**CNN-BASED APPROACH TO CLASSIFY**

**MELANOMA AND KERATOSIS IMAGES**

***Submitted by***

**Dharshini Thangapushparaja,** 67617845

**Pooja Gadigeppagouda Hiregoudar,** 35880566

**Sydney Agbonkhese,** 63508188

**Vishal Balasubramanian,** 68971506

***for the course***

**ENGR 501: DEEP AND REINFORCEMENT LEARNING FOR ENGINEERS**

**INSTRUCTOR: DR. CHEN FENG**



**School of Engineering**

**Faculty of Applied Science**

**University of British Columbia, Okanagan**

**APRIL 2022**

**ABSTRACT**

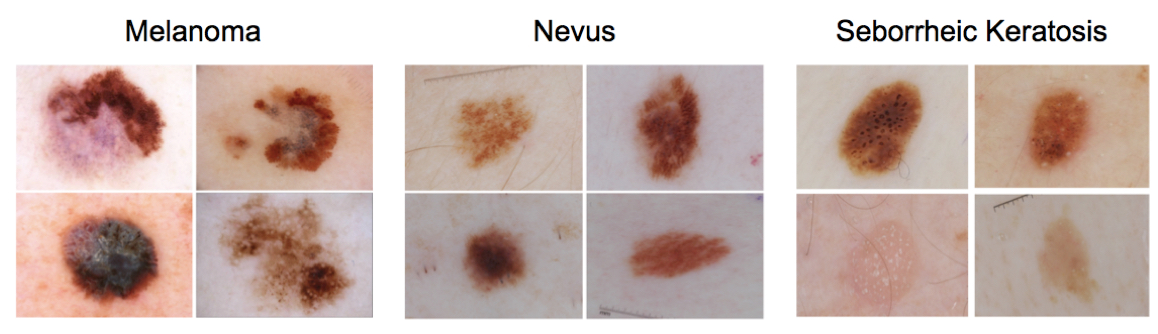
Keratosis is a non-threatening skin condition that looks very similar to melanoma (skin cancer which can be a life-threat). Doctors often get confused between the two owing to the similarity in their appearance. The difference can be found out only after biopsy. This is particularly an issue in third world countries and remote regions where access to healthcare is challenging. To solve this issue, we propose a Smart Convolutional Neural Network (CNN) based image classification algorithm which leverages Keras for efficient classification between melanoma and keratosis (non-melanoma). A CNN image classification takes an input image, processes it and classifies it under certain categories. Deep learning CNN models train and test, each input image will pass it through a series of convolution layers with filters. A Deep learning model is developed to identify the category of the skin disease. The application is the classification of melanoma and keratosis (non-melanoma) since it is the most mistaken and wrongly identified skin disorder.

***Keywords -*** CNN, Melanoma, Keratosis, Skin Cancer Classification

**1. INTRODUCTION**

Skin Cancer is one of the most lethal types of cancer which looks like an unorganized callous of cells. The main causes of skin cancer are Ultraviolet (UV) radiation, genetic disorders, and tumor cell blood transfusion. Early detection of skin cancer can help in efficient treatment with greater chances of complete cure. Skin Cancer is classified into three major types - Basal Cell Carcinoma, Squamous Cell Carcinoma and Melanoma. Among these three types, melanoma is the most indistinguishable since it can develop in any body part and resembles a mole but is cancerous. The easiest way to identify melanoma is just by looking at it. They are darker when compared to the average skin tone and can sometimes be itchy [1].

Melanoma is particularly notorious since it looks extremely like skin conditions such as keratosis as seen in Figure 1. Keratosis is a skin disease that is characterized by scaly and rough patches of skin contrasting the regular skin tone. Like melanoma, keratosis can develop in any body part. Unlike melanoma, keratosis is non-lethal and can be cured easily. People in third world countries and remote regions adopt homemade recipes to cure keratosis [2]. This is mainly due to the lack of professional medical practitioners and dermatologists and high cost of diagnosis. Identifying melanoma as keratosis and taking treatment for keratosis can do more damage, leading to faster spread of the tumor [3]. Hence, classification of keratosis (non-lethal) and melanoma (lethal) is essential. The standard procedure to differentiate between the two involved taking a biopsy sample and then processing it [4]. This not only takes time but also is an issue for people living in third world countries and remote regions, who do not have proper access to medical facilities.



**Figure 1.** Images of Melanoma and Keratosis (Nevus and Seborrheic)

Several image classification approaches have been tried and tested in the past [5-10]. The major distinguishing factor in these image processing approaches is that melanoma is asymmetric and has fuzzy borders, while keratosis is symmetric and has clear borders. Approaches that include the size and color of the condition were also considered [11-15]. However, these approaches have low accuracy owing to the non-homogeneity in the images. Hence Machine Learning (ML) and Deep Learning (DL) approaches were introduced and tested to enhance the classification accuracy [16-19]. The issue with most of these ML/DL based approaches is that the dataset used is very less for training. The lack of a sizable amount of dataset and at the same time, data that is balanced while training is a challenge in developing a classification model. Data must be purified and augmented to make the model more efficient. This step of data augmentation consists of two steps - Data Cleansing and Data Duplication. In this paper, we propose a CNN based approach to address the accuracy issue in classification of keratosis and melanoma images.

**2. EXISTING METHODS AND PROPOSED WORK**

Detection Of Skin Cancer Using Deep Neural Networks by Rahi, M.M.I., et al. In this research they used ImageNet's pre-trained data, including DENSENET121, RESNET50, and VGG11, were used in a transfer learning model and they used HAM10000 dataset. They used 4 models for training and chose the model which produced high accuracy. The models and the accuracy are listed below,

* CNN model – 79%
* ResNet50 – 90%
* DenseNet121 – 90%
* VGG11 – 85%

By comparing all the 4 models, it is given that ResNet50 showed consistent accurate results, so they implemented this model for evaluation. Skin Lesion Classification using Deep Learning Architectures by Salian, A.C., et al. They are classifying skin lesions involving augmenting labelled images, extracting features, and predicting skin lesions. In this research, PH2 and HAM10000 datasets were employed in pre-trained innovative models i.e., VGG16 and MobileNet, subjected to two conditions that are without augmentation and with augmentation. A tailor-made deep learning architecture was then constructed and evaluated against the two innovative models' performance. A conjecture that a thoughtful model design from the scratch would perform just as fine. From the results, MobileNet and the tailor-made model fared well inaccuracy rate performance. Despite this, the result indicated an insignificant effect of data augmentation in comparison to non-augmented data in classification.

Predominantly, past classification solutions inclined towards employing complex and sophisticated models to improve accuracy rate in detection. Research evidence on intraclass dissimilarity and inter-class similarity in lesion features is rather sparse in the literature. Employing a sophisticated model with a large computation overhead might render challenges to applicability in the real world. So, we are going with creating the model from scratch. The approach is to build the project Image Recognition model for Skin Cancer Identification using CNN which are discussed in the following steps:

* Explore the HAM10K dataset
* Build a Baseline VGG model
* Apply Data augmentation and Regularization techniques
* Train and validate the model accuracy

The method followed for building this model involves- Importing and training the data set, creating a test harness to conduct a thorough evaluation of a model and define a performance baseline for a classification task, boosting learning and model capacity by extending a baseline model, creating a finished model, and testing its performance. The chosen metric for evaluating the model performance through simulation is the accuracy, which is appropriate for the Melanoma identification from the HAM10K dataset. A Stochastic Gradient Descent (SDG) optimizer was used initially to compile the accuracy metric of the model.

The model accuracy has increased or improvised by the following the below ideas:

* Training more Epochs
* Using Data Augmentation
* Designing a larger network topology
* Tuning the learning rate

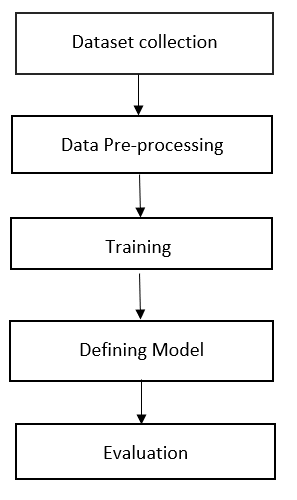
Several parameters are used to be tuned up in our proposed CNN model. However, the principal once that we have used here is given below.

* Batch size (32): the number of processed images in every iteration
* Learning rate (0.001): It is the amount that the weights are updated during the training process.
* Optimizer: Adam is the optimizer that we utilized. It is a stochastic gradient descent substitution optimization approach for minimizing the loss function for training DL models. It is computationally efficient, uses little memory, and is well suited to issues with a lot of data, parameters.
* Loss Function: ReLU is used as a loss function. It measures the performance of a classification model whose output is a probability value between 0 and 1.
* The number of epochs (50): The number of times the using dataset is passed to the created model.

During training, Dropout drops out neurons at random with a certain probability. Every iteration of the training algorithm employs a random subset of the network. This method enables neurons to learn useful traits without the assistance of other neurons. The complete network is used for inference once it has been trained. To avoid overfitting, we can either limit the network's effective capacity by employing methods like dropout, or we can increase the amount of training data by utilizing data augmentation to generate changed versions of the samples we currently have.

**3. METHODOLOGY**

Our proposed model consists of various processes as seen in Figure 2.



**Figure 2.** Design process of proposed CNN model

**3.1 DATASET COLLECTION**

The dataset for this CNN came from Kaggle titled Melanoma. They used the original data from the HAM10k pictures dataset, Human Against Machine with 10,000 training images. The dermoscopic images in the HAM10k dataset were filtered and normalized in terms of luminance, colors, and resolution. Histopathology (also known as a source of truth) supported the real diagnosis in more than 50 percent of cases, which is twice as many as previously available skin lesion datasets. The remainder of the lesions were diagnosed based on dermatologists' consensus. Rather than attempting to categorize seven skin lesions using a very unbalanced dataset, they focused on the diagnosis of melanoma versus non-melanoma. Even after dividing the above categories into two groups, the resulting dataset contains 1,113 images of melanoma and 8,902 images of non-melanoma. Thus, in order to balance the dataset, we used data augmentation on the melanoma group in order to equalize the number of images with the non-melanoma group [20].

After data augmentation, the previously imbalanced 10,000 images became a 17,800 balanced dataset of the two (2) groups (melanoma and non-melanoma). These 17,800 images were organized into three directories: training set, testing set, and validation set. The training directory contains 5,431 images for the melanoma class and 5,431 images for the non-melanoma class, the testing directory comprises 1,781 images for the melanoma class and 1,780 images for the non-melanoma class, while the validation directory contains 1,781 images classified as melanoma and 1,781 images classified as non-melanoma [20].

**3.2 DATA PREPROCESSING**

Data preprocessing is the first and crucial step in preparing the raw data to make it compatible with the CNN model. The main aim of applying the preprocessing technique in our dataset is to remove the noise present in the images and to get high accuracy during classification. The preprocessing techniques that we used are normalization, resizing all the images to common size and converting the string data of label into numerical data. Normalization is used to convert the pixel value of an image from 0-255 to the range of 0-1. By this way, the computation becomes faster and easier.

Most computer vision tasks may benefit from additional data. Unlike some other domains (machine learning applications, for example), computer vision typically lacks sufficient datasets to work with. As a result, data augmentation is a frequently utilized approach for enhancing the performance of computer vision systems. And this is true regardless of whether you're utilizing transfer learning by starting with someone else's pre-trained parameters or attempting to train from scratch [21]. Data augmentation is a technique for artificially boosting the amount of data by adding slightly altered copies of current training data without actually gathering new data [22].

The simplest data augmentation technique is to mirror the vertical and horizontal axes. For instance, reflecting an image of an automobile retains the appearance of the automobile. Another often-used technique for data augmentation is random cropping. When given a dataset or image, random cropping selects a few random crops from the dataset or image to feed into the training sample. However, you may wind up selecting a crop of the image or dataset that may not keep the original image or dataset's properties. Thus, while random cropping is not an ideal strategy for data augmentation, it is a useful technique when the crops represent sufficiently large subsets of the original image [21].

Rotation, local wrapping, shearing, and color-shifting are further data augmentation techniques. Color-shifting is the process of distorting a dataset or image's RGB channels. For instance, by adding distortions to the red and blue channels and subtracting from the green channel, an image becomes more purply, which results in the creation of a distorted image for the training set [21]. The Principal Component Analysis (PCA) algorithm, referred to in the AlexNet publication as PCA Color Augmentation, is one method for implementing color distortion. The rough idea is that, if an image is predominantly purple, that is with significant red and blue tints and very little green, PCA Color Augmentation will add and subtract significantly to the red and blue channels but relatively little to the green channel, maintaining the overall color of the tint [21, 23].

Furthermore, like other aspects of deep neural network training, the data augmentation process contains a few hyperparameters, such as the amount of color shifting implemented, and the exact parameters used for random cropping [21]. The following parameters were used for data augmentation for the convolutional neural network for skin cancer classification:

* Rotation range (15): Generating input data with rotation from -15 to 15.
* Horizontal flip: Flipping the image randomly to the horizontal axis and
* Vertical flip: Flipping the image randomly to the vertical axis.

The project initially involves processing the dataset image by loading the image and seeding it to get a repeatable result every time. The loaded data is normalized from 0-255 to 0, 1 and the categorical data variable is converted to a Machine learning algorithm using one-hot encoding. This improves prediction and classification accuracy. An alternative method that can be used is the Dummy Coding Scheme. Processed data or image was initially tested with the VGG baseline model.

The whole project was worked out with 4 stages:

* Baseline model
* Baseline model + 0.4 Dropout
* Baseline model + 0.4 Dropout + Batch Normalization
* Baseline model + 0.4 Dropout + Batch Normalization + Data augmentation + 50 epoch

The learning rate (lrate) parameter is tunable by step size to move towards the minimum loss function. The ideal value of lrate is between 1 and 10-6. Optimization technique SGD, Stochastic Gradient Descent is replaced by ADAM, which is a combination of Adaptive Moment Optimization and RMS Propagation. The reason for the replacement is that SGD is comparatively more unstable than ADAM and ADAM help in fast convergence. The metrics that are commonly reviewed to learn the performance in the classification are Accuracy, Precision, Recall, and F-1 score.

The data is separated into two parts, validation data and training data, to check for overfitting. As the model is trained, the error is plotted on the training and validation statistics. Overfitting was identified by a substantial discrepancy in the error graph between the training and validation datasets. Reduce model complexity using regularization or enhance dataset using approaches such as data augmentation to reduce overfitting. Adding a component to our cost function that penalizes too complex models is one method to apply regularization.

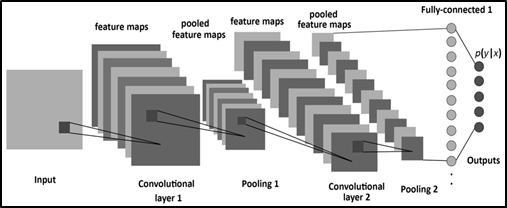
**Table 1.** Classification of loss curves

|  |  |  |
| --- | --- | --- |
| **Training Loss** | **Validation Loss** | **Loss curves** |
| Low | Low | Good fit |
| Low | High | Overfitting |
| High | Low | Unlikely |
| High | High | Underfitting |

The variance of underfitting data is low, but the bias is large, whereas the variance of an overfitted model is low, but the bias is high. Our model had a low training loss but a large validation loss at first, indicating that it was overfitting. Overfitting can be avoided by reducing capacity through regularization and Dropout while increasing data through data augmentation.

**3.3 MODEL CREATION AND TRAINING**





**Figure 3.** Typical Block Diagram of 2-layer CNN

Figure 3 depicts the basic architecture and block diagram of a convolutional neural network. Convolution layers, max-pooling layers, and fully connected dense layers along with batch normalization layers and dropout layers and a flatten layer make up the final network.

1. Convolutional Layers: The input image and a stack of filters/features make up the convolution layer. The image array and the filter are convoluted in 2D, resulting in a stack of filtered image arrays at the output. Numbers reflect how closely the filter fitted that section of the image in the filtered photos. This becomes the location map for the feature. The convolution layer refers to the process of convolving an image with several filters to create a stack of filtered images.
2. Batch Normalization Layers: This layer particularly allows every layer to do learning more independently. It also reduces the internal covariance shift and affects training speed by increasing it.
3. Max Pooling Layers: After every two convolution layers, the network architecture employs max-pooling layers. There are four pooling levels total, each measuring 2x2. By maintaining the maximum value, these pooling layers reduce the image stack size. To produce a lower stack of filtered photos, Max Pooling is performed to the full stack of filtered images.
4. Dropout and Flatten Layers: Dropout is a technique where a selected neuron is ignored during training to prevent the neural network from overfitting. Flatten layer converts the multi-dimensional array to a 1-D array.
5. Dense Layer: The layer collects every single neuron from the preceding layer and creates a fully connected layer in which every output depends on every input. Sigmoid activation function is used at the end of the dense layer function in the model.

The main goal of the project is to create a model and train it to identify and classify the HAM10k dataset images into the 2-class segregated as melanoma(0) and not melanoma(1) . The focus is mainly to increase the accuracy metric as much as possible. This includes high optimization techniques. The project is developed from a basic VGG model baseline to applying Data augmentation and Regularization techniques.

**4. RESULTS**

The baseline model includes 6 convolutional layers, 3 max-pooling layers, a flatten and 3 dense layers and was trained with SGD optimization, categorical cross entropy and ReLU activation. Using a dropout layer of 20% after every block increased the accuracy. Next step of the project is to introduce a regularization method, batch normalization after every convolutional layer of the basic model, which indeed increased the accuracy. Data augmentation is an important part of the project as it trains the model to classify the image into the 10 classes even when the orientation of the image is different. Finally, epochs are increased as the dataset is large, to get more accurate results.

**Table 2.** Accuracy Comparison in percentage

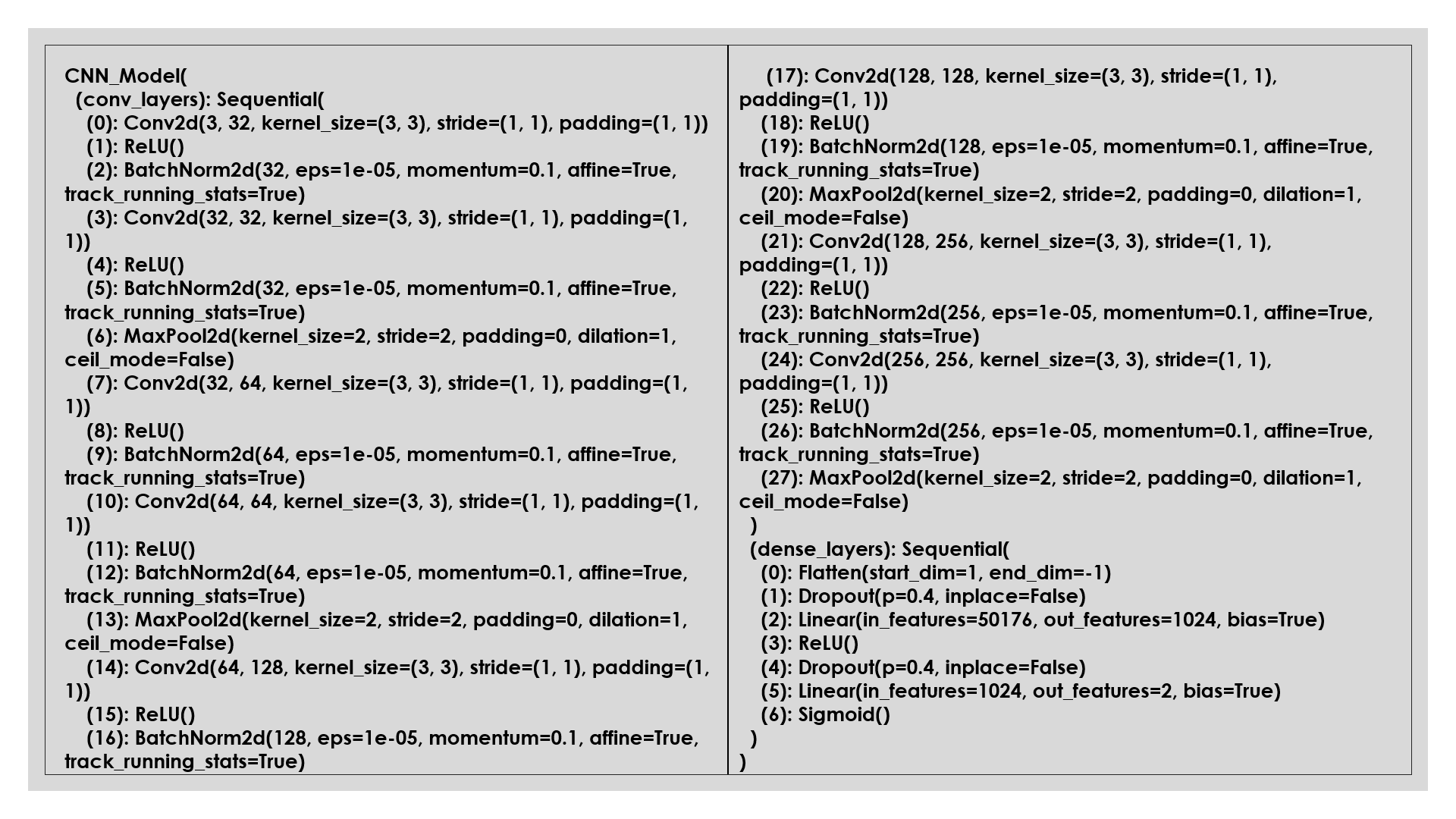
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Method** | **Acc** | **Val\_Acc** | **Loss** | **Val\_Loss** |
| **BL** | 78 | 76 | 43 | 85 |
| **BL+DO** | 88 | 82 | 32 | 53 |
| **BL+DO+BN** | 86 | 85 | 38 | 46 |
| **BL+DO+BN+DA** | 90 | 91 | 25 | 30 |
| **BL+DO+BN+50e+DA** | 93 | 93 | 17 | 16 |

Where BL, DO, BN, DA, 50e, Acc and Val, indicates Baseline, Dropout, Batch Normalization, Data Augmentation, 50 epochs, Accuracy and Validation respectively.

From table 1, our model has developed from overfitting to good fit by making significant changes and from table 2, it is clearly visible that every method used in the model has increased the training and validation accuracy. It has also reduced the losses to a minimum significant level.

The CIFAR-10 data set was used to train the final model. The dataset is split into two parts: training data and testing data. For 50 epochs, the network is trained, 32 batches were created from the training data. As a result, one epoch is completed when the network has been trained on all batches, each batch containing 32 images. The model is updated to a VGG model, and there are two methods for doing so: coding the model or importing a previously trained model via Transfer learning.

**Table 3.** Model Summary



The results of training for 50 epochs are shown below in table 4 about the accuracy metric and classification report. Additionally, the confusion matrix is also used to summarize the results. We were able to identify the Melanoma with 100% accuracy.

**Table 4.** Accuracy

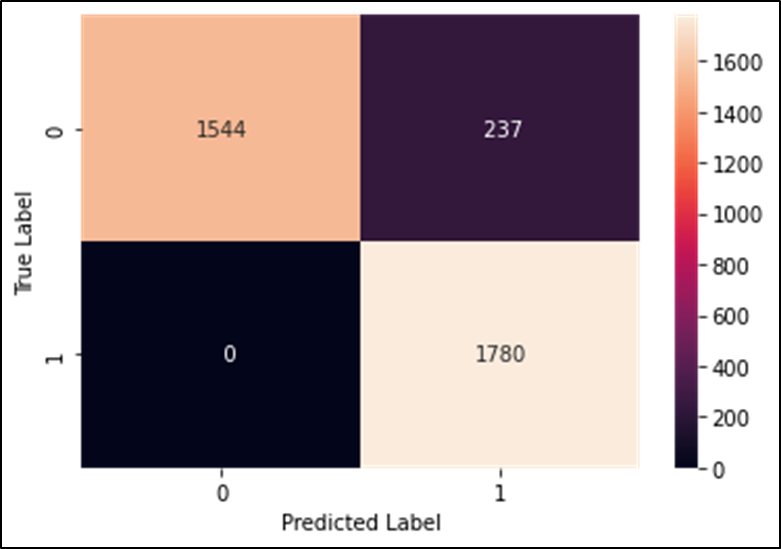
|  |  |
| --- | --- |
| Validation Accuracy | 93.9360 % |
| Train Accuracy | 93.7759% |

A classification report of table 5 is a performance evaluation metric in machine learning. It is used to show the precision, recall, F1 Score, and support of your trained classification model.

**Table 5.** Classification report

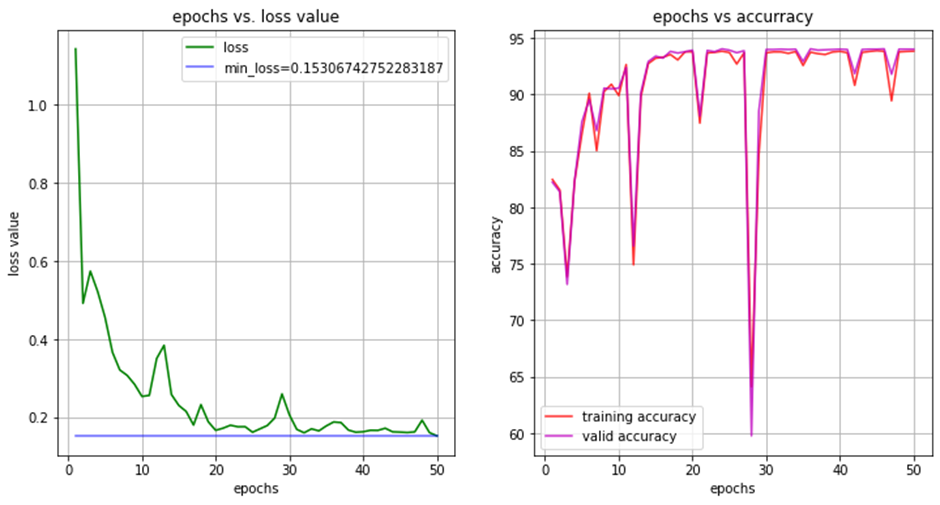
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Precision** | **Recall** | **F1 score** | **Support** |
| 0 | 1.00 | 0.87 | 0.93 | 1781 |
| 1 | 0.88 | 1.00 | 0.94 | 1780 |
| **Accuracy** |  |  | 0.93 | 3561 |
| **Macro avg** | 0.94 | 0.93 | 0.93 | 3561 |
| **Weighted avg** | 0.94 | 0.93 | 0.93 | 3561 |

Figure 4 below shows the confusion matrix for the final model, which indicates the true positives, false negatives, false positives, and true negatives. This is an additional factor to verify the classification report.



**Figure 4.** Confusion Matrix

The figures below show the effect of applying data augmentation to the final network where the overfitting is reduced.



**Figure 5.** Accuracy vs Epochs curve for better fit mode

**5. CONCLUSION AND FUTURE WORK**

A CNN model was developed to distinguish melanoma from keratosis. We achieved an efficiency of over 93%. The developed model can be confidently used for identifying keratosis since we were able to obtain an accuracy of 100% with zero errors while identifying an image as melanoma. Hence this model can be used to easily detect the presence of melanoma. Unlike melanoma, keratosis is a non-threatening skin condition. The developed model can be used in remote locations and third world countries with slower and lesser access to medical facilities. Usage of data augmentation has enhanced the accuracy when compared to other pre-existing methods. Using a dropout layer of 20% after every block increased the accuracy. Next step of the project is to introduce a regularization method, batch normalization after every convolutional layer of the basic model, which indeed increased the accuracy. The reason for the errors can be due to overfitting. Overfitting can be further reduced by introducing an ensemble and deeper neural network. Different methods that can be implemented are - Transfer Learning: On this dataset, we can try employing transfer learning, such as a pre-trained VGG-16 model, including prediction for the model, and Increasing epochs and dataset to improvise the training and learning.

**REFERENCES**

1. https://www.mayoclinic.org/diseases-conditions/skin-cancer/symptoms-causes/syc-20377605#:~:text=There%20are%20three%20major%20types,squamous%20cell%20carcinoma%20and%20melanoma
2. https://www.healthline.com/health/seborrheic-keratosis
3. https://www.healthline.com/health/skin-disorders/seborrheic-keratosis-vs-melanoma#prevention
4. https://www.mayoclinic.org/diseases-conditions/actinic-keratosis/symptoms-causes/syc-20354969#:~:text=Actinic%20keratoses%20are%20scaly%20spots,from%20years%20of%20sun%20exposure
5. Bisla D, Choromanska A, Berman RS, Stein JA, Polsky D. Towards automated melanoma detection with deep learning: Data purification and augmentation. InProceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops 2019
6. Winkler JK, Sies K, Fink C, Toberer F, Enk A, Deinlein T, Hofmann-Wellenhof R, Thomas L, Lallas A, Blum A, Stolz W. Melanoma recognition by a deep learning convolutional neural network—performance in different melanoma subtypes and localisations. European Journal of Cancer. 2020 Mar 1;127:21-9.
7. Naeem A, Farooq MS, Khelifi A, Abid A. Malignant melanoma classification using deep learning: datasets, performance measurements, challenges and opportunities. IEEE Access. 2020 Jun 10;8:110575-97.
8. Sultana NN, Puhan NB. Recent deep learning methods for melanoma detection: a review. InInternational Conference on Mathematics and Computing 2018 Jan 9 (pp. 118-132). Springer, Singapore.
9. Serte S, Serener A, Al‐Turjman F. Deep learning in medical imaging: A brief review. Transactions on Emerging Telecommunications Technologies. 2020 Aug 27:e4080.
10. Serte S, Demirel H. Gabor wavelet-based deep learning for skin lesion classification. Computers in biology and medicine. 2019 Oct 1;113:103423.
11. Goyal M, Oakley A, Bansal P, Dancey D, Yap MH. Skin lesion segmentation in dermoscopic images with ensemble deep learning methods. IEEE Access. 2019 Dec 18;8:4171-81.
12. Hagerty JR, Stanley RJ, Almubarak HA, Lama N, Kasmi R, Guo P, Drugge RJ, Rabinovitz HS, Oliviero M, Stoecker WV. Deep learning and handcrafted method fusion: higher diagnostic accuracy for melanoma dermoscopy images. IEEE journal of biomedical and health informatics. 2019 Jan 4;23(4):1385-91.
13. Rezvantalab, A., Safigholi, H., & Karimijeshni, S. (2018). Dermatologist level dermoscopy skin cancer classification using different deep learning convolutional neural networks algorithms. arXiv preprint arXiv:1810.10348.
14. Adegun A, Viriri S. Deep learning techniques for skin lesion analysis and melanoma cancer detection: a survey of state-of-the-art. Artificial Intelligence Review. 2021 Feb;54(2):811-41.
15. Xiang A, Wang F. Towards interpretable skin lesion classification with deep learning models. InAMIA annual symposium proceedings 2019 (Vol. 2019, p. 1246). American Medical Informatics Association.
16. Yap J, Yolland W, Tschandl P. Multimodal skin lesion classification using deep learning. Experimental dermatology. 2018 Nov;27(11):1261-7.
17. Spyridonos P, Gaitanis G, Likas A, Bassukas I. Characterizing Malignant Melanoma Clinically Resembling Seborrheic Keratosis Using Deep Knowledge Transfer. Cancers. 2021 Jan;13(24):6300.
18. Al-Masni MA, Al-Antari MA, Park HM, Park NH, Kim TS. A deep learning model integrating FrCN and residual convolutional networks for skin lesion segmentation and classification. In2019 IEEE Eurasia Conference on Biomedical Engineering, Healthcare and Sustainability (ECBIOS) 2019 May 31 (pp. 95-98). IEEE.
19. DeVries T, Ramachandram D. Skin lesion classification using deep multi-scale convolutional neural networks. arXiv preprint arXiv:1703.01402. 2017 Mar 4.
20. Augmented dermoscopic pigmented skin lesion from HAM10k https://www.kaggle.com/datasets/drscarlat/melanoma
21. https://www.coursera.org/learn/convolutional-neural-networks/lecture/AYzbX/data-augmentation
22. Data Augmentation | How to use Deep Learning when you have Limited Data - Part 2 https://nanonets.com/blog/data-augmentation-how-to-use-deep-learning-when-you-have-limited-data-part-2/
23. Imagenet classification with deep convolutional neural networks. A Krizhevsky, I Sutskever, GE Hinton - Advances in neural information processing systems, 2012.