I.J. Intelligent Systems and Applications, **2023, 2, 23-36**
Published Online on April 8, 2023 by MECS Press (http://www.mecs-press.org/)
DOI: 10.5815/ijisa.2023.02.03

Classification of Images of Skin Lesion Using Deep Learning

Momina Shaheen*
School of Arts and Digital Industries, University of Roehampton, London, United Kingdom, SW15 2YB
E-mail: momina.shaheen@roehampton.ac.uk
ORCID iD: https://orcid.org/0000-0001-9424-9787
*Corresponding Author

Usman Saif
School of System and Technology, University of Management and Technology Lahore, Pakistan
E-mail: f2017114016@umt.edu.pk
ORCID iD: https://orcid.org/0000-0002-4547-044X

Shahid M. Awan
University of West Scotland, Paisley, Glasgow United Kingdom PA1 2BE
E-mail: Shahid.Awan@uws.ac.uk

Faizan Ahmad
Cardiff Metropolitan University, Llandaff Campus, United Kingdom CF5 2YB
E-mail: FAhmad@cardiffmet.ac.uk
ORCID iD: https://orcid.org/0000-0003-0595-5343

Aimen Anum
Islamia University Bahawalpur, Bahawal Nagar Campus 063200
E-mail: Aiman.anum@gmail.com
ORCID iD: https://orcid.org/0000-0002-0894-5266

Received: 17 August 2022; Revised: 09 November 2022; Accepted: 04 January 2023; Published: 08 April 2023

Abstract: Skin cancer is among common and rapidly increasing human malignancies, which can be diagnosed visually. The diagnosis begins with preliminary medical screening and by dermoscopic examination, histopathological examination, and proceeding to the biopsy. This screening and diagnosis can be automated using machine learning tools and techniques. Artificial neural networks are helping a lot in medical diagnosis applications. In this research, skin images are classified into 7 different classes of skin cancer using deep learning methodology, then analyzed the results w.r.t to their respective precision, recall, support, and accuracy to find its practical applicability. This model is efficient in comparison to the detection of skin cancer with human eyes. Human eyes detection can be 79% accurate at most. Thus, having a scientific method of diagnosis can help the doctors and practitioners to accurately identify the cancer and its type. The model provides 80% accuracy on average for all 7 types of skin diseases, thus being more reliable than human eye examination. It will help the doctors to diagnose the skin diseases more confidently. The model has only 2 misclassified predictions for Basal cell carcinoma and Vascular lesions. However, Actinic keratosis diagnosis is most accurately predicted.

Index Terms: Biomedical, Convolutional Neural Network, Deep Learning, Skin Cancer Diagnosis.

1. Introduction

Machine learning is has an applicability over multiple areas such as robotics [1], astronomy, banking and finance[2], computational biology, natural language processing[3], marketing, sentiment analysis, data mining[4]. However, its most literature is seen from the medical imaging [5], where the images are used to train the ML algorithm to diagnose the related disease.

The diagnosis of healthcare issues has led to a new era, with the increase in the amount of data, where the data-driven approaches are now used to diagnose MRI or CT-scan [6]. Artificial neural networks are helping a lot in the area of medical diagnosis applications. The machines use DL framework to process from highly complex models for data exploration and representation, which in turns performs accurate data analysis. Linear and Non-linear functions of the input data, weighted by the model parameters, are further hierarchically computed [7].

Machine learning, being evolved as one of the most powerful tools, will play an integral role in health care diagnosis in the future, which will be even practiced by most of the dermatologists [8]. Dermoscopy examines different types of skins to identify multiple skin diseases by capturing the images of the skin [9]. The technique, however, advances the diagnosis of the malignant skin lesions as compared to the one completed with the solo eye. Dermatoscopic images are also similarly processed and are a suitable source for the training of artificial neural networks (ANN), which are used to diagnose the pigmented skin lesions automatically. Machine learning (ML) techniques have set new yardsticks in this area of neural networks, which has even expanded the prospects that these automated diagnostic systems will diagnose types of pigmented skin lesions without any necessity of human expertise.

1.1. Skin Cancer Diagnosis and Neural Networks

Machine learning algorithms are getting complex each day, but they are still like machines. Even after a lot of expertise on the domain and human intervention, they can do only what they are designed for, nothing more and nothing less than that. Here deep learning holds a bit promising place for Artificial Intelligence designers and the rest of the world. Machine learning algorithms, in combination with neural networks, are now designing vigorous predicting systems that have brought machine learning ahead into its processing techniques[10,11] [Fig. 1]. Hence, Deep Learning is becoming more popular due to its higher accuracy precision, then others, if trained with a large amount of data. Deep learning technique uses its hidden layer architecture to learn the categories, starting from defining the basic low level categories like letters, moving towards a bit higher categories like words and finally moving to higher categories like sentences. Despite of traditional Machine learning algorithms, Deep Learning involves high-end machines. Hence GPU has now emerged as an integral part for processing Deep Learning algorithms.

The conventional machine learning algorithms require domain expertise to decrease data complexity by identifying most of the features and defining clear patterns for learning algorithms to work (shown in Fig 2). However, the Deep Learning algorithms learn high-level features incrementally from the data, thus eliminating the need of any domain experts or any core feature extraction.

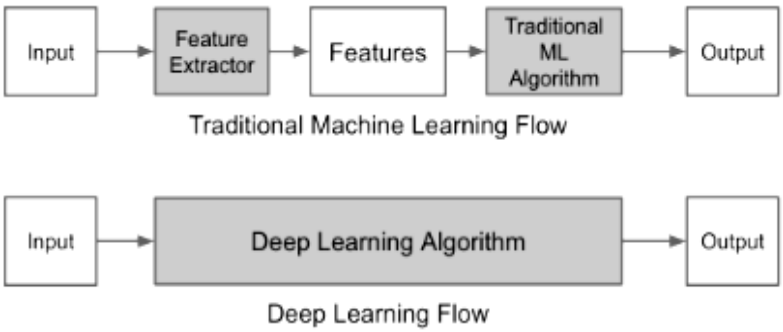

Fig.1. A Comparison between Classic Machine Learning and Deep Learning.

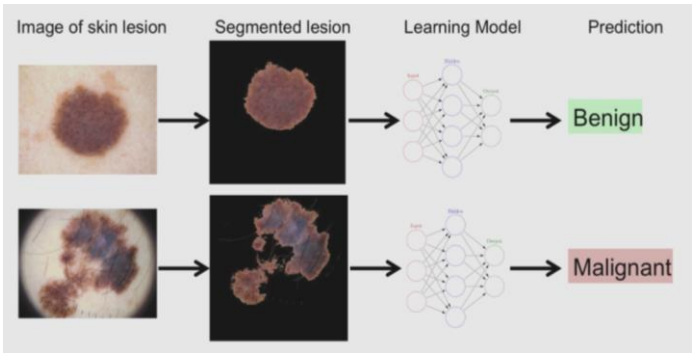

Fig.2. ML diagnostic tool for doctors to screen suspicious skin lesions and moles.

Preliminary studies have shown that biologist, medical residents and dermatologist from Stanford University's hospital could classify the skin lesions accurately for 46% and 52% of the cases [12]. With this low classification rate a formal diagnosis cannot rely on manual visual inspection by naked eye; therefore, biopsies od lesions are needed. The biopsy takes time and money, additionally pain and scars to the patient, which is not always needed. Different ML based approaches are employed for the Melanoma classification problem and skin disease classification in general

attempts include traditional models such as ANNs and SVM while two more recent studies employed fine-tuned VGG-16 Convolutional Nets, which moved the latter approach of this study [13].

1.2. Research Objectives

This work emphasizes the use of deep learning for skin lesson classification purpose. The primary goal of this research study is to offer a helping hand for dermatologists for diagnosing exact cancer type in less time. It is only possible to have better tools that can help in diagnosis. The skin lesion should have been well segmented and deep learning techniques are used in a way to get maximum accuracy from the model. Based on this the objectives for this research study can be formed. These research objectives are as follows:

RO1. To achieve accurate skin lesion segmentation.
RO2. To use Deep Learning Techniques for the identification of skin cancer and its type.
RO3. To find the accuracy of detection by using multiple layers of CNN.
RO4. To achieve more accuracy of predicting skin cancer type from skin lesion images.

1.3. Research Questions

These research objectives can be shaped into some research question, as shown in the Table 1:

Table 1. Research Questions.

Sr#	Research Questions
RQ1	Can we achieve better accuracy than human professionals?
RQ2	How helpful this model can be for dermatologists?
RQ3	Can this model outperform already available skin cancer detection models?

1.4. Skin Lesion Segmentation

A skin lesion is an abnormal single bounded region, which can be distinguished from the regular skin employing dissimilar color or texture. The skin lesion is the skin region of interest that can be used for further processing. Segmentation refers to isolating lesion (abnormal skin) from non-lesion (normal skin) [14]. Lesion segmentation performs a key role in analyzing dermoscopy images, as it identifies multiple lesion-specific morphological features and thus also identifies a confined region for local clinical feature segmentation that can be done at the advanced stage [Fig. 3]. The border or boundary of the segmented region also provides multiple features that can be used in the lesion analysis. A region of normal skin can also be recognized by correctly identifying the area with no lesion by overlooking artifacts existing in skin images. This part of even skin can be used for computing relative colors and many other significant features[15].

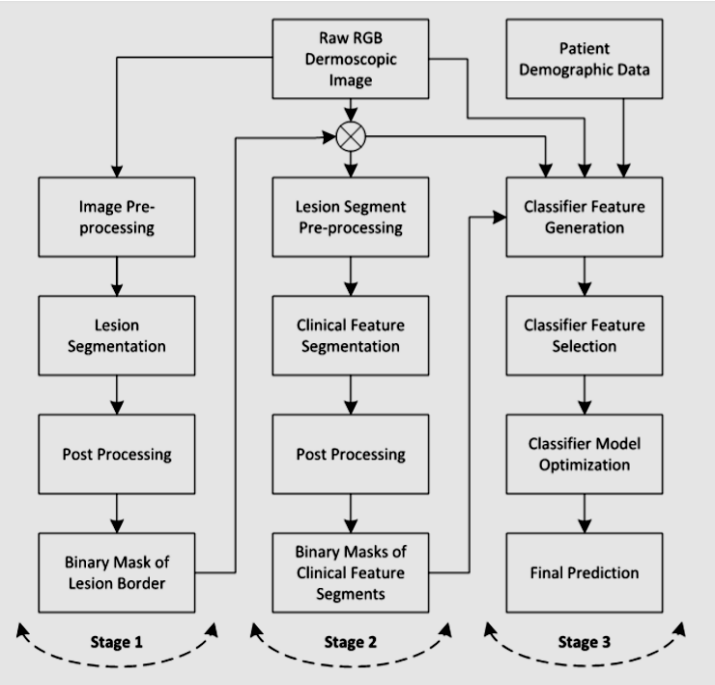

Fig.3. Detection of melanoma from dermoscopy images.

1.5. Organization of this Article

This article is divided in five more sections, the section 2 refers to the related work followed by the section 3 which explains materials and the methodology employed for this research study. In the 4th section the step-wise implementation of the algorithms on the skin cancer images is explained. The results and findings of this research experiment are discussed in fifth section. At last, this research is concluded in the 6th section.

2. Related Works

Skin cancer is a malignant disease that is causing numerous deaths globally. More than a million of skin care cases are reported worldwide every year. Diagnosis of skin lesion still remains a challenge; however, the early diagnosis increases the possibility of recovery[16]. Several techniques and approaches are used to help the clinicians to classify the skin lesion by automated systems. But the automatic differentiation between diverse types of lesion categories and further recommending the appropriate treatment is still a challenge [17].

Wang et al. proposed a method for health care applications to decrease the effect of confounders to generalized the performance of neural networks more efficiently[6]. The aim to design such a method is to accomplish a predictive model which is free of cofounding factor. The distinct feature of the method is that it is easily adoptable by any network model with minimal changes. The proposed method named "Confounder Filtering CF" was evaluated with four different healthcare data sets, which includes heart MRA/CT of the lung, EEG brainwave data and brain. All the experiments proved significant improvement in the accuracy of prediction by using the neural networks where the method is adopted. The only limitation of the method is its repeated training process.

Romero-Lopez designed a solution to assist the dermatologists in diagnosing the skin lesions [18]. A two-class classifier is implemented that inputs the images of skin lesion categorized as malignant or benign, builds a model using VGGNet, convolutional neural network architecture. This model is then finally used against the unseen image of skin lesion to predict whether the image belongs to benign category, or it is malignant. Their experiment shows 76.74% precision.

Marvdashti et al. reported first and fully automated detection of a most common kind of skin cancer, basal carcinoma (BCC), in human skin [19]. PS-OCT (Polarized Sensitive Coherence Tomography) is used. The study proposed an ML classifier based on image features, extracted from image contrasts that were presented by PS-OCT. The proposed machine learning classifier was, however, built by extracting new features as well as the previously proposed features and gives an exceptional performance of 95.4% accuracy to detect BCC in human skin.

Haenssle et al. compared man against machine [20]. They compared the diagnostic performance of Google Inception CNN over a group of 58 international dermatologists including 30 skin experts. The results were significantly unique, as CNN overpowered most of the dermatologists.

Masood [21] proposed an automated self-advised and semi-supervised novel learning model to diagnose skin cancer by recognizing melanoma using images. A deep neural network is constructed using labelled and unlabeled data samples along with applying exponential loss function to enhance the classification of labelled data.

In [22], it demonstrated a real-world medical diagnosis application. An active approach for high dimensional data, based on the advancement of the intersection of deep learning and Bayesian model for active learning of image data is presented. The performance of techniques was assessed and showed that even with long running time, the technique reduces costs and running time.

In [23], the authors addressed two fundamental classification problems of skin lesion, which includes use of high-res images alongwith standard image classification architectures and the other includes the higher imbalance of classes in real-world datasets. The former problem is solved by proposing an architecture based on patch attention. The later problem is solved by comparing class-specific loss weighting, oversampling and batch sampling.

In [24], the authors summarize the automated detection of skin cancer, melanoma in dermoscopy images. Formally, segmentation of skin lesion followed by clinical feature segmentation is discussed. Later, the existence of skin cancer melanoma is predicted by applying machine learning algorithms and techniques to attributed produced from segment features.

Kalouche proposes using deep learning algorithms to detect skin cancer, explicitly melanoma, from photos of skin wounds captured with a regular camera [12]. The dataset was trained on three different learning models, each with increasing accuracy. Deep Learning Model; Logistic regression; a fine-tuned, pre-trained VGG-16 Convolutional Neural Networks were among the three models. Using a fine-tuned VGG-16 CNN, they demonstrated the developed algorithm's capacity to separate moles from images with 70 percent accuracy and diagnose skin lesions as melanoma with 78 percent balanced accuracy.

In [17], an ensemble-based approach is presented to differentiate different seven types of skin categories. The deep learning approach has been introduced to help the dematologists in categorizing multiple skin types. The approach used a combination of two architectures i.e. Inception V3 and ResNet-50, for the classification of the lesion. This approach involves no prior segmentation, grey-scale conversion or pre-processing and shows significantly higher results as compared to other approaches used for lesion classification.

In [25], the authors use CNN (convolutional neural nets) to demonstrate the classification of pigmented skin

lesions. The neural network is trained against a dataset of 129,450 images and inputs disease labels and pixels. The images include clinical images of 2032 different diseases. The performance of Convolutional Neural Network is tested against 21 dermatologists certified images proven by biopsy, over two use cases. Hence, the effectiveness of deep learning in the area of dermatology is demonstrated.

In [26], the authors explored the collective use of OCT (optical coherence tomography) features from actinic keratosis (AK) and basal cell carcinomas(BCC). AK and BCC lesions are studied in 34 patients. The diagnosis accuracy was evaluated using a machine learning tool. The results recommends that 73% (AK) and 81% (BCC) accuracies of classification are accomplished, whenever multiplicity of features is involved. They proposed that diagnosis data can be extracted with reasonable accuracy by using OCT images architecture if applied in combination.

3. Material and Methods

This work uses the ISIC 2018 labelled skin lesion dataset to investigate a categorization of seven different forms of skin cancer, as stated in the aims and introduction. Classifiers are trained for this three-class classification job using the labelled dataset, which includes seven different skin lesion classes, from which binary (melanoma/non-melanoma, as illustrated in Fig. 4) classification results are derived. To ensure unbiased testing, a labelled test set is kept always separate. All the tests are done in Keras using Tensorflow as the backend. In terms of pretraining, three primary frameworks are investigated: two in the transfer learning line and one involving the use of unlabeled data via auto-encoders. Along with the three frameworks, techniques for dealing with overfitting and class imbalance are investigated.

3.1. Material

The HAM10000 ("Human Against Machine with 10000 images") dataset is used for the classification [27]. It contains 10015 dermoscopic images, available publically in the ISIC archive, for academic purposes. It contains the following attributes; lesion_id, image_id, ge, sex, localization and dx type: Technical Validation field (ground truth).

This dataset is newly composed and contains 10,000 images of skin lesions. Training with data can help the model to learn about the 7 categories of skin lesions. Thus, it leads to better training and accuracy of results than other skin lesion datasets.

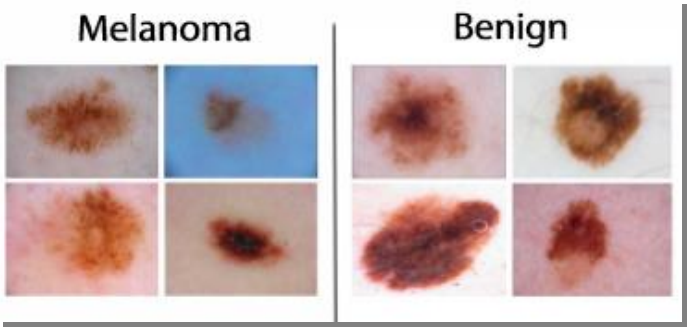

Fig.4. Sample images from HAM 1000 dataset.

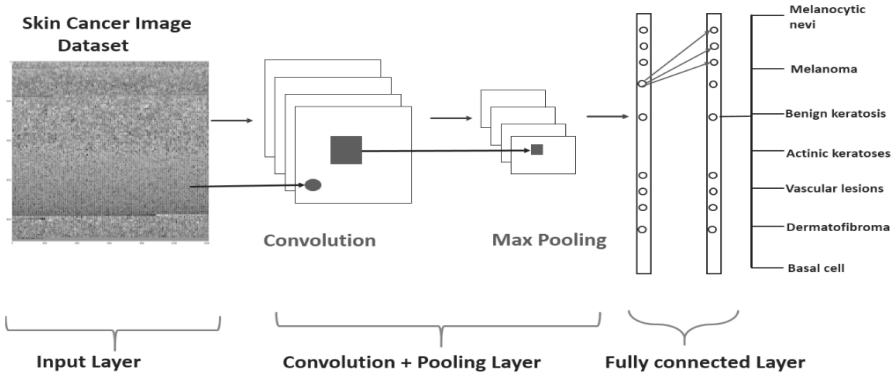

Fig.5. Demonstration of classification of dermatoscopic images.

From this dataset, it can be extracted that it has seven classes of skin diseases, listed below:

- Basal cell carcinoma
- Actinic keratoses
- Melanocytic nevi

- Melanoma
- Dermatofibroma
- Benign keratosis-like lesions
- Vascular Lesions

3.2 Method

Skin photos are classified into 7 different classifications of skin cancer in this study utilizing Convolution Neural Networks using Keras and TensorFlow in the backend, and the results are then analyzed to evaluate how the model might be beneficial in a practical scenario, as shown in Fig. 5. This research went through a step-by-step method to classify seven types of cancer. The following are the steps that are charted when constructing and evaluating a model:

- Reading and preprocessing Data
- Exploratory data analysis (EDA)
- Images loading & resizing
- Train-test set split and normalization
- Model building (CNN)
- Setting optimizer & annealing
- Fitting the model
- Model Evaluation (validation of confusion matrix, accuracy, precision, recall, support and analysis of misclassified instances etc.)

4. Implementation

4.1. Reading & Preprocessing Data

The image dictionary was made by merging the folder path from base_skin_dir (base directory) and then merging the .jpg images from both parts of dataset "HAM10000_images_part1" and "HAM10000_images_part2". Following libraries are used for this classification and diagnosis:

- Pandas
- Keras
- TensorFlow
- Glob
- sklearn
- Matplotlib

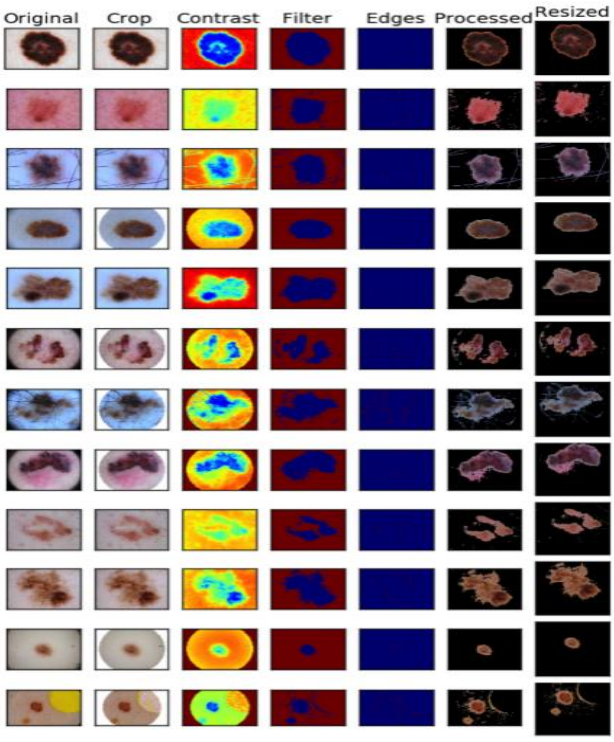

Fig.6. Demonstration of pre-processing, lesion segmentation, and formatting of some images.

The csv is read by linking the path to the image folder, which is designated base_skin_dir and contains all of the images. Following that, we created some new columns that are easy to understand for future reference, such as column path, which contains the image id, cell type, which contains the short name of the lesion type, and finally, the column cell type idx, in which the lesion type is divided into codes ranging from 0 to 6. Then each field is examined for missing values and data types.

4.2. Exploratory Data Analysis (EDA)

We explored the diverse features of the dataset, actual counts plot and distribution to analyze a distribution of seven distinct classes of cell types (Fig. 7.), in this stage. It seems, cell type Melanocytic nevi has an enormous number of samples in contrast to other types.

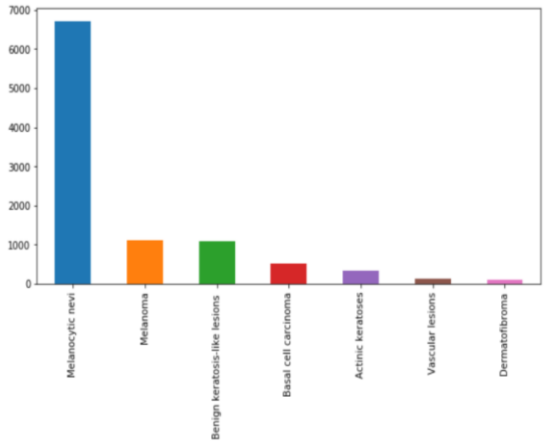

Fig.7. The distribution of target classes of cell type.

Plotting of Technical Validation (ground truth) i.e., dx_type to insight the distribution of its four classes (summarized in Fig. 8) named Histopathology, Confocal, Follow-up, Consensus.

1. Histopathology (History): Histopathologic diagnoses of abrasions, done by specialised dermatopathologists.
2. Confocal: This microscopy, is a novel imaging technique that provides noninvasive, in-vivo imaging of skin. It comes with several facial benign with a grey-world assumption and a resolution approaching the cellular level. These training images remain in Lab-color space prior to and after the manual histogram changes.
3. Follow-up: Biologists consider it as indication of biologic benignity if no change is monitored by digital Dermatoscopy in nevi even after 1.5 years or 3 follow up visits. As dermatologists don't monitor vascular lesions, dermatofibromas or seborrheic keratosis, therefore they can't be labelling any other except nevi, with such ground-truth.
4. Consensus: For distinctive benign cases, biologists provide ratings by consensus of experts of authors HK and PT. This label is applied only if both authors gave the same clear benign diagnosis. Lesions with this dx-type did not need further biopsy or follow-up.

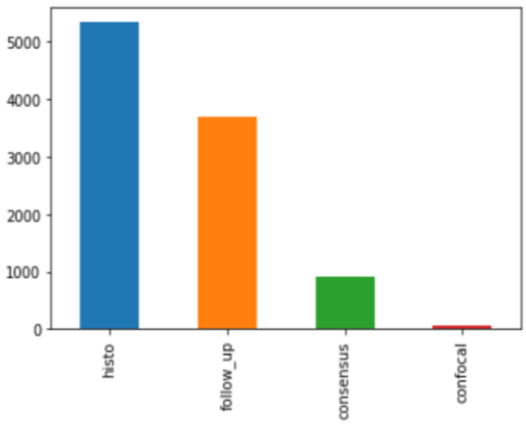

Fig.8. Scheming of Technical Validation (ground truth), dx_type to analyze distribution of its four types.

Plotting the distribution of localization field is done to identify which body part is most affected by the skin lesion. It seems lower extremity, back, upper extremity, and are usually affected regions of skin cancer as compared to neck and scalp or ears.

When the distribution of age is analysed, it seems that there are larger instances of patients are from the age group 30 to 60 years and a relatively small number of cases have occurred till the age of 20 years (represented by Fig. 10.).

The distribution of males and females is visualized in the age-wise distribution of skin cancer types. And it seems that males are more affected by skin lesions than females. Apparently, males are mostly affected by skin cancer as compared to female (as depicted in Fig. 11.).

The data is visualised age-wise for distribution of skin cancer types, and it provided an insight that Melanocytic nevi, Basal cell carcinoma, dermatofibroma, and Vascular lesions, are not much prevalent below the age of 20 years (as shown in Fig. 12.).

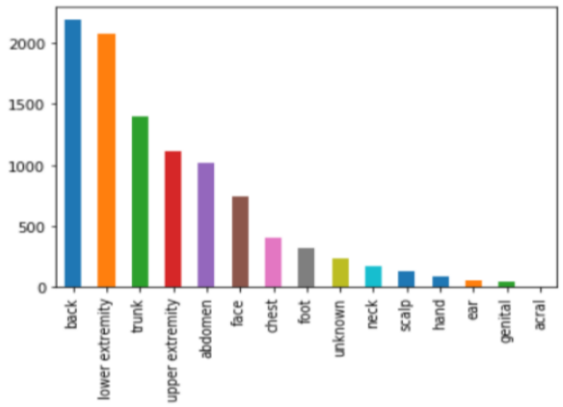

Fig.9. Plotting of localization of pigmented skin lesion.

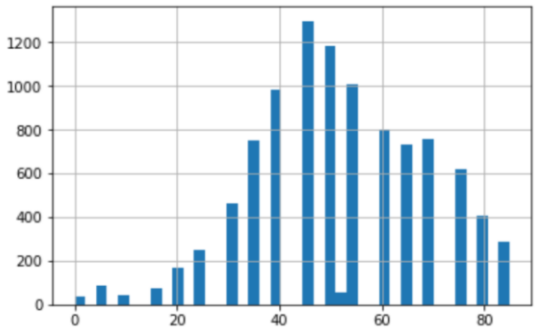

Fig.10. Plotting of Skin lesion occurrence over the attribute of age.

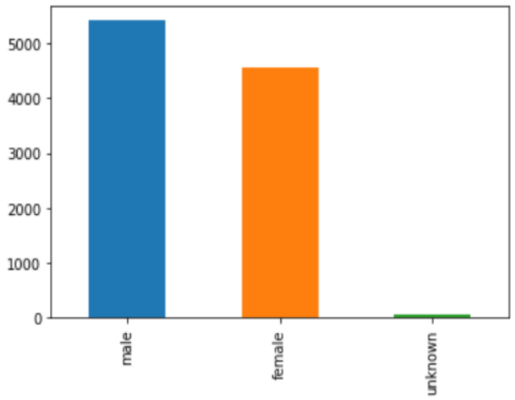

Fig.11. Distribution of males and females is plotted.

4.3. Loading and Resizing of Images

Images from the folder's image path were loaded into the column named image in this phase. Then we downsized the photos because their original dimensions are 450 x 600 x3, which TensorFlow could not process quickly, thus we resized the images to 100 x 75 pixels. After resizing 10015 photos to 100 x 75 pixels a sample of photos is photographs of each cancer type below.

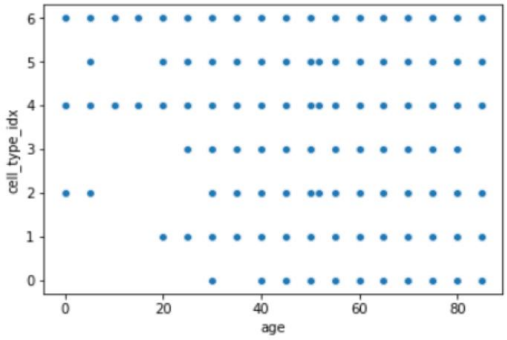

Fig.12. Age-wise for distribution of skin cancer types.

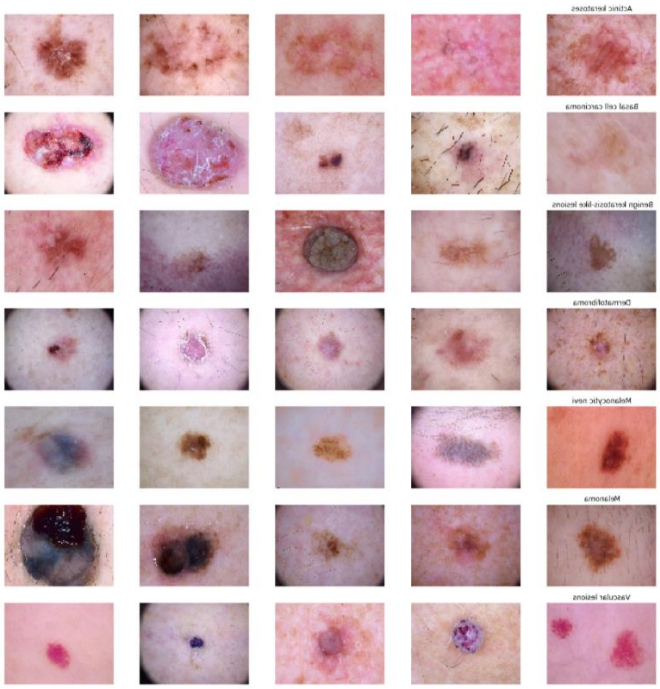

Fig.13. Images of seven different types of skin cancers.

4.4. Train Test Split and Normalization

For reliability of the results produced from the model implementation it is necessary to test the model. By testing we can find if the model is producing accurate and precise results. The dataset is divided into two sets: a training set (x train) for model learning and a test set (x-test) for model validation, with a ratio of 80% and 20%, respectively. The x train and x test are then normalised by subtracting their mean values from their SD (standard deviation). The data is labelled with seven different classes of skin cancer types ranging from 0 to 6, as seen in Table 2 and Fig. 13. These labels must be encoded into a single vector.

Table 2. Target Classes arranged according to their respective codes.

Code	Class
0	Melanocytic nevi
1	Melanoma
2	‘Benign keratosis-like lesions
3	Basic cell carcinoma
4	Actinic keratosis
5	Vascular lesions
6	Dermatofibroma

4.5. Model Building

Convolutional Neural Networks: The first is the convolutional layer (Conv2D) similar to learnable filters set. A set

of 32 filters for the two conv2D layers and last two having 64 filters. An area of the image is transmuted by the filter using the kernel filter, which is functional on the whole image. CNN can separate features that are useful from these transformed images.

Pooling layer (MaxPool2D) is the second layer in CNN, which also works as down-sampling filter, picks the maximal value from 2 neighboring pixels. This reduces the computational cost and the overfitting to some context. The pooling size is then selected if the pooling dimension is high, then down sampling becomes more important.

CNN can combine local features and determine global aspects of images by combining the pooling and convolutional layers. Dropout is a regularization approach. Some of the network is dropped at random, which aids in the extraction of features in a dispersed pattern if a percentage of nodes in each layer is disregarded at random for each sample. This method also improves generalization while reducing overfitting. The rectifier (activation function, max(0,x)) is called 'relu.' This rectifier activation function is used to make the network non-linear.

To convert the final feature plots into a single 1D vector, the Flatten layer is used. This flattening phase allows fully linked layers to be used after some convolutional/max pool layers. All of the local features of other levels are syndicated.

This categorization is done with the Keras Sequential API (where you only add one layer at a time, starting with the input). Finally, the features from the two dense (completely linked) layers (ANN classifier) are employed. The net produces the probability distribution of the class in the last layer (Dense (10, activation="softmax")).

4.6. Setting Optimizer and Annealer

Following the addition of the above layers to the model, an optimization method, as well as a score and loss function, is created. The model's poor performance on photos with known labels is determined by the loss function. The error rate between observed and expected labels is what defines it. The "categorical cross entropy" is a special form for categorical classifications with more than two classes. The optimizer, as the most important function, improves parameters such as weights and filters kernel values in order to reduce loss. For this study, the Adam optimizer is utilized because it combines the benefits of two independent versions of the stochastic gradient descent technique.

AdaGrad: AdaGrad (Adaptive-Gradient Algorithm) maintains the learning rate according to the gradient value of the independent variable, and thus it avoids manual tuning of the learning rate[28].

RMSProp: RMSProp (Root Mean Square Propagation) maintains learning rates for each parameter that are improved according to the average of current degrees of the gradients for the weight[29]. Hence, the algorithm can perform well on non-stationary and online problems (e.g. noise).

Adam: Adam combines the advantages of both algorithms, AdaGrad and RMSProp, and as a result, Adam is widely used since it produces outstanding results quickly.

Our model's performance is further assessed using the metric function "accuracy." The metric function is very similar to the loss function; however, the metric evaluation results are only used to evaluate the model; they are not utilized to train the model.

We used annealing method of the learning rate (LR) is used in this study to make the optimizer converge quicker and nearest to the global minimum of the loss function. The LR is the step through which the optimiser walks through the 'loss landscape'. Higher the LR, larger are the steps, and the more rapidly is the convergence. However, the sampling is weaker with a high LR, and perhaps the optimizer could drop into local minima.

It's better to have less learning rate during the training of the model to reach the global minimum of the loss function efficiently. To keep the advantage of the fast computation time with a high LR, we decreased the LR dynamically every X step (epochs) depending on if it is necessary (when accuracy is not improved). With the Reduce LROnPlateau function from Keras. Callbacks, I choose to reduce the LR by half if there is no improvement in accuracy after 3 epochs.

Data Augmentation: It's an optional step to avoid the overfitting problem. To solve HAM 10000, the dataset is artificially extended. This may result in the present dataset becoming even larger. The goal is to alter the training set in order to replicate the variations. Data augmentation strategies update the array representation of the training set while keeping the label the same. Vertical and horizontal flips, grayscales, random cropping, translations, colour jitters, and rotations are all common augmentations. We can easily double or quadruple the number of training examples and develop a very robust model by applying only a few of these adjustments to our training data.

We selected to randomly rotate some training photos by 10 degrees, randomly zoom certain training images by 10%, and then randomly move images horizontally by 10% of the width for data augmentation. Similarly, photos are shifted vertically by 10% of their height at random. The next stage is to fit the model for further processing after it has been created.

4.7. Fitting the Model

In this step, the model is fit into x_train, y_train and batch size of 10 and 50 epochs can be executed as small as batch size. To make it more efficient, 50 epochs are executed to give the model sufficient epochs to train.

4.8. Model Evaluation

This stage involves checking the model's testing and validation accuracy, plotting a confusion matrix, and verifying the number of misclassified images. Because of the problems with other classification systems, the study used

accuracy as its primary metric throughout the research. For testing purposes, the test data consists of more than 10015 photos. The confusion matrix, which displays how much predicted test data falls under the same results as the actual class, is used to evaluate a classifier's performance. It can also be quantified by the number of test data records sorted incorrectly into target classes. The data provided by the Confusion Matrix (Table 3) aids in the identification of evaluation measures, such as accuracy, precision and recall (represented in table 4) etc.

Table 3. Confusion Metrics.

		PREDICTED CLASS	
ACTUAL CLASS		P	N
	P	TP (True Positive)	FN (False Negative)
	N	FP (False Positives)	TN (True Negatives)

Accuracy: Accuracy being an intuitive performance measure, signifies how much correct predictions are identified from the total number of predictions.

Precision: Precision represents the ratio of correctly predicted observations to the total number of predicted observations.

Recall: Recall signifies the ratio of all positive observations that are predicted correctly to all the observations in the actual class.

F1-Score: The F1 score identifies "a weighted average of the precision and recall, where an F1 score reaches its best value at 1 and the worst score at 0". F1 score from 0 to 1 can illustrate a 0 to 100% measure.

Table 4. Evaluations Measures.

Evaluation Measure	Formulae	Evaluation Measure	Formulae
ACCURACY	(TP + TN)/(TP + TN+ FP + FN)	RECALL	(TP) / (TP + FN)
PRECISION	*(TP) / (TP + FP)*	F1-SCORE	*(2 x PRECISION x RECALL) / (PRECISION + RECALL)*

5. Results

Implementing CNN for the cancer data set provides higher accuracy in skin cancer diagnosis. This accuracy can be increased in future by obtaining more data and training the model with a larger data. By increased accuracy of skin disease, the medical community can easily produce the diagnosis of skin lesion. However, it appears that the model has a maximum number of misclassified predictions for Code 3 i.e. Basal cell carcinoma, Vascular lesions code 5 and Melanocytic nevi code 0 whereas, Actinic keratoses code 4 has least misclassified type (Fig 14.). We can also further tune our model to straightforwardly attain the accuracy above 80%, and it can still be claimed that this model is efficient in comparison to detection with human eyes having 79% accuracy as shown in Fig. 15. The same result is shown in the confusion metrics as 04 class is accurately classified (Fig.16).

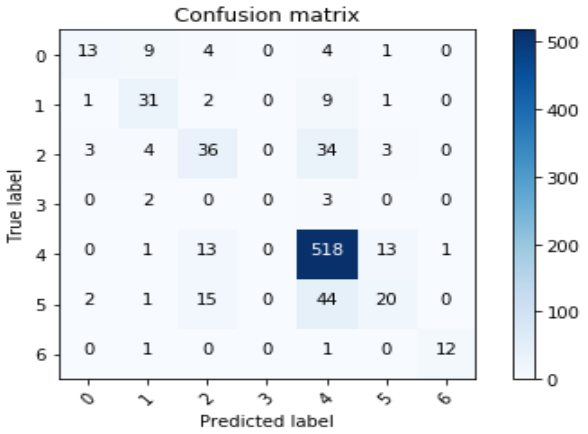

Fig.14. Representation of Model results in Confusion Metrics.

Similarly, the class with code 6 is also classified, class 3, 5 and 0 are most of the misclassification as compared to the other four classes (as shown in Fig.16).

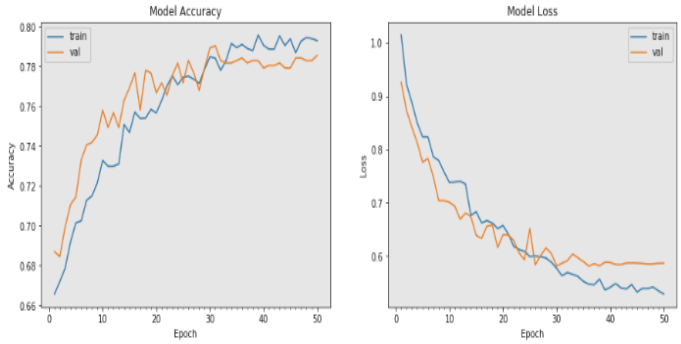

Fig.15. Representation of the model according to the epochs.

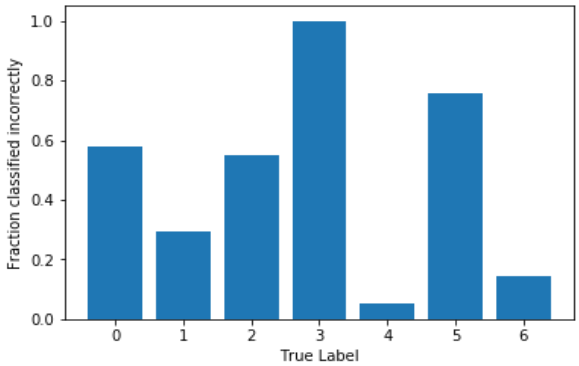

Fig.16. Fraction of misclassification against true labels.

5.1. Evaluation Measures Calculated

The data represented using Confusion Matrix helped to identify other evaluation measures too, such as Precision, Recall, F1 Measure and support etc. These measures are as follows:

Table 5. Evaluation of the classification done by CNN model.

Classes	Precision	Recall	F1 score	Support
nv	0.60	0.39	0.47	31
mel	0.60	0.59	0.60	44
bkl	0.51	0.45	0.48	80
bbc	0.00	0.00	0.00	5
akiec	0.84	0.95	0.89	546
vasc	0.65	0.32	0.43	82
df	0.91	0.71	0.80	14

6. Conclusions

This research exploration underlines the utilization of profound learning for skin malignant growth grouping reason. The primary goal of this examination is to give some assistance to dermatologists for diagnosing definite malignant growth type quicker than expected. For this reason, skin pictures are characterized into 7 distinct classes of skin malignant growth utilizing Convolution Neural Network. From that point forward, another model is executed which contain CNN with optimization and annealing of the outcome to see which the model can be helpful in a useful situation. It is providing better diagnosis and suitable for quick diagnosis before autopsy. It can be furthermore trained upon acquisition of more skin lesion data and we can acquire more accuracy. As it appears to has some misclassification for Basal cell carcinoma, Vascular sores code 5 and Melanocytic nevi, code 0. But Actinic keratosis code 4 has least misclassified type and for other remaining classes, our model produces good accuracy and precision. We can likewise furthermore tune our model to effortlessly accomplish the accuracy above 80%, and still this model is proficient in contrast with location with natural eyes having 79% accuracy. In the future, we are intending to update the model to get better accuracy for the diagnosis of skin diseases.

References

[1] R. H. Taylor and D. Stoianovici, "Medical Robotics in Computer-Integrated Surgery," IEEE Trans. Robot. Autom., vol. 19, no. 5, pp. 765–781, 2003.
[2] G. Long, Y. Tan, J. Jiang, and C. Zhang, "Federated Learning for Open Banking," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 12500 LNCS, pp. 240–254, Aug. 2021.
[3] S. Ramaswamy, R. Mathews, K. Rao, and F. Beaufays, "Federated Learning for Emoji Prediction in a Mobile Keyboard," 2019.
[4] O. Ardhapure, G. Patil, D. Udani, and K. Jetha, "Comparative Study of Classification Algorithm for Text Based Categorization," Int. J. Res. Eng. Technol., pp. 217–220, 2016.
[5] M. Ahmed and A. K. M. Najmul Islam, "Deep Learning: Hope or Hype," Ann. Data Sci., vol. 7, no. 3, pp. 427–432, Sep. 2020.
[6] H. Wang, Z. Wu, and E. P. Xing, "Removing Confounding Factors Associated Weights in Deep Neural Networks Improves the Prediction Accuracy for Healthcare Applications."
[7] F. Altaf, S. M. S. Islam, N. Akhtar, and N. K. Janjua, "Going Deep in Medical Image Analysis: Concepts, Methods, Challenges and Future Directions."
[8] A. S. Adamson and A. Smith, "Machine Learning and Health Care Disparities in Dermatology," JAMA Dermatology, vol. 154, no. 11, p. 1247, Nov. 2018.
[9] J. Premaladha and K. S. Ravichandran, "Novel Approaches for Diagnosing Melanoma Skin Lesions Through Supervised and Deep Learning Algorithms," J. Med. Syst., vol. 40, no. 4, pp. 1–12, 2016.
[10] Y. Khan, Y. Khan, and Y. M. Student, "AI , are dermatologists ' days numbered ?," pp. 1–11.
[11] B. Monien, R. Preis, and S. Schamberger, "ImageNet Classification with Deep Convolutional Neural Neural Networks," Handb. Approx. Algorithms Metaheuristics, pp. 60-1-60–16, 2007.
[12] S. Kalouche, "Vision-Based Classification of Skin Cancer using Deep Learning."
[13] H. Liao, "CSC 400-Graduate Problem Seminar-Project Report A Deep Learning Approach to Universal Skin Disease Classification."
[14] R. J. Friedman, D. S. Rigel, and A. W. Kopf, "Early Detection of Malignant Melanoma: The Role of Physician Examination and Self-Examination of the Skin," CA. Cancer J. Clin., vol. 35, no. 3, pp. 130–151, 1985.
[15] Y. Cheng et al., "Skin lesion classification using relative color features," Ski. Res. Technol., vol. 14, no. 1, pp. 53–64, 2008.
[16] M. Emre Celebi, Q. Wen, H. Iyatomi, K. Shimizu, H. Zhou, and G. Schaefer, "A state-of-the-art survey on lesion border detection in dermoscopy images," Dermoscopy Image Anal., no. September, pp. 97–130, 2015.
[17] A. H. Shahin and A. Kamal, "Deep Ensemble Learning for Skin Lesion Classification from Dermoscopic Images," 2018 9th Cairo Int. Biomed. Eng. Conf., pp. 150–153, 2018.
[18] A. Romero-Lopez, X. Giro-i-Nieto, J. Burdick, and O. Marques, "Skin Lesion Classification from Dermoscopic Images Using Deep Learning Techniques," in Biomedical Engineering, 2017.
[19] T. Marvdashti, L. Duan, S. Z. Aasi, J. Y. Tang, and A. K. Ellerbee Bowden, "Machine-learning detection of basal cell carcinoma in human skin using polarization sensitive optical coherence tomography," vol. 7, no. 9, p. JM4A.5, 2016.
[20] H. A. Haenssle et al., "Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists," Ann. Oncol., vol. 29, no. 8, pp. 1836–1842, Aug. 2018.
[21] A. Masood, A. Al- Jumaily, and K. Anam, "Self-supervised learning model for skin cancer diagnosis," in 2015 7th International IEEE/EMBS Conference on Neural Engineering (NER), 2015, pp. 1012–1015.
[22] Y. Gal, R. Islam, and Z. Ghahramani, "Deep Bayesian active learning with image data," Proceedings of the 34th International Conference on Machine Learning - Volume 70. JMLR.org, pp. 1183–1192, 2017.
[23] N. Gessert et al., "IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING, PREPRINT 1 Skin Lesion Classification Using CNNs with Patch-Based Attention and Diagnosis-Guided Loss Weighting," 2017.
[24] N. K. Mishra and M. E. Celebi, "An Overview of Melanoma Detection in Dermoscopy Images Using Image Processing and Machine Learning," Jan. 2016.
[25] A. Esteva et al., "Dermatologist-level classification of skin cancer with deep neural networks," Nature, vol. 542, no. 7639, pp. 115–118, Feb. 2017.
[26] T. M. Jørgensen, A. Tycho, M. Mogensen, P. Bjerring, and G. B. E. Jemec, "Machine-learning classification of non-melanoma skin cancers from image features obtained by optical coherence tomography," Ski. Res. Technol., vol. 14, no. 3, pp. 364–369, 2008.
[27] P. Tschandl, C. Rosendahl, and H. Kittler, "The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions," Sci. Data, vol. 5, p. 180161, Aug. 2018.
[28] A. Agnes, L. #1, and F. Sagayaraj Francis, "Adagrad-An Optimizer for Stochastic Gradient Descent."
[29] T. Kurbiel and S. Khaleghian, "Training of Deep Neural Networks based on Distance Measures using RMSProp."

Authors' Profiles

Momina Shaheen is Lecturer of Computing, in University of Roehampton, London, United Kingdom. She is a Ph.D Scholar of University of Management and Technology, Lahore, Pakistan. Her research interests are Machine Learning, Data Science, Artificial Intelligence, Agent-based Modeling, Cognitive Sciences and Distributed Systems.

She has more than 6 years of teaching and research experience in Higher Education Institutes. She has supervised and initiated a number of projects in her career. She earned her master's in software engineering,

from Bahria University Islamabad Campus in 2016. She did her bachelor's in information technology from Islamia University Bahawalpur, Pakistan.

Usman Saif is currently a lead Software Engineer at Codility Solutions. He remained a part of multiple projects, completed in this organization. He completed his master's in software engineering from University of Management and Technology, Lahore, Pakistan in 2019 and worked in multiple data science projects throughout his masters and career. He completed his Bachelors in Software Engineering from UMT, Lahore in 2016.

He did his thesis in Deep Learning, Machine Learning and Artificial Intelligence domain. Some of his other areas of interest are Natural Language Processing and Robotics.

Shahid M. Awan is Assistant Professor/ Lecturer (Data Analytics) in University of West of Scotland, United Kingdom. He is a well-experienced data scientist thought leader with a demonstrated working history in the information technology and services industry. He is skilled in Machine Learning, Natural Language Processing, Computer Vision, IoT, Real-Time Analytics, and AI Solution design and implementation over the edge and cloud systems.

He has worked with textual, numerical, imagery and video data sets. He had carried out industrial projects on numerical forecasting for consumer electric demand, sales, solar energy generation, and disease spread forecasting. He applied natural language processing algorithms for text summarization, text generation, named entity resolution, chatbot development, text classification, sentiment analysis, and social media mining.

Faizan Ahmad is an academic from a computer science background with a total of 11 years of teaching and research experience, specializing in Human-Computer Interaction (HCI), motivated to serve in assistive and education technologies by proposing next-generation gamification solutions.

I have made significant contributions to HCI and published 14 high-quality research articles (including 8 as a first author) in prestigious journals and conferences such as Smart Learning Environments (SMLE), Interactive Learning Environments (NILE), International Journal of Games-Based Learning (IJGBL), ACM PRICAI, and OzCHI. So far, I have had the opportunity to supervise 17 bachelor's, 23 master's, and 01 doctoral dissertations.

Aiman Anum is working as Ad-hoc Lecturer in Islamia University Bahawalpur, Bahawalnagar Campus, Punjab Pakistan since 2019. She is working with research group of distributed computing in Islamia university of Bahawalpur, Bahawalnagar campus. She has published 4 journals articles in reputed computer science journals. Her research interests are Machine Learning, Federated Machine Learning and Data Science.

She received her MS Computer Science degree from Comsats University Islamabad Sahiwal Campus. Her bachelor's in computer science is from Islamia University Bahawalpur, Bahawalnagar Campus, Pakistan.

How to cite this paper: Momina Shaheen, Usman Saif, Shahid M. Awan, Faizan Ahmad, Aimen Anum, "Classification of Images of Skin Lesion Using Deep Learning", International Journal of Intelligent Systems and Applications(IJISA), Vol.15, No.2, pp.23-36, 2023. DOI:10.5815/ijisa.2023.02.03