IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.DOI

Skin Cancer Classification using Convolutional Neural Networks

MD. H. ZAMAN[1], JUMANA AHMED[2], AND ZANNATUL M. TURIN.[3]
[1]North South University, Bashundhara, Dhaka-1229, Bangladesh (e-mail: registrar@northsouth.edu)
[2]Department of Physics, North South University, Bashundhara, Dhaka-1229, Bangladesh
[3]Electrical and Computer Engineering Department, North South University, Bashundhara, Dhaka-1229, Bangladesh

ABSTRACT Skin cancer is the abnormal growth of skin cells and is visible on the outer layer of the skin, this happens mainly to skin that has been exposed to the sun. The incidence of skin cancer is increasing at an alarming rate. Morbidity and mortality rates of skin cancers are increasing and, therefore, pose a significant public health concern. That is why it has become necessary to diagnose skin cancer early to detect the rapid rate of growth of skin cancer, its high treatment costs, and mortality rate. In most cases, these cancer cells are detected manually and it takes time to cure. Therefore, several computer vision methods are introduced lately, which not only aids early detection and diagnosis of various types of skin cancer, but are also cost-effective and fairly accurate. This paper proposed an artificial skin cancer detection system using image processing and machine learning method. The features of the affected skin cells are extracted after the segmentation of the dermoscopic images using feature extraction technique. A deep learning based method convolutional neural network is used to identify skin cancer presence in patients.
Convolutional neural networks (CNNs) are a branch of deep learning which have been turned into one of the popular methods in different applications, especially medical imaging. One of the significant applications in this category is to help specialists make an early detection of skin cancer in dermoscopy and can reduce mortality rate. Image processing is applied on raw images before feeding them to the neural network. However, there are a lot of reasons that affect system diagnosis accuracy.

INDEX TERMS challenge, Convolutional Neural Network, dataset, deep learning, HAM10000 dataset, image processing, Medical Image Analysis, skin cancer, Skin Lesion Analysis.

I. INTRODUCTION

SKIN cancer is defined as the abnormal growth of skin cells that is most often developed on skin exposed to the sun, primarily the areas of the scalp, face, lips, ears, neck, chest, arms, hands, and legs. It affects people of all skin tones, including those with darker complexions. It represents a major public health concern as it affects many people around the world, where it is the most common of all cancer types.

In recent years, with the advancement of technology, particularly artificial intelligence, suitable methods have been developed for this issue. In the meantime, image processing techniques are progressing as successful methods. Automated and accurate analysis could be done via machine learning (ML), where the application of image processing and computer vision for automatically identifying the patterns like cancer from images reduces human errors and increases the speed of detection. In addition, the importance of medical image processing can be considered as it helps physicians and radiologists to more easily diagnose the disease, thus protecting the patient against irreparable risks that will come about.

In this paper, Convolutional Neural Networks (CNN) were utilized to accurately classify pigmented skin lesions in dermoscopic images to detect the malignant skin lesions such as melanoma, the deadliest type of skin cancer, basal cell carcinoma, and some vascular lesions as early as possible. A CNN is a class of deep neural networks that use convolution in place of general matrix multiplication in at least one of their layers. It excels in analyzing visual imagery as they are fully connected (FC) feed forward neural networks that reduce the number of parameters very efficiently without losing out on the quality of models. Three different CNN models were employed to analyze and predict the pigmented lesions classes. The models used were Simple Sequential Analysis, InceptionV3 and ResNet152V2. Moreover, we wanted to also

apply VGG16 as a model, but unfortunately we could not as it was too much load for the device already so could not train further models. We also agreed on not going for imageNet, transfer learning, GoogleNet and AlexNet.

The models were trained and then compared using the metrics described in section II-F. This paper is organized as follows: section II details the evaluation metrics that were used for the models, as well as some basic information on how the models operate; section III contains information on similar research in this field. The data set used in this paper is "The HAM10000 dataset, a large collection of multi-source dermoscopic images of common pigmented skin lesions", and is described in the 'Data Set' section IV. Section V discusses the pre-processing methods performed on the data; section VI details the specifics of the models used; section VII discusses all of the model's performances; section VIII contains the final conclusions derived from the research; Lastly, section IX will discuss the future work and next steps for this project.

II. BACKGROUND

In this section we will discuss in detail about the required background information about CNNs, the different models used, and the comparison metrics applied to evaluate them. Dermatologists can enable lifesaving and fast diagnoses by identifying images of skin cancer, even outside the hospital via mobile application and software.

A. CNN

Remarkable progress has been made in image recognition, primarily due to the availability of large-scale annotated datasets and deep convolutional neural networks (CNNs). CNNs enable learning data-driven, highly representative, hierarchical image features from sufficient training data. A CNN is a class of deep neural networks that comprise of FC feed forward neural networks and utilize convolution in place of general matrix multiplication in at least one of their layers. This allows the network to reduce the number of parameters very efficiently without losing out on the quality of models, thriving in analyzing visual imagery. The network keeps on learning new higher dimensionality and more complex features with every layer. It comprises of several kinds of layers:

1) Convolutional Layer

Applies a filter that scans the whole image a couple of pixels at a time, producing a feature map for class probabilities predictions of each feature.

2) Pooling Layer

A layer that scales down the information produced from the convolution layer yet still maintaining the most essential information. Different types of pooling include max pooling and average pooling.

3) FC Input Layer

Flattens previous layer's outputs into one vector for use in the next FC layers.

4) FC Output Layer

Determines the image class by generating the final probabilities.

B. SIMPLE SEQUENTIAL ANALYSIS

A Sequential model is appropriate for a plain stack of layers where each layer has exactly one input tensor and one output tensor.

C. INCEPTIONV3

InceptionV3 is the third version roll out of the widely used Google's Inception CNN. It is a 48-layered model that was initially introduced during the ImageNet recognition challenge in 2015, where it was the 1st runner up. It attained an accuracy of 78.1% and greater in the challenge.

D. RESNET152V2

ResNet152V2 is a 152-layered residual neural network (ResNet) developed by Microsoft's research team. It builds on constructs known from pyramidal cells in the cerebral cortex, by skipping connections or jumping over layers. ResNet won the 1st place in the ImageNet recognition challenge in 2015, having achieved a 3.57 percent top-5 error. It focused on increasing the network depth rather than width, as well as reducing the effect of the vanishing gradient problem.

E. COMPARISON METRICS

Comparison metrics were employed to evaluate the models used and compare their performances. Five different metrics were used in this paper.

1) Confusion Matrix

A confusion matrix is a table that is used to describe the performance of a classification model on a set of test data for which the true values are known. It provides insights into the true and false positives and negatives of a model. The goal is to maximize the values of the diagonal of the matrix which encompass the true positives and true negatives, while minimizing the values that are off the diagonal that encompass the false positives and false negatives

2) Accuracy

The accuracy of a model is defined as the ratio of true predictions to total predictions. The higher the accuracy, the more true predictions the model produced for that class.

3) Recall

The recall of a model is defined as the ratio of true positives to the total true positives and false negatives. The higher the recall, the lower false negative predictions the model produced for that class.

4) Precision

The precision of a model is defined as the ratio of true positives to the total true positives and false positives. The higher the precision, the lower the false positive predictions the model produced for that class.

5) F1-Score

The F1-Score of a model is a balanced measure between precision and recall, utilizing harmonic mean to weigh the lower value heavier.

III. RELATED WORK

We searched the Google Scholar, IEEE Xplore and Web of Science databases for systematic reviews and original research articles published in English. Only papers that reported sufficient scientific proceedings are included in this review. Currently skin cancer is being detected by doctors by manually checking the pattern and area of the affected area. The diagnosis is a visual process, which relies on the long-winded procedure of clinical screenings, followed by dermoscopic analysis, and finally a histopathological examination. This process easily takes months and the need for many medical professionals and still is only about 77% accurate. Since this method is time consuming and is prone to human errors, research is being conducted to detect skin cancer automatically.

Image Processing techniques are extensively being used for the purpose. However due to limitations of Image Processing where certain parameters have to be sent manually, this approach for automated detection turns out practically to be inefficient. The need is to move towards an efficient and robust automated skin cancer classification system, which can provide highly accurate and speedy predictions. Thus it would be apt to train the machine to identify the cancer areas without using complex algorithms, which in turn increases the accuracy and efficiency of the system/algorithms.

Here are some of the projects' literature review:

used the MobileNet model from the HAM10000 dataset, and received an overall accuracy of 83.1%.

Achieved a 92.9 percent accuracy with fast.ai and using ResNet34. This is 2.9% lower than the ISIC 2018 challenge obtained with a considerably more complex approach, employing a machine learning algorithm on top 18 different neural networks.

With another project the accuracy was found to be 90.5% with ResNet model and 78% with VGG16 model, while with other models the accuracies were way lower around 65%.

A modified VGGNet was also utilized, where the classification of melanoma versus nevi or lentigines was addressed using dermatoscopic images. The authors compared the classification accuracy of i. a CNN trained from scratch, ii. a pre-trained CNN with transfer learning and frozen layers, iii. and a pre-trained CNN with transfer learning and fine-tuning of the weighting parameters. All three configurations were tested with 379 images from the ISBI 2016 Challenge dataset, and the last-mentioned configuration achieved the highest accuracy of 81.33%.

Mobile Net v2: 81.9%, Inception v3: 84.3%, ResNet50 v2: 81.3%, ResNet 152V2: 83.9% The top accuracy achieved on dilated versions of VGG16, VGG19, MobileNet, and InceptionV3 is 87.42%, 85.02%, 88.22%, and 89.81% respectively. Dilated Inception V3 achieved the highest classification accuracy. Mobile Net also has high classification accuracy while having the lightest computational complexities.

IV. DATA SET

As mentioned earlier, the data set used in this paper was "The HAM10000 dataset, a large collection of multi-source dermoscopic images of common pigmented skin lesions". It consisted of 10,015 dermoscopic skin pigmented lesion 600 by 450 pixel images, digitized and stored as JPEG images. They were initially manually cropped and centered around the lesion, as well as adjusted to enhance visual contrast and color reproduction. With each image and patient, the data set included 7 features, which were:

- Patient's Age [age]
- Patient's Sex [sex]
- A Lesion ID [lesion id]
- An Image ID [image id]
- A Technical Validation Field Type [dx type]
- The Localization of the Skin Lesion [localization]
- A Diagnostic Skin Lesion Category [dx]

Firstly, the data set was statistically interpreted. Figure 1 shows the main descriptive statistics of the data set contents.

	lesion_id	image_id	dx	dx_type	sex	localization	path	cell_type
count	10015	10015	10015	10015	10015	10015	10015	10015
unique	7470	10015	7	4	3	15	10015	7
top	HAM_0000835	ISIC_0033874	nv	histo	male	back	C:\Users\Abacus_171\Documents\GitHub\CSE498R_S...	Melanocytic nevi
freq	6	1	6705	5340	5406	2192	1	6705

	age	cell_type_idx
count	9958.000000	10015.000000
mean	51.863828	3.623964
std	16.968614	1.208859
min	0.000000	0.000000
25%	40.000000	4.000000
50%	50.000000	4.000000
75%	65.000000	4.000000
max	85.000000	6.000000

FIGURE 1. Main descriptive statistics of the data set contents

We can see how there was a unique image id, but not a unique lesion id for each entry. This indicated that there were duplicate images for the same lesion id but at a different distortion, such as an angle, shear, or zoom distortion. In addition, class nv dominated the skin lesion categories, by having a frequency of 6,705 out of the 10,015 images we got, directly signaling a class imbalance issue in the data

set. Before analyzing the skin lesion categories further, let us analyze the rest of the attributes first. Figure 2 shows the distribution of the patients' age. It can be seen that the majority of patients resided between the ages of 35 and 70, with the highest number of patients at the age of around 45.

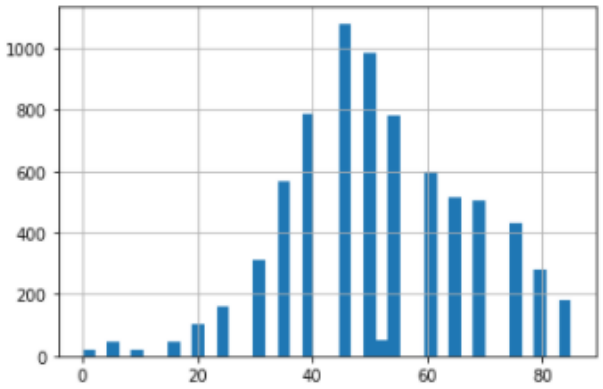

FIGURE 2. Patient's Age Distribution

Figure 3 shows the distribution of the patients' sex. It can be seen that there was almost an equal amount of male to female patients.

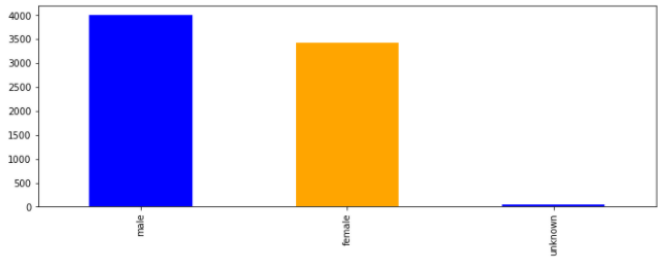

FIGURE 3. Patient's Sex Distribution

The technical validation field category represented the ground truth of the data set and indicated how the skin lesion diagnosis was made. The publishers defined four different types of ground truths which were:

- Histopathology: Diagnoses of excised lesions have been performed by specialized dermatopathologists.
- Confocal: Diagnoses of excised lesions have been based
- Follow-up: This type is limited to nevi class only, where digital dermatoscopy did not show any changes during 3 follow-up visits or 1.5 years.
- Consensus: Based on experts consensus. This type is defined for typical benign cases without histopathology or follow-up, and where two experts have provided same unequivocal benign diagnosis.

Figure 4 shows the distribution of the technical validations. As shown in the figure, more than 50% of the skin lesion diagnosis were based on histopathology, and the highest was at follow-up.

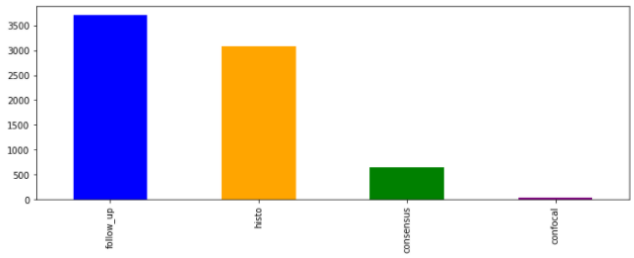

FIGURE 4. Technical Validation Distribution

Figure 5 shows the localization distribution of the data set. It can be seen that the back, lower extremity, and trunk have the highest skin cancer exposure rates.

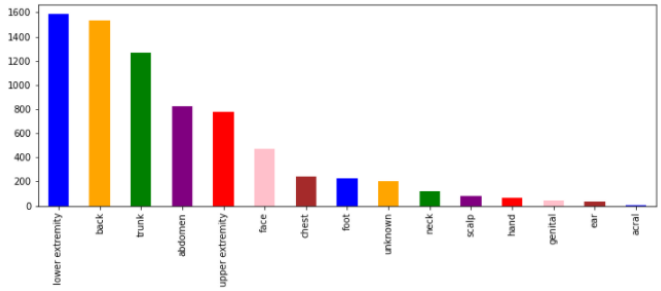

FIGURE 5. Localization Distribution

There were seven different classes of diagnostic skin lesion categories were present in the data set. The seven categories were:

- Melanocytic Nevi [nv]: Benign neoplasms of melanocytes and appear in a myriad of variants. The variants may differ significantly from a dermatoscopic point of view. [6705 images]
- Melanoma [mel]: Malignant neoplasm derived from melanocytes that may appear in different variants. If excised in an early stage, it can be cured by simple surgical excision. [1113 images]
- Benign Keratosis-like Lesions [bkl]: Can be regarded as a flat variant of seborrheic keratosis and lichen-planus like keratoses (LPLK), which corresponds to a seborrheic keratosis or a solar lentigo with inflammation and regression. [1099 images]
- Basal Cell Carcinoma [bcc]: A common variant of epithelial skin cancer that rarely metastasizes but grows destructively if untreated. [514 images]
- Actinic Keratoses [akiec]: Common non-invasive, variants of squamous cell carcinoma that can be treated locally without surgery. [327 images]
- Vascular Lesions [vasc]: These range from cherry angiomas to angiokeratomas and pyogenic granulomas, meaning they could be benign or malignant. [142 images]
- Dermatofibroma [df]: Benign skin lesions regarded as either a benign proliferation or an inflammatory reaction to minimal trauma. [115 images]

Figures 6 and 7 shows sample images from the data set for each class. Figure 6 shows the sample images for the seven skin lesion categories for Simple Sequential model. Whereas Figure 7 shows Sample images for the seven skin lesion categories for the models Inception V3 and ResNet152V2.

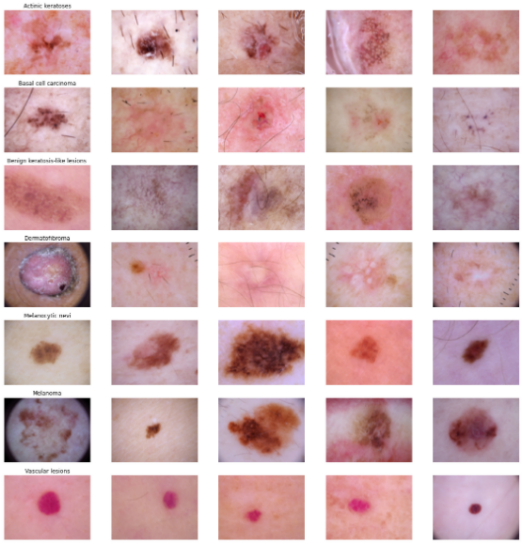

FIGURE 6. Sample images for the seven skin lesion categories for Simple Sequential model

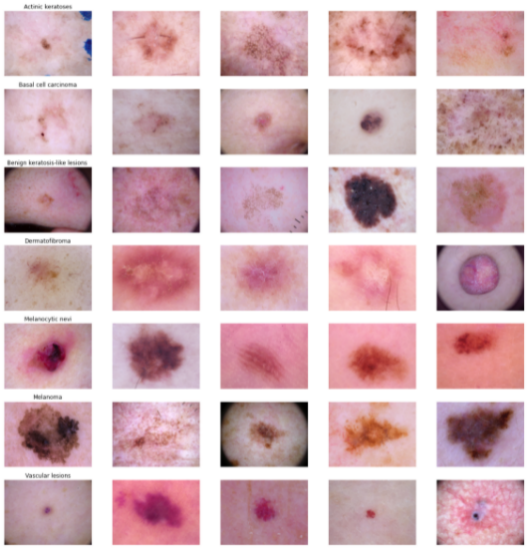

FIGURE 7. Sample images for the seven skin lesion categories for Inception V3 and ResNet152V2

Figure 8 illustrates the distribution of the skin lesion categories. As already mentioned, clearly melanocytic nevi [nv] represented the majority of the data set entries, and dermatofibroma had the fewest representations. (This demonstrated a key class imbalance challenge which was addressed in the next section in this paper.)

V. DATA PRE-PROCESSING

A. DATA EDITING AND CLEANSING

1) Created two dictionaries for the images and their labels. The first dictionary included image names that were retrieved from the various image folders of the downloaded data set. Afterwards, a second dictionary was created to map the diagnostic skin lesion categories code to the category full name.

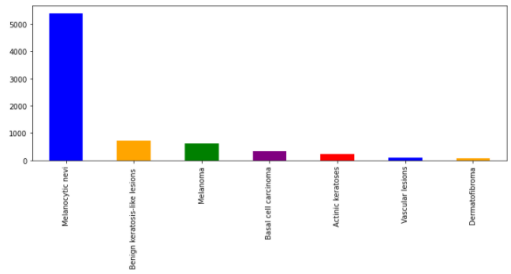

FIGURE 8. Skin Lesion Categories Distribution

2) Data cleansing. This involved dropping lesion id's with duplicate images, keeping only one image for each lesion id. These images are of the same exact lesion but at a different angle, zoom etc. In addition, there were fifty-two NA values in the age entries. These were replaced by the age mean of the data set. That ensured reasonable estimation of what the missing values could be rather than ignoring the entire patient's record.
3) Unique numeric codes were created for each skin lesion category to assist with the predictions to be made later on, as integers were easier to handle than strings.
4) Images were resized and processed. Width and height sizes would have be an issue when it came to training our CNN models, due to the huge amount of images on hand. To speed up the process and ensure our CNN models worked smoothly, images were resized by a factor of 0.25. The new images were 150 by 112.5 pixels. Afterwards, the image was flattened and stored as a numeric image list.
5) Applied one-hot encoding to the skin lesion categories unique numeric codes that were created. This is where the integer encoded variable was removed and a new binary variable was added for each unique integer value, producing the (1,7) row vector in our case as we had seven unique label integer values. This was needed as there was no ordinal relationship between the label's integer values. It eliminated any natural integer ordering assumption the CNN algorithm may conclude.

B. DATA SPLITTING

1) Feature and target split. The feature used in this project was the flattened numeric image lists created in the previous subsection. The target was the one-hot encoding created for the skin lesion categories.
2) Training 70 / Validation 10 / Test 20 split. The data was split 70:10:20 respectively across each class indi-

vidually to ensure there was enough samples from each class in each split for accurate modeling. This was necessary due to the huge class imbalance demonstrated by the data set. A 70:10:20 is a split commonly used when the data is less than 100,000 entries.

C. FEATURE NORMALIZATION

The image feature was then normalized. Normalization is a scaling technique in which values are shifted and re-scaled so that they end up ranging between 0 and 1. The normalization of each image was done by subtracting it's values from the training's mean value and then dividing by the training's standard deviation. The normalization formula is equation 1.

$$x_{(new)} = \frac{x - \mu}{\sigma} \tag{1}$$

D. DATA AUGMENTATION

All the original images were transformed and augmented every epoch and then used for training to avoid overfitting. This allowed the model to be more robust and accurate, as it was trained on different variations of the same image. The number of images in each epoch was equal to the number of original images. The images were:

- Randomly rotated by 20%
- Randomly shifted horizontally by 20%
- Randomly shifted vertically by 20%
- Randomly sheared by 10%
- Randomly zoomed by 10%
- Randomly channel shifted by 10%

This method was preferred over random oversampling, which was one way of dealing with class imbalances. Random oversampling consisted of re-sampling less frequent samples to adjust their amount in comparison with predominant samples. However, the distribution of classes would change significantly, where the smaller classes would be much less variable, and the larger classes would have richer variations.

E. FURTHER DATA EDITING

Flattened images were reshaped back to [width x height x depth] to be fed into our models.

VI. MODELS AND IMPLEMENTATIONS

The models were built and implemented in Python 3.7.4, using the Tensorflow and Keras libraries.

A. SIMPLE SEQUENTIAL MODEL, INCEPTIONV3, RESNET152V2

These models were pre-built and loaded via the keras.applications library package. Extra layers were added to each model's base network to align and flatten the model's inputs, the number of parameters, and the output.

B. MODELS HYPERPARAMETERS

All the CNN models employed had so many parameters, meaning there were so many possible changes in the architecture. In addition, training them with the huge data set in hand took quite a long time. Performing hyperparameter optimization in our case would have been a total overkill, specially with the computing resources we had available to us. Therefore, we utilized some common hyperparameter values and looked further into techniques to achieve a better model evaluation. Table 2 highlights the hyperparameter values used across all four CNN models. The following explains the reasoning behind the hyperparameters values selected:

- Optimizer: Adam is the most common optimization algorithm used today for training deep neural networks, as it is straightforward to implement, computationally very efficient, and is very effective in dealing with large data and parameters. Adam can be looked as a combina-

Layer	Layer	Hyperparameters
1	Conv2D	32 Filters, 3x3 Filter Size ReLU Activation, Sam
2	Conv2D	32 Filters, 3x3 Filter Size ReLU Activation, Sam
3	MaxPool2D	2x2 Pool Size
4	Dropout (Core Layer)	0.25 Neurons
5	Conv2D Conv2D	64 Filters, 3x3 Filter Size ReLU Activation, Sam
6	Conv2D	64 Filters, 3x3 Filter Size ReLU Activation, Sam
7	MaxPool2D	2x2 Pool Size
8	Dropout (Core Layer)	0.4 Neurons
9	Flatten (Core Layer)	-
10	Dense	128 Units, ReLU Activation
11	Dropout (Core Layer)	0.5 Neurons
12	Dense	7 Units, Softmax Activation

TABLE 1. Layers and Hyperparameters

Hyperparameter	Value
Optimizer	Adam
Loss Function	Categorical Cross-Entropy
Epochs	50
Batch Size	10
Learning Rate	0.001 - 0.00001 (Reduces on Plateau)

TABLE 2. Models Hyperparameters of ResNet 152V2 and Inception V3

Hyperparameter	Value
Optimizer	Adam
Loss Function	Categorical Cross-Entropy
Epochs	30
Batch Size	16
Learning Rate	0.000001 - 1e-06 (Reduces on Plateau)

TABLE 3. Models Hyperparameters of Simple Sequential Analysis

Skin Lesion Category	Weight
Actinic keratoses [akiec]	1.0
Basal cell carcinoma [bcc]	1.0
Benign keratosis-like lesions [bkl]	1.0
Melanoma [mel]	1.0
Melanocytic nevi [nv]	3.0
Vascular lesions [vasc]	1.0
Dermatofibroma [df]	1.0

TABLE 4. Class Weights

tion of RMSprop and Stochastic Gradient Descent with momentum.

- Loss Function: Categorical cross-entropy is a loss function that is used for single label categorization. This is when only one category is applicable for each data point. This worked perfectly here as one example could only belong to one of the seven skin lesion categories.
- Epochs: Upon multiple initial trials with values of 20, 50, 100, 150 and 200, 50 epochs was sufficient to get to the most optimum results for Inception V3 and ResNet152V2. For Simple Sequential model, 30 epochs gave us the optimum results.
- Batch Size: Upon multiple initial trials with values of 5, 10, 20, and 40, a batch size of 10 produced the most optimum results for Inception V3 and ResNet152V2. For Simple Sequential model, a batch size of 16 produced the best results.
- Learning Rate: A learning rate annealer was utilized here. A decreasing learning rate during training enabled the global minimum of a loss function to be reached efficiently. Learning rate started at 0.001 for Inception V3 and ResNet152V2 and decreased by factor of 0.5 if the validation accuracy was not improved after 3 epochs (patience). However, for Simple Sequential model, the learning rate was 0.000001 - 1e-06.

C. CLASS WEIGHTS

Weights were added to make the models more sensitive to the Melanocytic nevi [nv] skin lesion category due to the classes imbalance, as it represented around two-thirds of the data. The weights were the factors by which the loss value was multiplied internally for use in the backpropagation algorithm. Equation 2 demonstrates how it worked.

$$weighedlossclass[i] = loss[i] \times classweights[i] \quad (2)$$

Table 4 illustrates the weights associated with each skin lesion category. In reference to table 4 and equation 2, this made the model penalize any Melanocytic nevi [nv] classification mistake by a factor of three.

VII. RESULTS

All three CNN models were run on a machine which comprises of Nvidia GeForce GTX 1660 - 6GB, DDR4-1600 32 GB, and Intel Core i-7700 processor.

In our GitHub there are 2 files for implementation:

1) "skinCancerDetection.ipynb" for implementation of Simple Sequential Analysis model.
2) "SkinCancerDetection v2.ipynb" file for ResNet152V2 and Inception V3.

An accuracy of 75.6% and the training accuracy of 76.9% have been achieved after applying the publicly available data set. Simple Sequential analysis provided the highest accuracy among all other models. We applied 3 models Simple Sequential Analysis, Inception V3 and ResNet152V2. We planned on using VGG16 as a model as well but the accuracy was way too low and also our device have become dead even when we tried to imply ResNet152V2 after Inception V3. So we have decided to not implement any new models further.

A. MODELS ACCURACY, LOSS, ERRORS, AND TIME TAKEN FOR THE TRAINING AND VALIDATION SETS

Table 5 shows the weighted training and validation sets accuracy, loss, errors and the total training time associated with each model. All models had a low bias, as the training set errors are all below 1% and close to the optimal Bayes error of 0%. In addition, all models had a low variance, as training and validation accuracies were within 1% of each other. This meant all models were neither underfitting or overfitting. Furthermore, Simple Sequential Analysis took the shortest to train in around 1,860 seconds, and ResNet152V2 took the longest to train in around 34,695 seconds. That was expected due to the huge number of layers difference between the models.

B. MODELS HISTORY

Figure 9 shows the models' training history against the number of epochs. That was an important visualization to ensure the model kept learning with every epoch, improving its accuracy and reducing its losses and errors, while trying to optimize the objective function. The other two models' improvements with every epoch were verified as well through this visualization. The figures can be viewed in the code.

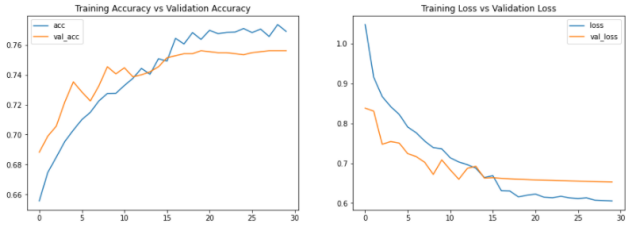

FIGURE 9. Simple Sequential analysis training history against epochs

		Simple Sequential Analysis	InceptionV3	ResNet152V2
Accuracy (%)	Train	76.90	72.32	72.32
	Val	75.60	72.32	72.32
Loss	Train	0.605	1.091	0.936
	Val	0.653	0.893	0.836
RMSE	Train	-	0.156	0.139
	Val	-	0.156	0.141
MSE	Train	-	0.0610	0.0525
	Val	-	0.0599	0.0546
Total Train Time (s)		1860	4321	34695

FIGURE 10. Table 5: Models Training and Validation

Metrics	Model		
	Simple Sequential Analysis	InceptionV3	ResNet152V2
Accuracy (%)	76.90	72.32	72.32
Precision (%)	-	57.00	52.00
Recall (%)	-	30.00	72.00
F1-score (%)	-	37.00	61.00

FIGURE 11. Table 6: Models Performance on the Testing Set

C. MODELS PERFORMANCE COMPARISONS FOR THE TESTING SET

The metrics discussed in section II were utilized. Table 6 shows the weighted testing sets accuracy, recall, precision, and f1-score associated with each model. Firstly, all of the testing accuracies were comparable to the validation accuracies, further solidating the fact that all models were neither underfitting or overfitting. Simple Sequential Analysis topped each metric score, scoring 76.90% accuracy. Simple Sequential Analysis performance scores were followed by ResNet152V2 and InceptionV3. Where ResNet152V2 scoring 72.32% accuracy, 52% precision, 72% recall, and 61% f1-score. And Inception V3 scoring 72.32% accuracy, 57% precision, 30% recall, and 37% f1-score. Taking total training time into account however, Simple Sequential model's performance was a contrast to ResNet152V2 and Inception V3's performance. Simple Sequential model trained in about 31 minutes, whereas ResNet152V2 took over 9 hours and 40 minutes, and Inception V3 over 1 hour. Simple Sequential model is more time efficient than ResNet152V2 and Inception V3. In addition, Simple Sequential model achieved over 4% higher accuracy than the other models.

VIII. CONCLUSION

CNNs display a high performance as state-of-the-art skin lesion classifiers. Unfortunately, it is difficult to compare different classification methods because some approaches use nonpublic datasets for training and/or testing, thereby making reproducibility difficult. Future publications should use publicly available benchmarks and fully disclose methods used for training to allow comparability.

This work investigated the use of different CNN architectures to predict skin lesion categories based on skin lesion images. The data set was first pre-processed via data editing and cleaning, then split into the feature and target values, before applying feature normalization and data augmentation. We used Simple Sequential analysis alongside InceptionV3 and ResNet152V2 which were pre-trained CNN models on the ImageNet data set. Hyperparameters were preselected and kept constant across all models. Weights were added to address the major class imbalance in the data set, by making the models more sensitive to Melanocytic nevi [nv] as it represented around two-thirds of the data. As for the results, Simple Sequential Analysis had the highest accuracy of 76.9%, while ResNet152V2 topped the metric scores overall with ResNet152V2 scoring 72.32% accuracy, 52% precision, 72% recall, and 61% f1-score. However, it is clear that the more training data available for the skin lesion category, the better the models were at predicting them during test time. More malignant skin lesion images had to be provided for more accurate results and better malignancy classification, as the main aim in the medical field is to reduce the number of patients dying. Overall, this project showed the highest accuracy for simple sequential model.

IX. FUTURE WORK

To improve upon these results, a couple of next steps are required:

- Improve model accuracy
- Utilize all the data in this data set, such as age and sex, via other machine learning techniques to extract useful information that can be combined with the work demonstrated in this paper.
- Increase dataset and explore other skin lesion data sets to obtain more training data to feed into our models, specially for the underrepresented and malignant skin lesion classes.
- Explore other pre-trained CNN models to look for higher accuracy.
- Implement different models.

APPENDIX A

GROUP MEMBERS CONTRIBUTION

Each member contributed equally to all parts of the project. This includes the literature research and survey, data set exploration, data pre-processing, models building and implementations, and result comparisons. Meetings were conducted every week via zoom to discuss our project progress.

ACKNOWLEDGEMENT

The authors would like to thank Dr. Sifat Momen for providing us the platform to conduct a project of such magnitude and for all their guidance and support.

REFERENCES

[1] @articlesaba2019region, title=Region extraction and classification of skin cancer: A heterogeneous framework of deep CNN features fusion and reduction, author=Saba, Tanzila and Khan, Muhammad Attique and Rehman, Amjad and Marie-Sainte, Souad Larabi, journal=Journal of medical systems, volume=43, number=9, pages=1–19, year=2019, publisher=Springer

[2] @articlebrinker2018skin, title=Skin cancer classification using convolutional neural networks: systematic review, author=Brinker, Titus Josef and Hekler, Achim and Utikal, Jochen Sven and Grabe, Niels and Schadendorf, Dirk and Klode, Joachim and Berking, Carola and Steeb, Theresa and Enk, Alexander H and Von Kalle, Christof, journal=Journal of medical Internet research, volume=20, number=10, pages=e11936, year=2018, publisher=JMIR Publications Inc., Toronto, Canada

[3] @articleesteva2017dermatologist, title=Dermatologist-level classification of skin cancer with deep neural networks, author=Esteva, Andre and Kuprel, Brett and Novoa, Roberto A and Ko, Justin and Swetter, Susan

M and Blau, Helen M and Thrun, Sebastian, journal=nature, volume=542, number=7639, pages=115–118, year=2017, publisher=Nature Publishing Group

[4] @incollectionnahata2020deep, title=Deep learning solutions for skin cancer detection and diagnosis, author=Nahata, Hardik and Singh, Satya P, booktitle=Machine Learning with Health Care Perspective, pages=159–182, year=2020, publisher=Springer

[5] @articlenunnari2019cnn, title=A CNN toolbox for skin cancer classification, author=Nunnari, Fabrizio and Sonntag, Daniel, journal=arXiv preprint arXiv:1908.08187, year=2019

[6] @inproceedingshartanto2020development, title=Development of Mobile Skin Cancer Detection using Faster R-CNN and MobileNet V2 Model, author=Hartanto, Cahyo Adhi and Wibowo, Adi, booktitle=2020 7th International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE), pages=58–63, year=2020, organization=IEEE

[7] @articlehekler2019superior, title=Superior skin cancer classification by the combination of human and artificial intelligence, author=Hekler, Achim and Utikal, Jochen S and Enk, Alexander H and Hauschild, Axel and Weichenthal, Michael and Maron, Roman C and Berking, Carola and Haferkamp, Sebastian and Klode, Joachim and Schadendorf, Dirk and others, journal=European Journal of Cancer, volume=120, pages=114–121, year=2019, publisher=Elsevier

[8] @inproceedingsaima2019predictive, title=Predictive Approach for Melanoma Skin Cancer Detection using CNN, author=Aima, Arneesh and Sharma, Akhilesh Kumar, booktitle=Proceedings of International Conference on Sustainable Computing in Science, Technology and Management (SUSCOM), Amity University Rajasthan, Jaipur-India, year=2019

[9] @inproceedingspham2018deep, title=Deep CNN and data augmentation for skin lesion classification, author=Pham, Tri-Cong and Luong, Chi-Mai and Visani, Muriel and Hoang, Van-Dung, booktitle=Asian Conference on Intelligent Information and Database Systems, pages=573–582, year=2018, organization=Springer

[10] @articlezhang2019optimization, title=Optimization of the convolutional neural networks for automatic detection of skin cancer, author=Zhang, Long and Gao, Hong Jie and Zhang, Jianhua and Badami, Benjamin, journal=Open Medicine, volume=15, number=1, pages=27–37, year=2019, publisher=De Gruyter

[11] https://github.com/TareqTayeh/Skin-Lesion-Classification-using-CNN

[12] @articlegloster1996epidemiology, title=The epidemiology of skin cancer, author=Gloster Jr, Hugh M and Brodland, David G, journal=Dermatologic Surgery, volume=22, number=3, pages=217–226, year=1996, publisher=Wiley Online Library

[13] @articlenarayanan2010ultraviolet, title=Ultraviolet radiation and skin cancer, author=Narayanan, Deevya L and Saladi, Rao N and Fox, Joshua L, journal=International journal of dermatology, volume=49, number=9, pages=978–986, year=2010, publisher=Wiley Online Library

[14] @articletschandl2018ham10000, title=The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions, author=Tschandl, Philipp and Rosendahl, Cliff and Kittler, Harald, journal=Scientific data, volume=5, number=1, pages=1–9, year=2018, publisher=Nature Publishing Group

[15] @articleshin2016deep, title=Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning, author=Shin, Hoo-Chang and Roth, Holger R and Gao, Mingchen and Lu, Le and Xu, Ziyue and Nogues, Isabella and Yao, Jianhua and Mollura, Daniel and Summers, Ronald M, journal=IEEE transactions on medical imaging, volume=35, number=5, pages=1285–1298, year=2016, publisher=IEEE

[16] @articlekadampur2020skin, title=Skin cancer detection: Applying a deep learning based model driven architecture in the cloud for classifying dermal cell images, author=Kadampur, Mohammad Ali and Al Riyaee, Sulaiman, journal=Informatics in Medicine Unlocked, volume=18, pages=100282, year=2020, publisher=Elsevier

[17] @articleyu2018acral, title=Acral melanoma detection using a convolutional neural network for dermoscopy images, author=Yu, Chanki and Yang, Sejung and Kim, Wonoh and Jung, Jinwoong and Chung, Kee-Yang and Lee, Sang Wook and Oh, Byungho, journal=PloS one, volume=13, number=3, pages=e0193321, year=2018, publisher=Public Library of Science San Francisco, CA USA

[18] @articlemunir2019cancer, title=Cancer diagnosis using deep learning: a bibliographic review, author=Munir, Khushboo and Elahi, Hassan and Ayub, Afsheen and Frezza, Fabrizio and Rizzi, Antonello, journal=Cancers, volume=11, number=9, pages=1235, year=2019, publisher=Multidisciplinary Digital Publishing Institute

[19] @miscboman2018evaluating, title=Evaluating a deep convolutional neural network for classification of skin cancer, author=Boman, Joakim and Volminger, Alexander, year=2018

[20] @articledildar2021skin, title=Skin Cancer Detection: A Review Using Deep Learning Techniques, author=Dildar, Mehwish and Akram, Shumaila and Irfan, Muhammad and Khan, Hikmat Ullah and Ramzan, Muhammad and Mahmood, Abdur Rehman and Alsaiari, Soliman Ayed and Saeed, Abdul Hakeem M and Alraddadi, Mohammed Olaythah and Mahnashi, Mater Hussen, journal=International journal of environmental research and public health, volume=18, number=10, pages=5479, year=2021, publisher=Multidisciplinary Digital Publishing Institute

[21] @inproceedingscodella2018skin, title=Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (isbi), hosted by the international skin imaging collaboration (isic), author=Codella, Noel CF and Gutman, David and Celebi, M Emre and Helba, Brian and Marchetti, Michael A and Dusza, Stephen W and Kalloo, Aadi and Liopyris, Konstantinos and Mishra, Nabin and Kittler, Harald and others, booktitle=2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018), pages=168–172, year=2018, organization=IEEE

[22] @articleli2018skin, title=Skin lesion analysis towards melanoma detection via end-to-end deep learning of convolutional neural networks, author=Li, Katherine M and Li, Evelyn C, journal=arXiv preprint arXiv:1807.08332, year=2018

[23] @inproceedingshasan2019skin, title=Skin cancer detection using convolutional neural network, author=Hasan, Mahamudul and Barman, Surajit Das and Islam, Samia and Reza, Ahmed Wasif, booktitle=Proceedings of the 2019 5th International Conference on Computing and Artificial Intelligence, pages=254–258, year=2019

[24] @incollectiongarg2021decision, title=Decision support system for detection and classification of skin cancer using CNN, author=Garg, Rishu and Maheshwari, Saumil and Shukla, Anupam, booktitle=Innovations in Computational Intelligence and Computer Vision, pages=578–586, year=2021, publisher=Springer

[25] N. Codella et al., "Deep learning ensembles for melanoma recognition in dermoscopy images", IBM Journal of Research and Development, vol.61, no. 45, pp. 5:1-5:15, 2017. Available: 10.1147/jrd.2017.2708299.

[26] https://www.saudis.cc/jug/using-inception-v3-for-image-classification