these techniques are indications that they are deficient and there is need to improve on their performances.

Many of the existing scaling techniques were hybridized with the augmentation method of generating data, but this has not convincingly improved the performances of the classifiers. Hence, there is need to look inward to other scaling technique hybridized by another method of data generation to improve on the performance of image classifiers.

Therefore, this study proposes Mean Pixel Division technique. The technique is capable of extracting the pixels from dermatological images, calculating some statistics on the pixel values, synthesizing additional samples to the minority classes using a more robust image sample generating technique, processing them and classifying the skin samples, on high performance rates, into seven classes of skin related issues. Hence, the aim of this research work is to develop a hybridized image scaling, synthetic oversampling and CNN techniques to detect dermatological diseases.

The study has been able to contribute to knowledge by designing and implementing a CNN image based model enhanced with Mean Pixel Division technique for classifying skin diseases.

The remaining part of this paper is organised as follows: The related literatures from the previous works are reviewed in Section 2. Section 3 describes the details of the procedure carried out in solving the identified problem, while results of various experiments carried out in section 3 are presented and discussed in section 4. Section 5 is the conclusive part where the conclusion and recommendation on future works are provided.

2. Related works

Deep Learning is fast becoming a key instrument in Artificial Intelligent (LeCun et al., 2015). One of the problem areas where deep learning excels is image

classification (Rawat & Wang, 2017). The goal of image classification is to classify a specific image according to a set of possible categories. The attempts were to produce medicated systems that are capable of diagnosing skin related issues, just as performed by human dermatologists, but in an effective and more accurate manner. Several studies carried out for classifying the skin diseases have established that the use of instant based algorithm is more robust and accurate than the traditional method of classification carried out by human (Xiong et al., 2019; Tschandl et al., 2020). The most relevant of these studies are reviewed below.

Reddy (2018) classified skin related issues with transfer learning approach, substituting the last layer of the pre-trained ResNet architecture with a combination of max pooling layer and an average pooling layer. These two pooling layers were followed by two fully connected layers. The study used the HAM10000 dermatological dataset in evaluating the performance of this new created architecture. In preprocessing the image samples, the sizes of the images were enlarged to 224 by 224 to increase resolution of the image features. Normalization was performed on the instances to meet the standard of ImageNet and were augmented to increase the amount of the image samples for the training. The results obtained show that the system performed 91% accuracy, which can still be improved upon.

Danilo and Nilton (2018) as well proposed a model for the classification of twelve skins related issues. This study chose three sets of dermatological datasets, the first is the MED-MODE, containing 170 skin samples which were classified into 100 nevus and 70 melanoma cases. Another dermatological dataset used by the study is Edinburg, made available by the Edinburg Dermofit Image Library (EDIL). It consists of 1,300 skin related samples of ten types of skin diseases. Also, used are the skin samples from the Atlas Skin Image dataset, which contains skin issues of six types. The collection of the three datasets did not produce enough samples for training, hence were augmented and fed into the pre-trained RESNET152 architecture for classification. The results of the experiment showed that the accuracy of the training was 78% which can still be improved on.

Also, Agilandeeswari et al., (2019) used texture-based segmentation to classify skin lesions. In approaching the classification process, they applied filters to segregate the lesion areas of the skin and the unwanted background from the skin image. The Gray Level Co-occurence Matrix (GLCM) was after this used to extract the features from the lesions area before being forwarded for processing in a CNN classifier. The study obtained an average of 96% accuracy of skin classification. The study did not explore the image pixels and the accuracy can be improved upon. Sriwong et al. (2019) employed the integration of skin image data with patients' details to determine the performance of CNN. The study made use of the skin images collected from the Medical University of Austra, Harvard and published by Tschandl et al. (2018). A deep learning approach was utilised in designing the process of the skin diseases classification. To achieve their aim, Feature Extraction Support Vector Machine (FESVM) was pre-trained along the AlexNet CNN image classifier that was employed for the training. At the training and validation stage, five rounds of experiments were performed, one each for the AlexNet transfer learning, FESVM, FESVM+Age, FESVM+Age+Sex and FESVM+Age+Sex+Location. The results reveal that the pre-trained AlexNet model produced training accuracy of 84.94% and testing accuracy of 79.29%, while the FESVM performed 100% training accuracy and 78.7% testing accuracy. The results obtained when the description of the patients was added indicate that each of the FESVM+Age and FESVM+Age+Sex produced 100% training accuracy and 80.16 testing accuracy. The training and testing accuracies obtained from the addition of the three attributes to the image samples are 100% and 80.39% respectively; hence the combination that performs the best among the five scenarios. The concern here is that the system may not be efficient when the details of the patients are not available.

A system proposed by (Khan et al, 2019) for classifying skin related diseases using features extraction transfer learning of deep convolutional neural networks and a feature selector. The system used three different datasets which are HAM10000 dermatological image dataset, ISBI 2016 dataset and ISBI 2017 image dataset to train some selected pre-trained RESNET neural networks. The information of these deep neural networks were fused, the most suitable features were selected and transmitted into SVM supervised learning technique for samples' classification. The results obtained after the experiments indicate that the developed system performed 89.8% accuracy on the HAM10000 image dataset, 90.2% accuracy on the ISBI 2016 dataset and 95.6% accuracy on the ISBI image dataset. This approach, apart from the low accuracy recorded, is too deep which in turn consumes much time automated technique aims at providing solution to.

Adegun and Viriri (2020) proposed a framework capable of performing segmentation and classification to detect skin related issues. The system designed to integrate conditional random Field that confirm the contours and localise the samples' boundaries, this is possible with the presence of a series of Gaussian filters that are mounted for detecting edges. The study evaluated architecture of DesNet with the HAM10000 dermotological dataset achieving accuracy of 98%, 98.5% recall and AUC 99%. However, the use of multiple stages in carrying out the classification can lead to complexity in the detection process. And most recently, Garg, et al (2021) came up with a system to detect and classify skin diseases to various types using Convolutional Neural Network. The researchers use the dermatological HAM10000 image dataset to evaluate the residues of each of the VGG and ResNet architectures having frozen some layers of the pre-trained image classifiers. The samples of the images were prepared for the experiment by improving the quality of the contrast with the aid of Histogram Equalizer, and by augmenting the samples by copying through factors such as rotation, zooming and translation. The study as well evaluated the new approach using Random Forest algorithm, Support Vector Machine and XGBoost. The results obtained from evaluating the system show that it performed 0.88 precision, 0.74 recall and 90.51% accuracy. The results obtained can still be improved upon using more efficient preprocessing approach.

The results obtained by many of the reviewed studies indicate that various scaling techniques applied on the skin samples were defective. The deficiency in these techniques is a gap that required urgent bridging. If the gaps are bridged, there is clear indication that the performances can be improved upon, which is what this study seeks to address.

3. Methodology

The proposed is designed as a data preprocessing step to enhance the prediction of diseases. Foremost, the image data were preprocessed by scaling the pixel values in order to bring the range of the pixel values to a bearable size. The sample distribution shows that the classes are imbalanced, classifiers will be biased towards the minority classes while more recognition is accorded the majority classes. The issue of class imbalancing is addressed by synthesizing the samples in the minority classes using Synthetic Minority Oversampling Technique (SMOTE). The sample synthesis technique balances the amount of data samples across all classes, which enables fair recognition to all classes as each contributes to the performance of the classifier. The CNN architecture designed for this study is then trained with the synthesized values and tested with the testing set to evaluate the performance of the proposed technique. Performance metrics used to evaluate the model are accuracy, sensitivity and specificity. Figure 1 shows the procedural structure of the proposed skin diseases classification process.

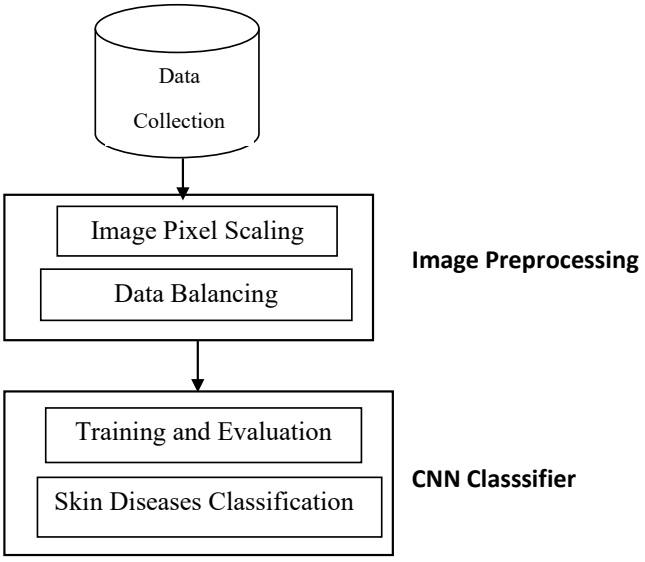


Figure 1: The procedural structure for the detection process

(A) Data preprocessing

Preprocessing of data is highly required in the development of a machine learning algorithm as a result of the high range of the pixel values. This is necessary to improve the quality of the image features by removing undesirable representations of the images. It also enhances the performance of the image classifier. In this study, some data preprocessing techniques were applied on the HAM10000 dataset in simplifying the samples for processing and improving the performance of the proposed technique. The preprocessing techniques are discussed below.

Mean Pixel Division: The mean of the entire image channels is determined. Each of the pixel values in the particular image is divided by the mean to obtain an array of quotient values. The quotients generated are fed into the classifier for training and evaluation. The

function of the proposed scaling is defined as stated in Equation (4).

$$f(x) = \frac{x}{\mu} \quad (1)$$

where x = the value of an image pixel
$f(x)$ = the function defined on the pixel values,
μ = mean of the colour channel,

Pixel scaling algorithm

The following algorithm denotes the procedure used in coming up with the proposed scaling technique. The procedural statements in the box find the quotient of the pixel values and the mean of each of the colour channels.

Algorithm 1: MeanPixelDivision(X[0, …, n-1], μ)

//Finding the division of the pixel values by the mean $\mu\mu$
//Input: Image $X \in \Re^{a \times b \times n}$
//Output: $X_{Scaled} \in \Re^{a \times b \times n}$
k = a*b*n
For each image *g*,
findMean (a,b,n)
Sum = 0
For i in range (0, k) // each pixel
Sum $+= g[i]$
mean = $\frac{sum}{k}$
for each image g,
findDivision (g[i], mean)
For i in range (0, k) // each pixel
mean_division = $\frac{g[i]}{mean}$

Data balancing

The SMOTE tool is applied on the original data distribution to increase the number of instances in the minority classes. Application of the tool equally boosts the performance of the proposed technique. The use of SMOTE technique of sampling balancing enables balancing of the data samples across all classes. The training set of 9,015 instances is synthesized to 42,469. Table 1 reveals the number of samples, before and after synthesizing in each class.

Table 1: Original and synthesized samples of the training set

SN	Skin Type Acronyms	Original Frequency	Frequency before SMOTE	Frequency after SMOTE
1	Akiec	327	292	6,067
2	Bcc	514	449	6,067
3	Bkl	1,099	985	6,067
4	Df	115	103	6,067
5	Mel	1,113	991	6,067
6	Nv	6,705	6,067	6,067
7	Vasc	142	128	6,067
Total		10,015	9015	42,469

(B) Implementing the CNN Architecture

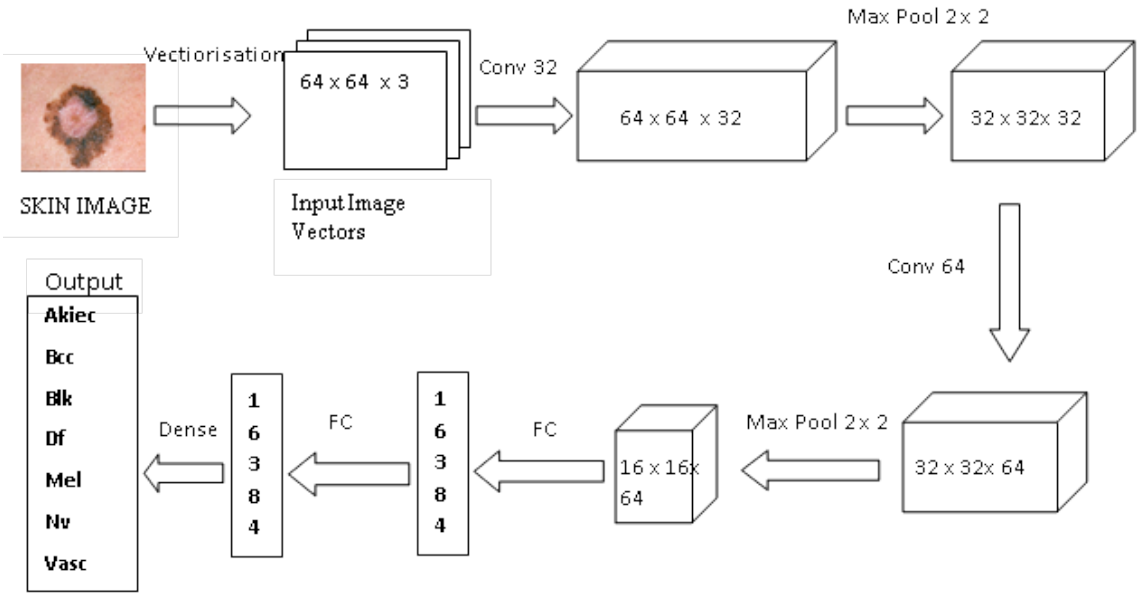
A CNN architecture was designed with the first layer designed to use a convolutional filter of size 32, while the second convolutional filter size is 64. Average pooling filter was used to reduce the spatial features transmitted from a convolution layer to the other. The choice of considering Max pooling can be justified for its ability to recognize and present the most important regions of an image to the network. The output node used is *softmax* because of its probabilistic interpretation in classifying values (Ayyuce, 2019), especially when the number of labels is multiple as it is the case in this study.

The architecture is implemented with Python 3.5.2 version (via Anaconda IDE), and supported by other libraries including Keras 1.1.0 and TensorFlow backend 0.3 high level API. The implementation was conducted on a personal computer having Intel Core i5 processor and 16 GB of RAM.

The architecture of the designed CNN image-based classifier is detailed in Table 2 and as shown in Figure 2.

Table 2: The highlight of the CNN architecture designed for this study

Convolutional Layers	Filter Size	Activation Function
Conv2D	(32, (3,3))	Relu
Conv2D	(32, (3,3))	Relu
MaxPooling2D	(2, 2)	
Conv2D	(64, (3, 3))	Relu
Conv2D	(64, (3, 3))	Relu
MaxPooling2D	(2, 2)	
Dense	128	Relu
Dropout	0.2	
Dense	64	Relu
Dropout	0.2	
Dense	7	Softmax
Number of Parameters	1,680,807	



FC = Fully Connected

Figure 2: The architecture of the CNN classifier implemented for this study

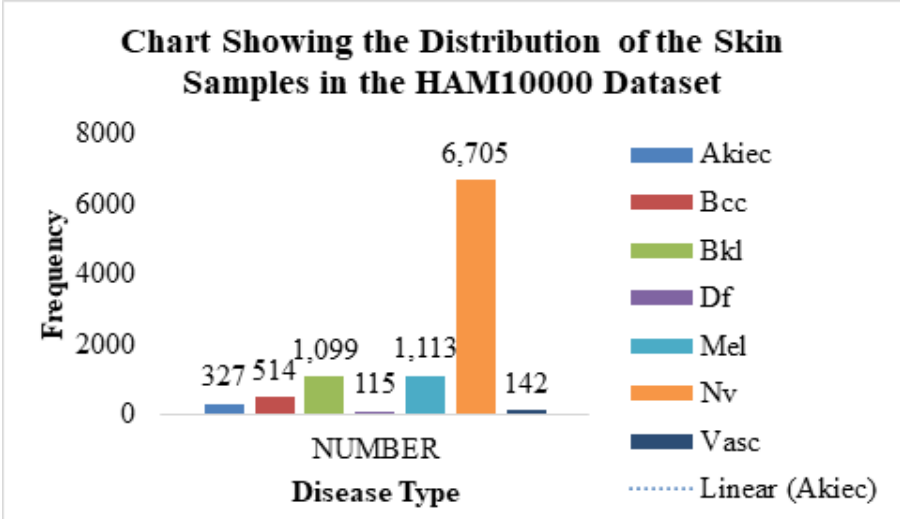


Figure 3: A chart representing the original skin distribution in the dataset

(C) Experiment

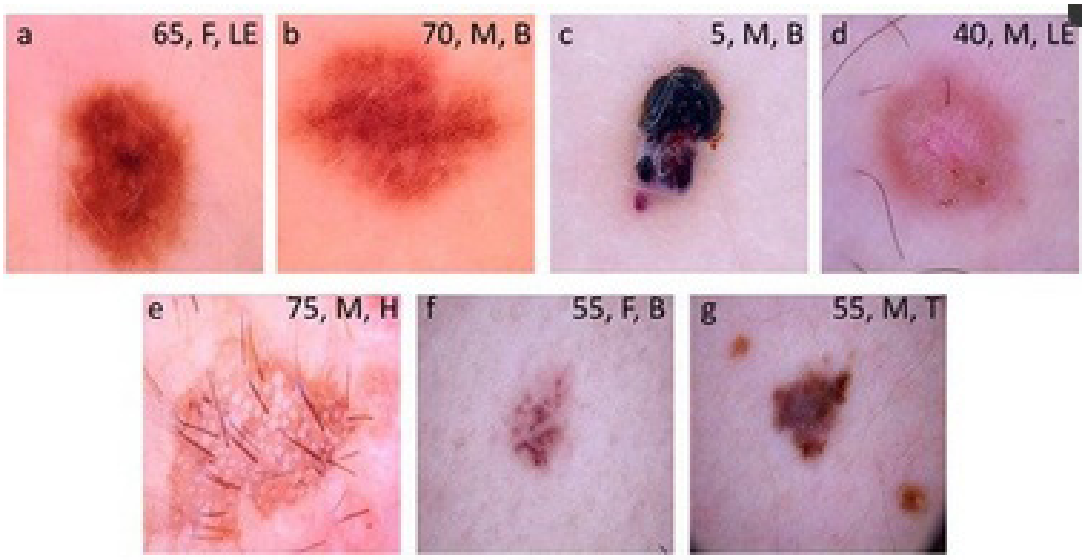
Dataset

The dermatological HAM10000 dataset collected from the Medical University of Vienna, Austria published by Tschandl, et al. (2018) is used for the study. It consists of a total number of 10,015 images collected from different populations over different locations and races and is presented in Table 3.

he distribution of the diseases in the HAM10000 dataset is represented in a bar chart as shown in Figure 3. The samples of the dermatological images are as shown in Figure 4.

Table 3: The distribution table for the skin disease types

S/N	Disease	Acronym	Number
1.	Actinic keratoses and intraepithelial carcinoma/Bowen's disease	Akiec	327
2.	Basal cell carcinoma	Bcc	514
3.	Benign keratosis-like lesions	Bkl	1,099
4.	Dermatofibroma	Df	115
5.	Melanoma	Mel	1,113
6.	Melanocytic nevi	Nv	6,705
7.	Vascular lesions	Vasc	142
Total			10,015



(a) Melanocytic nevus; (b) benign keratosis; (c) vascular lesion; (d) dermatofibroma; (e) intraepithelial carcinoma; (f) basal cell carcinoma; and (g) melanoma. Legends inside each image represents clinical data such as age, sex and localization associated to the image. F: female; M: male; LE: lower extremity; B: back; H: hand; T: trunk.

Figure 4: Sample picture of the seven skin types in the HAM10000 dataset

Training procedure

In training and testing the performance of the CNN classifier, the dataset was divided into 80:20 ratios for training and testing sets respectively. The training dataset was further divided into 90:10 ratios from training and validation datasets.

In the training session, values derived from applying some scaling techniques on the pixel values were fed into the classifier. The scaling techniques used include the *unscaled pixel values*, *local centering*, *global centering*, and *mean pixel division.*

(D) Performance evaluation

The training and testing accuracy performances of the pixels scaling techniques were monitored during the experiments. The performances of the techniques were further evaluated using accuracy, specificity and sensitivity performance metrics computed from various confusion matrices. The performance metrics used are discussed below.

Accuracy

Accuracy is measured by the proportion of the correct classified instances to the total instances available in the entire dataset. Equation 3.5 defines the function that performs the accuracy metric for the CNN image based classifier.

$$\text{Accuracy} = \left(\frac{True\ Positive + True\ Negative}{True\ Positive + True\ Negative + False\ Positive + False\ Negative}\right) \quad (5)$$

Sensitivity

Sensitivity of a classifier to the samples classified reveals how well it identifies the samples. The equation 3.6 shows the function for calculating the sensitivity of a classifier.

$$\text{Sensitivity} = \left(\frac{True\ Positive}{True\ Positive + False\ Negative}\right) \quad (6)$$

Specificity

Specificity measures the ability of a classifier to really classify samples that do not belong to a class rightly not belong to such class. The mathematical representation of specificity is as stated in Equation (7).

$$\text{Specificity} = \frac{True\ Negative}{True\ Negative + False\ Positive} \quad (7)$$

The performance evaluation of the image scaling techniques were performed and presented in Table 6.

4. Results and discussion

In this section, the results obtained from the experiments performed for this study are presented. In an attempt towards achieving the aim and objectives of this study, a CNN based image architecture was trained and evaluated overtraining and validation sets. The performance accuracies of the experiments on some existing techniques and the proposed scaling method as well as on various data structures were monitored.
The CNN architecture was trained with various forms of sample distributions which include the *original data structure of the dataset*, *weigh balancing technique* and *SMOTE*. The results obtained from various experiments are presented in Table 4.

Table 4: The results of the experiments

Statistics	Accuracies (%)		
	Original Data Structure	Weight Balancing	SMOTE
Unscaled Values	54.2	75.3	73.6
Local Centering	59.4	73.1	72.5
Global Centering	61.0	72.8	72.9
Mean Pixels Division (Proposed)	59.8	73.8	74.3

The *unscaled pixel values* technique performed 54.2% training accuracy, the *local centering* technique recorded accuracy of 59.4% and while with the technique of *global centering,* the accuracy is 61.0%. Similarly, the results reveal that the proposed scaling technique; *mean pixel division*, during the evaluation period recorded the training accuracy of 59.8% accuracy.

The performance of the *unscaled pixel* scaling technique, when trained and tested with balanced weights as shown Table 4 is 75.3%.Also from Table 4, it is observed that the accuracy of *local centering* is 73.1%, while that of the *global centering* model is 72.8% and that of the proposed technique; *mean pixel division* was able to make 73.8% accuracy.

The results presented in Table 4 as well reveal that the *unscaled pixel values* technique performed 73.6% accuracy, *local centering* performed 72.5% and the *global centering* recorded accuracy of 72.9%, *mean pixel division* technique recorded 74.3% accuracy.

It is observed from Table 4 that the classifier is sensitive to the data structure. The sensitivity to the structure of the dataset is connected to the class balancing and the amount of data used in training the classifier. According to the results presented, it is observed that the architecture performed better when the data were synthesized with the SMOTE. The confusion matrices for the techniques and SMOTE techniques are presented in presented in Figures 5, 6, 7 and 8.

Unscaled pixel values technique

The confusion matrix for the technique of *unscaled pixel values* is shown in Figure 5.

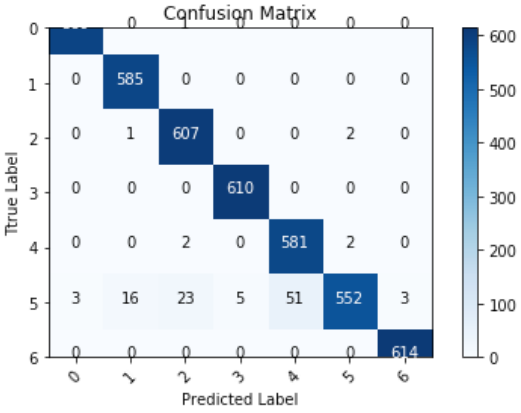


Figure 5: Confusion matrix for the *unscaled pixel values* technique

From Figure 5, 589 samples, representing 99.83%, were correctly classified as 'Akiec', while 1; representing 0.17% of the total samples was wrongly classified to be 'Bkl'. In the second row, 585 samples representing 100% were correctly classified as 'Bcc' and no instance from the class was wrongly classified. Out of 610 'Blk' present for classification, 607 were correctly classified and 3 were mis-classified to be 'Nv' and 'Bcc'. The model was able to classify the entire 610 'Df' samples correctly representing 100%. 581 of the 585 'Mel' samples, representing 99.32% were classified to be 'Mel', while 4, 2 each were classified as 'Bkl' and 'Nv' respectively. The *original pixel value* technique was unable to classify a total of 101, which is 15.47% of the samples belonging to class 'Nv' correctly, but was able classify up to 552 samples representing 84.53% correctly. However, the entire samples totalling 614 belonging to the 'Vasc' class was classified into 'Vasc' class. In overall, out of the total 4,247 samples classified, the model was able to classify a total of 4,138 samples correctly and 109 wrongly to achieve 99.27% performance accuracy.

Local centering scaling technique

Figure 6 shows the confusion matrix for the *local centering* method.

The confusion matrix for the method of *local centering* as shown in Figure 6 revealed that all 590 'Akiec' samples were correctly classified representing 100% classification. Similarly, the model achieved 100% classification in classifying the instances belonging to the class 'Bcc'. The model was able to classify 606 out of 610 samples of the class 'Bkl' correctly, and the remaining 4 which represent 0.66% of the samples in the class. Also, the entire 610 samples of the class 'Df' was correctly classified, but 98.46% of the of 'Mel' samples totalled 576 instances were classified correctly. In the 'Nv' class, the classification performance as shown in the table indicates that the model was able to classify a sum of 550 representing 78.24% of the class correctly while the remaining 103 samples were mis-classified. However, all the 614 samples contained in the class 'Vasc' were correctly classified. With these classifications, the model was able to classify a total number of 4,131 samples correctly and 116 wrongly to achieve a performance accuracy of 99.22%.

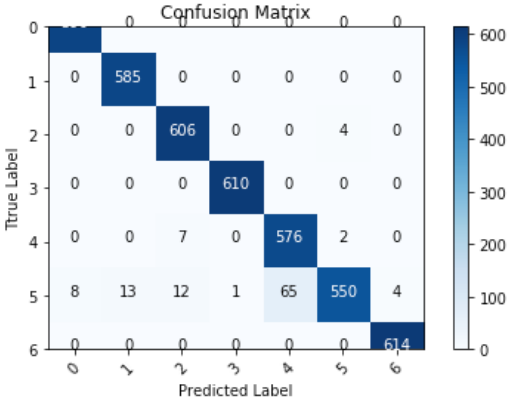


Figure 6: Confusion matrix for the *local centering* technique

Global centering technique

The confusion matrix of the *global centering* technique is shown in Figure 7.

The *global centering* technique classified all the samples belonging to 'Akiec' and 'Bcc' classes correctly to achieve 100% classification in each case. In the case of the 'Bkl' class, it was able to classify 99.67% standing for a sum of 608 out of 610 samples in the class correctly, and only 2 were not correctly classified. The method was able to classify the entire samples in the class 'Df' correctly as 'Df' samples hence, recording 100% accuracy. The instances of the class 'Mel' were 585, the method was able to classify up to 584 correctly but only 1 to the class 'Bkl', to achieve a performance accuracy of 99.83%. Also, it records 84.53% accuracy for classifying 552 samples belonging to the class 'Nv' correctly, while it mis-classified 101 of the samples representing 15.47% to other various classes. The classification performance of the technique in classifying the class 'Vasc' is accurate, classifying the entire 614 samples of the class correctly, to obtain 100% classification accuracy. In all, the technique was able to classify a sum of 4,143 instances correctly, which represents 99.30% of the classified samples and was unable to classify 104 samples representing 0.70% correctly.

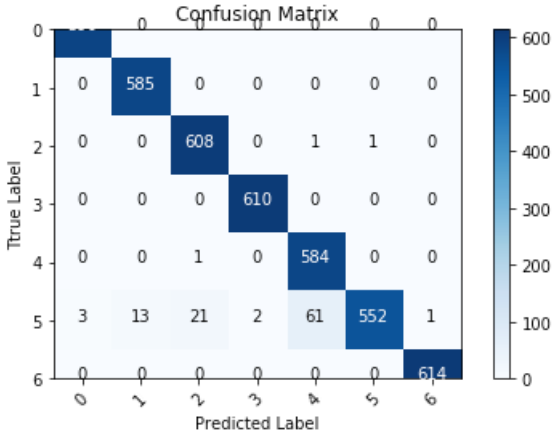


Figure 7: Confusion matrix for the method of *global centering*

The Mean pixels division technique

The confusion matrix for the method of *mean pixels division* is presented in Figure 8.

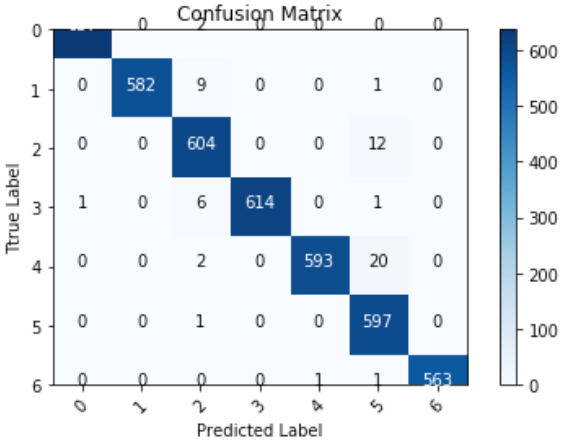


Figure 8: Confusion matrix for the technique *mean pixel division*

The classification results in Figure 8 show that the technique was able to classify 637 samples out

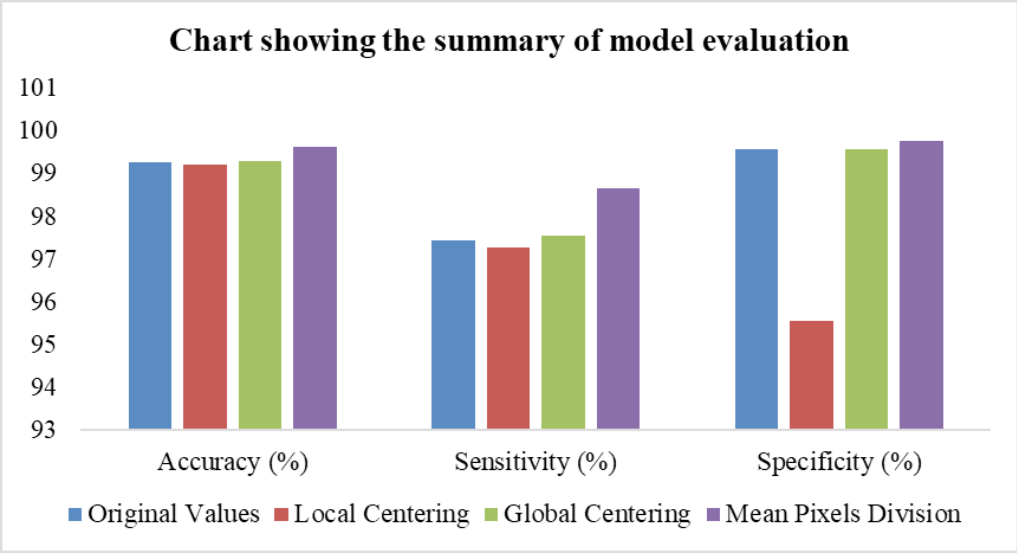


Figure 9: Chart showing the performances of the models

of the 639 of the class 'Akiec' correctly, to achieve 99.68% classification accuracy. The method records a classification accuracy of 89.19% while classifying the 'Bcc' class for have classified 582 instances correctly. Furthermore, 604 of the 616 samples of the 'Bkl' class were correctly classified to belong to the class, recording 98.05% classification accuracy. The technique records classification accuracy of 98.71% for classifying up to 614 samples of 622 samples contained in the class 'Df'. 593 samples were correctly classified to belong to 'Mel' class representing 96.42% classification accuracy, it misses 22 representing 3.58% of the entire samples in the class. 597 instances of the 598 samples present for classification in the class 'Nv' were correctly classified representing 99.83%, while 99.62% of the 565 samples of the class 'Vasc' was correctly classified to belong to the 'Vasc' class. Overall performance shows that the pixel scaling technique achieves a classification accuracy of 99.62%.

Table 5: Summary of the results for performance evaluation

Scaling Technique	Performance Measure (%)		
	Accuracy	Sensitivity	Specificity
Unscaled Pixel Value	99.27	97.43	99.57
Local Centering	99.22	97.27	95.55
Global Centering	99.30	97.55	99.59
Mean Pixel Division	99.62	98.66	99.78

Figure 9 shows in summary the visual presentations of the accuracy, sensitivity and specificity results of the performance evaluation.

From the results presented in Table 5 and as shown on Figure 8, it is obvious that the *Mean Pixels Division* scaling technique outperformed other models with 99.62% performance accuracy, 98.66% sensitivity and 99.78% specificity, hence can be adopted for the classification of dermatological skin diseases.

5. Conclusion

This study proposed and implemented a Convolutional Neural Network (CNN) based image classifier, enhanced with image scaling and image data synthesising techniques. The model was capable of classifying dermatological images of skin diseases into seven classes; Akiec, Bcc, Bkl, Df, Mel, Nv and Vasc. The CNN architecture was trained using a number of techniques, which are *unscaled value, local centering, global centering* and *mean pixel division*. The performances of the techniques were further evaluated to enable the adoption of the best performing one using accuracy, sensitivity and specificity performance measures. The evaluation results obtained indicated that the model *mean pixel division* outperforms others having scored 99.62% accuracy, 98.66% sensitivity and 99.78% specificity. Thus, considering the result obtained by Mean Pixel Division, we are optimistic that the time,

efforts and cost spent in diagnosing skin diseases can be minimised, while the danger posed by skin diseases will be reduced when the model proposed is put to use.

6. Future work

This study has been able to introduce a new pixel scaling technique in classifying skin diseases and by extension, image classification. However, there is room for improvement. Further research can be carried out to find the root of the division to see if the reduction in the pixel values will bring any improvement to the results obtained. In addition, the model can still be evaluated using other algorithms.

References

Adegun, A. A. & Viriri, S. (2020). FCN – Based DenseNet Framework for automated Detection and classification of skin Lesions in Dermoscopy Images. *IEEE*, 8, 150377-150395.

Agilandeeswari, L., Mahajan,T. S., & Keerthana, N. (2019). Skin Lesion Detection using Texture Based Segmentation and Classification by Convolutional Neural Networks (CNN). *International Journal of Innovative Technology and Exploring Engineering (IJITEEE)*, *9*(2), 2117-2120.

American Cancer Society. (2019) *Skin Cancer*. Retrieved from www.amp.cancer.org>cancer

AnaPath (2019). *Things to know about Histopathology*. Retrieved from https://anapath.ch/things-to-know-about-histopathology/

Ayyuce, K. (2019). *Comparison of Activation Functions for Deep Neural Networks*. Retrieved from https://www.medium

Belarouci, S. & Chikh, M. (2017). Medical Imbalanced Data Classification. *Advances in Science, Technology and Engineering Systems Journal*. DOI: 10.25046/aj020316. 2, 116-124.

Chevalier, M., Nicolas, T., Matheu, C., Jerome, F., Gilles, H., & Elodie, D. (2015). LR_CNN for Fine-Grained Classification with Varying Resolution, Image Processing (ICIP). *IEEE International Conference*.

Danilo, B. M., & Nilton, C. S. (2018). Skin Lesions Classification Using Convolutional Neural Networks in Clinical Images. *Computer Vision and Pattern Recognition*. arXiv:1812.02316.

Foody, G. M., McCulloch, M. B., & Yates, W. B. (1995). The effect of training set size and composition on artificial neural network classification. *Int. J. Remote Sensing,16*(9), 1707 – 1723.

Garg, R., Maheshwari, S., & Shukla, A. (2021). Decision Support System for Detection and Classification of Skin Cancer using CNN. *Innovations in Computational Intelligence and Computer Vision*. Sprinter, Singapore, arXiv preprint arXiv:1912.03798.2019, 578-586.

Khan, M. A., Javed, M. Y., Sharif, M. Saba, T, & Rehman, A. (2019). Multi-Model Deep Neural Network based Features Extraction and Optimal Selection Approach for Skin Lesion Classification. *2019 International Conferrence on Computer and Information Sciences (ICCIS)*, 1-7, Doi:10.1109/ICCISci.2019.8716400, 1-7

LeCun, Y., Bengio, Y. & Hinton, G. (2015). *Deep Learning*. Retrieved fromwww.nature.com, 436 – 444.

Mathieu, D., Karol, D., Kamel, A., Wassim, H., Lucc, M., & Maxime, P. (2017). Study of the Impact of Standard Image Compression Techniques on Performance of Image Classification with a Convolutional Neural Network. Diss. *INSA Rennes; Univ Rennes; IETR; InstitutPasca*.

Mayo Clinic (1998). *Diagnosis:Cancer Screening*. Retrieved from www.mayoclinic.org

Mendeley Careers. (2018). *Artificial Intelligence in Medicine*. Retrieved from https://www.mendeley.com/careers/news/careers-jobs-field/artificial-intelligence-medicine

National Breast Cancer Foundation (2019). *Inc. About Breast Cancer*. Retrieved from www.nationalbreastcancer.org

National Cancer Institute (2020). *Skin Cancer Treatment (PDQ). – Patient Version.* Retrieved from www.cancer.gov.

Randolph, S. (2002). When candida turns deadly: Thrush, diaper rash, and viginal yeast infection are all common manifestations of candida. But for immunocompromised patients, this common fungal species can cause a host of problems, including life threatening infections. *RN, 65*(3), 41-45.

Rawat, W., & Wang, Z. (2017). Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review. *Neural computation*, *29*(9), 2352–2449. https://doi.org/10.1162/NECO_a_00990Reddy, N. D. (2018). Classification of Dermoscopy Images using Deep Learning. *arXiv preprint arXiv:1808.01607*, scholargoogle.com

Reddy, N. D. (2018). Classification of Dermoscopy Images using Deep Learning. *arXiv preprint arXiv:1808.01607*, scholargoogle.com

Salah, I., Parkin, I. P., & Allan, E. (2021). Copper as an antimicrobial agent: recent advances. *Royal Society of Chemistry Advances*. DOI: 10.1039/DIRA02149AD. 11(30), 18179-18186.

Samuel, D., & Lina, K. (2016). Understanding How Image Quality Affects Deep Neural Networks. Quality of Multimedia Experience (QoMEX). *2016 Eighth International Conference on IEEE*.

Sanchez, A., Moreno, A. B., & Velez, J. (2016). Analysing the Influence of Contrast in Large-Scale Recognition of Neural Images. *Integrated Computer-Aided Engineering*, *23*(3), 221-235.

Shin, H. C., Roth, H. R., Lu, L. Xu, Z., Nogues, I., Yao, J., Mollura, D., & Summers, R. M. (2006). Deep Convolution Network for Computer Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning. *IEEE Transactions on Image Processing*, *15*(2), 430-444.

Soleyman, S. (2018). Effect of Dataset Size on Image Classification Accuracy. Online available: http://seansoleyman.com/effect-of-dataset-size-on-image-classification-accuracy/

Sriwong, K.,Bunrit, S., Kerdprasop, K. & Kerdprasop, N. (2019). Dermatological Classification Using Deep Learning of Skin Image and Patent Background Knowledge. *IJML&, 9*(6).

Suresh, P. K., & Gaurav, J. (2018). Effects of Varying Resolution on Performance of CNN Based Image Classification: An Experimental Study. *JCSE international Journal of Computer Sciences and Engineering*, *6*(9), 451-456.

Tschandl, P., Rinner, C., Apalla, Z., Argenziano, G., Codella, N., Halpern, A., Janda, M., Lalles, A., Longo, C., Malvely, J., Paoli, J., Puig, S., Rosendahl, C., Soyer, H. P., Zalaudek, I., & Kittler, H. (2020). Human-computer collaboration for skin cancer recognition. *Nature Medicine*. 26, 1229-1234.

Tschandl, P., Rosendahl, C., & Kittler, H. (2018). The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Scientific Data*. DOI: 10.1038/Sdata.2018.161. 5(1)

Wang, H., Chong, D., Huang, D., & Zou, Y. (2019). What Affects the Performance of Convolutional Neural Networks for Audio Event and Classification.*2019 8th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)*. doi: 10.1109/ACIIW.2019.8925277.140-146.

Wenhai, L. A., Jianzhong, Z. A., & Cao, Y. B. (2004). Detection of Treponema Pallidum in skin Lesions of Secondary and characterization of the inflammatory infiltrate. *Dermatology*. 208, 94-97.

Xiong, H., Lin, P., Yu, J., Ye, J., Xiao, L., Tao, Y., Jiang, Z., Lin, W., Liu, M., Xu, J., Hu, W., Lu, Y., Liu, H., Li, X., Zheng, Y., & Yang, H. (2019). Computer-aided diagnosis of laryngeal cancer via deep learning based on laryngoscopic images. *EBioMedicine*, *48*, 92-99.

Yanming, G., Yu, I., Ard,O., Songyang, L.,Song, W., & Micheal, S. L. (2016). Deep Learning for Visual Understanding: A Review. *Neurocomputing*, *187*, 27-48.