Classification of Skin Cancer Images Using Convolutional Neural Networks

Kartikeya Agarwal[*1] *and Tismeet Singh*[2]

Abstract

Skin cancer is the most common human malignancy according to American Cancer Society. It is primarily diagnosed visually, starting with an initial clinical screening and followed potentially by dermoscopic (related to skin) analysis, a biopsy and histopathological examination. Skin cancer occurs when errors (mutations) occur in the DNA of skin cells. The mutations cause cells to grow out of control and form a mass of cancer cells. The aim of this study was to try to classify images of skin lesions with the help of Convolutional Neural Networks. Deep neural networks show humongous potential for image classification while taking into account the large variability exhibited by the environment. Here, we trained images on the basis of pixel values and classified them on the basis of disease labels. The dataset was acquired from an Open Source Kaggle Repository (Kaggle Dataset) which itself was acquired from ISIC (International Skin Imaging Collaboration) archive. The training was performed on multiple models accompanied with Transfer Learning. The highest model accuracy achieved was over 86.65%. The dataset used is publicly available to ensure credibility and reproducibility of the aforementioned result.

Keywords : Benign, Computer Vision, Confusion Matrix, Convolutional Neural Network, Deep Learning, Gradient Class Activation Maps, Machine Learning, Malignant, Skin Cancer, Transfer Learning

I. INTRODUCTION

The skin is the largest organ of the body which protects all the internal organs from the harsh world outside. It helps regulate temperature and protects against infections. The skin consists of three layers:

✎ Epidermis

✎ Dermis

✎ Hypodermis

Cancer can be traced when a mass of healthy skin cells changes its property and starts growing uncontrollably forming a mass called tumour. This tumour can be a benign growth, meaning that there is no harm to life and it cannot grow or spread to other areas, whereas it can be malignant which means that it is interrupting the natural flow of things and can grow and spread to other parts of the body.

Skin cancer is the most common form of cancer generally associated with over exposure to sun or other forms of high energy particles which destroy the genetic build-up of the cell. Skin cancer is the most common human malignancy according to American Cancer Society [1].

As the world steps out into a new millennium, the amount of pollution in the air is growing at an exponential rate paving the way for skin diseases like skin cancer to thrive. The Ozone layer is the only protection from the harmful Ultraviolet rays of the sun, depletion of which is the largest most probable cause of skin cancer. Due to the growing levels of air pollution, this layer is depleting

K. Agarwal*[1], *Student*, Computer Science, Department of Computer Science and Engineering, Netaji Subhas University of Technology, Sector-3, Dwarka, Delhi, 110 078. Email : kartikeya72001@gmail.com ;
ORCID iD : https://orcid.org/0000-0002-4168-2533
T. Singh[2], *Student*, Computer Science, Department of Computer Science and Engineering, Netaji Subhas University of Technology, Sector-3, Dwarka, Delhi, 110 078. Email : tismeet14sh@gmail.com ;
ORCID iD : https://orcid.org/0000-0002-0829-3261

geometrically and thus, the number of skin cancer cases are on the rise. Skin cancer can be broadly classified into two major categories: Melanoma (Malignant) and non-melanoma (Benign). Melanoma is one of the deadliest kinds of cancer. However, the detection of this cancer at an early stage can help in improving the chances of survival [2].

Various methods were considered and studied before the selection of the actual functions for this thesis. The best and the most advanced of the available Machine Learning resources were employed.

Skin cancer detection through histopathology can be achieved by augmenting the power of deep learning [3]. The classification of skin cancer lesions through specialized Convolutional Neural Networks can help the medical community in detecting and possibly curing the disease in early stages. Some early stipulations related to this cancer are discolouration or inflammation of the skin, itching and bleeding of skin patches, moles or red, waxy bumps on the skin due to cancerous cells.

The dataset was extracted from an open Source Kaggle repository which itself was a part of the ISIC (International Skin Image Collaboration) (Kaggle Dataset). The images were extracted and were split into test and train segments. An adequate level of image augmentation was applied to expand the dataset with modified versions of images. The test set contained 350 images while the training dataset contained over 2900 images comprising of two distinct classes, namely, "benign" and "malignant".

Within the scope of this research, we found that classification using K-nearest Neighbours or Support Vector Machines or even Decision Trees yielded low precision as well as accuracies. On further studying the mathematics behind classification, it came to light that the most sophisticated method to achieve the desired results was to use Deep Learning models. We tried various mathematical models with and without the use of Transfer Learning, but in the end, it was concluded that the depth and quality of activation provided by pre-trained models were of no match. Hence, we paired our knowledge of mathematics and created a Dense Convolutional Network model which gave an accuracy of over 86.6%.

Deep Learning, a robust and powerful protégé of Artificial Intelligence, provided us with tools to perform effective and accurate image classification. Its structure and operation resembles the human brain with neurons firing across the brain, sending information, classifying data, drawing inferences, and producing results. The structure used by Deep Learning consists of a stack of layers collectively known as a Neural Network. As the name suggests, Neural Networks work like neurons in a brain in identifying patterns and forming predictions.

The aim of the study is to try to get a reasonably faster and more accessible method of detection of skin cancer. The earlier it is detected, the higher is the chance that the patient will recover. According to the American Academy of Dermatology Association, if found early, skin cancer is highly treatable. The model trained is just a precursor to a skin biopsy. The only way to know for sure, the presence of skin cancer is to visit a Dermatologist and get a skin biopsy done.

Related Works

To keep the paper as relevant as possible, a lot of other papers and journals were referred to and consulted. In [4] the researcher recorded a maximum accuracy of 0.74 with his application of mathematical models on an augmented dataset of 3000 images. The author in his paper aspires to promote early diagnosis of melanoma via further development of Machine Learning models.

Further, a few more papers [5], [6], and [7] were consulted. Here, the authors achieved 81% accuracy and the proposed system made use of non-dermoscopic images of lesions from which he extracted the lesion regions and computed the results based on the colour and texture.

II. MATERIALS AND METHODS

A. Dataset

A dataset of 2947 histopathology images was considered for this study. The dataset was acquired from an open source Kaggle Repository [8], a subset of the ISIC [9] (International Skin Imaging Collaboration) archive. The dataset consists of two classes namely, benign and malignant. The test dataset comprises 350 images taken from the Kaggle repository. The dataset chosen is publicly distributed and Open source to ensure better credibility of the models.

The division among the two classes were as follows:

↳ For training and validation purpose

(1) 1600 benign skin mole images

(2) 1347 malignant skin mole images
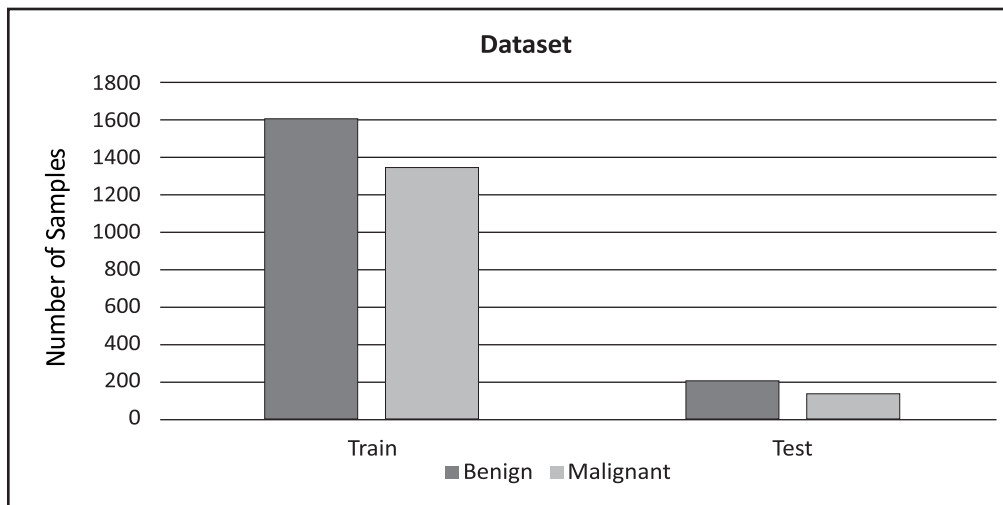
↳ For testing purpose

Fig. 1. Distribution of dataset

(1) 200 benign skin mole images

(2) 150 malignant skin mole images

All the images were resized to 176 x 176 for uniformity and improved testing metrics. Fig. 1 shows distribution of dataset.

B. Methodology

The aim of the study is to determine the efficiency and accuracy of different models in order to classify images between benign and malignant. The study takes into account multiple Transfer Learning models paired with a set of Dense Layers to predict the class of the image.

C. Augmentation

The dataset obtained was adequately augmented. Data Augmentation is a technique that can be used to exaggerate the dataset to introduce variability and expand the data so that a more robust and reliable model can be trained. The augmentation parameters were selected extremely carefully so as not to disturb the concerned image parameters that factor in deciding the outcomes. Augmentation included:

↪ The images were re-scaled to a re-scaling ratio of 1:255.0.

↪ The images underwent a shear of 0.2 degrees, which is the shear angle anticlockwise.

↪ The images were randomly flipped horizontally as well as vertically so as to preserve lesion marks and create new synthetic data.

↪ Zoom augmentation was introduced which randomly zooms the image in and either adds new pixel values around the image or interpolates pixel values.

All the images generated were passed through a generator and were passed into the model as a whole with a height and width of (176, 176). All the images were passed through all three channels, i.e., red, green and blue (R, G, B).

D. Convolutional Neural Network

Convolutional Neural Network is a class of neural networks applied in deep learning mostly to analyze image or visual data. A Convolutional Neuron is made up of learnable weights and biases which are tuned to generate desired results. Thousands and millions of these neurons when arranged in a specific order with other neurons result in a Convolutional Neural Network or CNN. CNNs are mainly applied to classify and order image data and cluster them together if they appear analogous, and then perform object recognition.

Filters are small matrices that we engage across the dataset with the help of a sliding window. The depth of the image is the same as the input, for a coloured RBG image value of depth is 4, a filter of depth 4 would also be applied to it. Fig. 2 shows visual representation of Convolution function.

The next part is, we apply a Rectifier function (generally Rectified Linear Unit) to increase the non-linearity of the Network. Finally, a pooling layer is used to

downsample the features, which is applied through the volume of the image. We finally use a fully connected layer which flattens the entire 3D feature matrix into a single column which is then fed further to the neural network for processing. All the layers, Convolutional layer, Activation layers, Pooling layers, and Fully Connected layers are interconnected as previously mentioned to form a Convolutional Neural Network. Fig. 3 shows visualization of a Neural Network.

E. ReLU

The most frequently utilized method of activation function in Deep Learning algorithms is the Rectified Linear Unit. The function returns 0 if it receives any negative input, but for any positive value x, it returns that value back. So, it can be written as:

$$(x) = \max(0, x) \tag{1}$$

F. Leaky ReLU

Leaky ReLU activation function definition says that instead of using 0 for negative values of inputs(x), we define it as an extremely small linear component of x. It helps bypassing the dead neuron problem caused by Rectified Linear activation. Leaky ReLU is a better
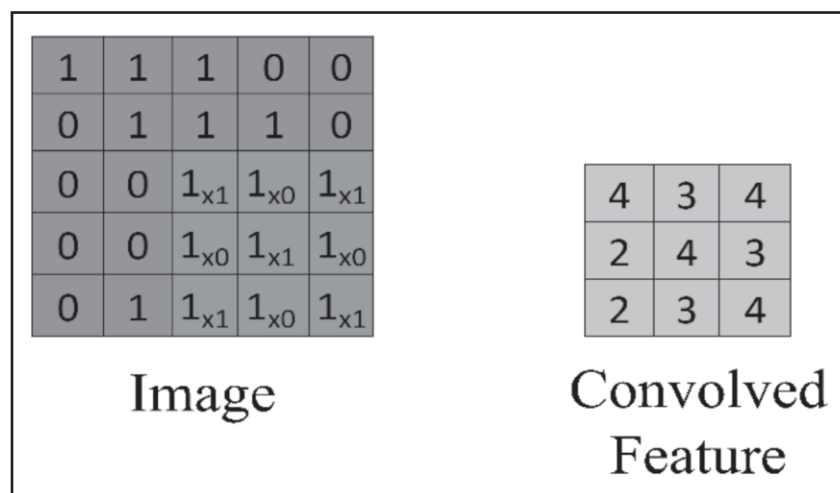

Fig. 2. Visual Representation of Convolution Function
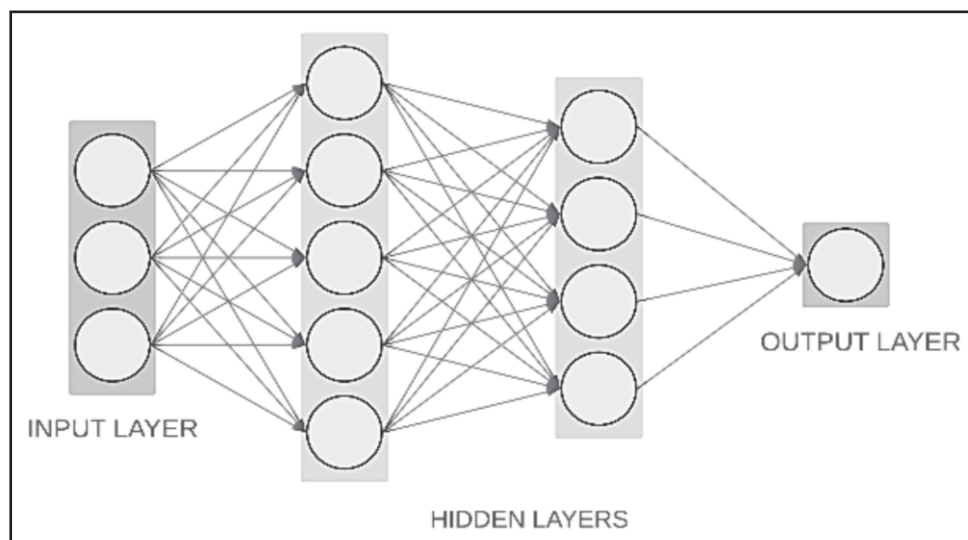

Fig.3. Visualization of a Neural Network

activation function as compared to Sigmoid and Tanh because it is less prone to saturation and has better sensitivity as compared to other activation functions. Leaky ReLU looks like a linear activation function but is actually a nonlinear function thereby, helping in defining complex associations in the data.

$$(x) = \max(a, x), \text{ where } a \text{ is a very small value} \quad (2)$$

G. Batch Normalization

Normalization is a technique of changing the variables in a particular type of data to a specific scale without affecting the contrast or shape of the image. Batch Normalization refers to a procedure of normalization of the effect to a layer for every batch. It helps in reducing the amount of computational power and number of training epochs required for a Deep Learning model. When a deep neural network is trained, the inputs of layers may change when the weights are updated per mini batch. This can cause internal covariate shift, that is changes in the distribution of inputs to a layer for each mini batch.

Batch Normalization for a Neural Network means scaling the output of the layer by normalizing the activations of each input variable per mini batch. Batch Normalization helps to make the model more robust and less prone to effects caused by hyperparameter tuning.

H. Flatten

Flatten is a technique adopted in the conversion of data to a 1D array as an input for the next layer. The output vector or the feature matrix map is flattened to get a single feature vector which can be used as an input to the next layer to be processed further.

I. Pooling

Pooling is one of the critical layers in a deep neural network. The pooling layer is responsible for the reduction of dimensionality or spatial size of a feature matrix. The pooling layer helps in reducing the computational power required to train a neural network. It is also critical in the extraction of dominant features in the data.

There are two major types of pooling layers which can be incorporated in a deep neural network: Max Pooling and Average Pooling (Fig. 4).

Max Pooling, as the name suggests, gives the maximum value for the image area covered by the kernel. Similarly, Average Pooling returns the average of the values for the image area covered by the kernel. In our study, we have used Max Pooling instead of Average Pooling as it helps in the elimination of noise or noisy features along with dimensionality reduction, whereas Average Pooling only helps in dimensionality reduction. Thus, the pooling layer helps in picking up meaningful information from a sea of data which is vital for helping a model learn the features in a dataset. Elimination of noise and provision of only useful features help in preventing overfitting and speeds up computation.
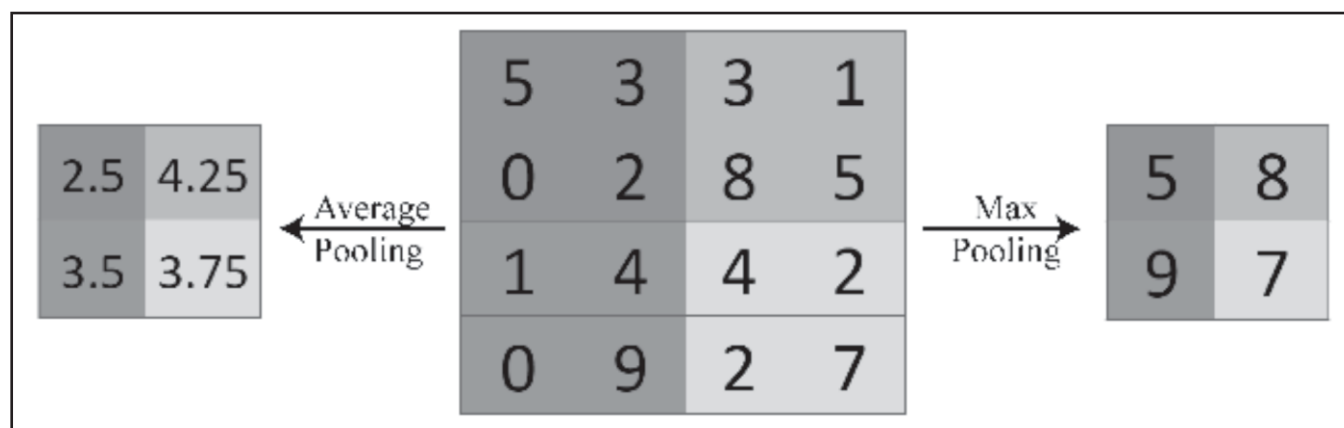

Fig. 4. Average Pooling vs Max Pooling

J. Dropout

Dropout Layers are used to decrease the level of overfitting in a large Network. Simply put, a dropout layer will ignore a specified number of random neurons while training so as to prevent overfitting and bias in the output of the model.

K. Categorical Cross Entropy

Categorical Cross Entropy is a loss function employed in a Network where there is a multiclass classification.

$$Loss = -\sum_{i=1}^{output\ size}(y_i \cdot log \hat{y}_i) \qquad (3)$$

Where,

y_i : ground truth, and

$\hat{y}$: predicted class.

This function is an excellent method of differentiating between different classes or discrete probability distributions.

L. Optimizer and Metrics

Optimizers help in tuning the weights such that the loss function is minimized. Their objective is reducing losses and providing better results. The 'adam' optimizer was used for training the model with a variable learning rate. The learning rate was initialized with a value of 0.0001 but was modified or more accurately, reduced after the validation-loss crossed a specific patience level with a reduction factor of 10^{-1}. The metric used for the model was 'accuracy' and Early Stopping was employed to prevent overfitting or overtraining of the model. Here too, the monitored hyper-parameter was the validation-loss with a high yield patience level.

Activation functions are generally used to normalize the output of a network to a probability distribution over the output classes. The activation function considered for the model's feature extraction layer was SoftMax.

$$\sigma(\vec{z})_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}} \qquad (4)$$

where,

σ = softmax;

$\vec{z}$ = input vector;

e^{z_i} = standard exponential function for input vector;

K = number of classes in the multi-class classifier;

e^{z_j} = standard exponential function for output vector.

M. Models Used

The study required meticulous comparison among a plethora of transfer learning models. Several transfer learning models were considered for this study namely, DenseNet, XceptionNet, ResNet, and MobileNet. DenseNet201 was chosen because of its strong gradient flow, better back propagation, and smooth decision boundaries. Xception Net encompasses modified depth wise separable convolution and has the same number of parameters as InceptionV3 but greater computational efficiency. Resnet50 illustrates the concept of skip connection which eliminates the problem of vanishing gradients (problem encountered while training very deep networks) and helps the model to learn the identity function properly for better results. MobileNetV2 also consists of depth wise separable convolution which provides reduced complexity and size.

N. Gradient Class Activation Maps (Grad CAM)

Grad CAM is a technique for analysis of the Activations generated by the trained model. Grad CAM is simply a heat-map generated by the Convolutional layers depicting the importance of a certain feature or the part of an image. It encapsulates the area or region of interest(ROI). Marked area in Fig. 5 depicts the region of interest, while the rest is area of low activation. Fig. 6 shows the structure of a Neural Network.

O. Training

The model was trained on the previously mentioned dataset after performing a training and validation split. The aforementioned models were trained with 60 epochs though, as mentioned earlier. Early Stopping was employed on account of the validation loss. The steps per epoch were decided as the batch size divided by the total number of images in the database. Different models yielded different results which are plotted in Section III.
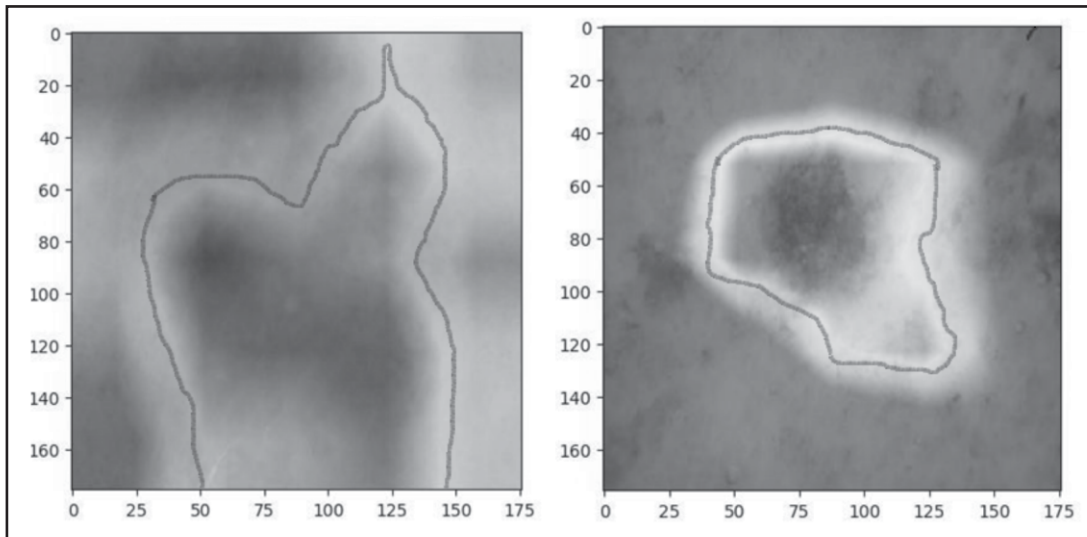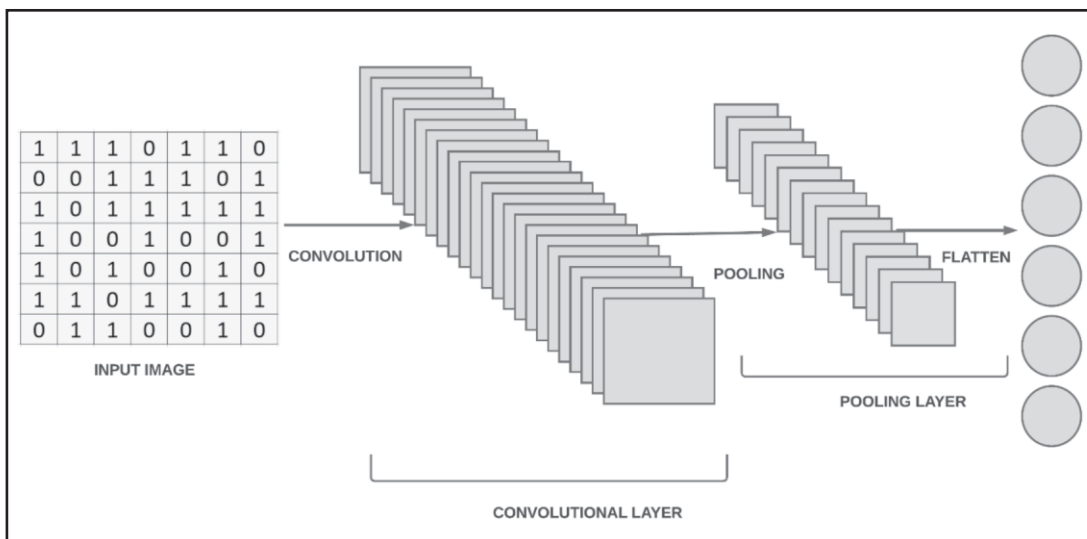
Fig. 5. Gradient Class Activation Maps

Fig. 6. Structure of a Neural Network

Early-stopping was employed to prevent over training It monitored validation loss as a parameter and at every point, when the model performed worse, the previous weights were restored. We also employed a conditional reduction in learning rate by 1/10 so that whenever the rate of convergence of the function plateaued, the learning rate would be reduced for a more precise reduction in the slope of the hyper-plane.

III. RESULTS

The models were tested on a test split of 350 images; 150 from malignant class and 200 from the benign class. Model accuracy, precession, recall, and $F1$-score were calculated for all the models for better insights of their efficiency and operability. Confusion matrix was plotted for each model based on the test data and the metrics were calculated according to the following formulae:

$$Precision = True\ Positive\ /\ (True\ Positive\ +\ False\ Positive) \qquad (5)$$

$$Recall\quad = True\ Positive\ /\ (True\ Positive\ +\ False\ Negative) \qquad (6)$$

$$Accuracy = True\ Positive\ +\ True\ Negative\ /\ (True\ Positive + False\ Negative + False\ Positive + True\ Negative) \qquad (7)$$

$$F1Score = (2 \text{ x } Recall \text{ x } Precision) / (Recall + Precision) \quad (8)$$

A. DenseNet201

DenseNet201 performed extremely well on the Validation split as well as on the final test split. The Neural Network paired with Dense Layers which in turn complemented with Leaky ReLU gave high accuracies as well as F1Score. DenseNet helped solve the problem of vanishing gradients and enabled feature reuse and propagation leading to substantial results and accuracy.

For visual insights, the Confusion Matrix derived from the test set and the graphical representation of the metrics used during training and validation steps are shown in Fig 7. The accuracy recorded over the test set for this model was 86%. Fig. 8 shows Confusion Matrix - DenseNet201.

The class activation map for this model was prepared where the activation provided by the last layer is represented as an overlayed heat map. Marked area depicts the region of interest while the rest represent area of low activation (Fig. 9).
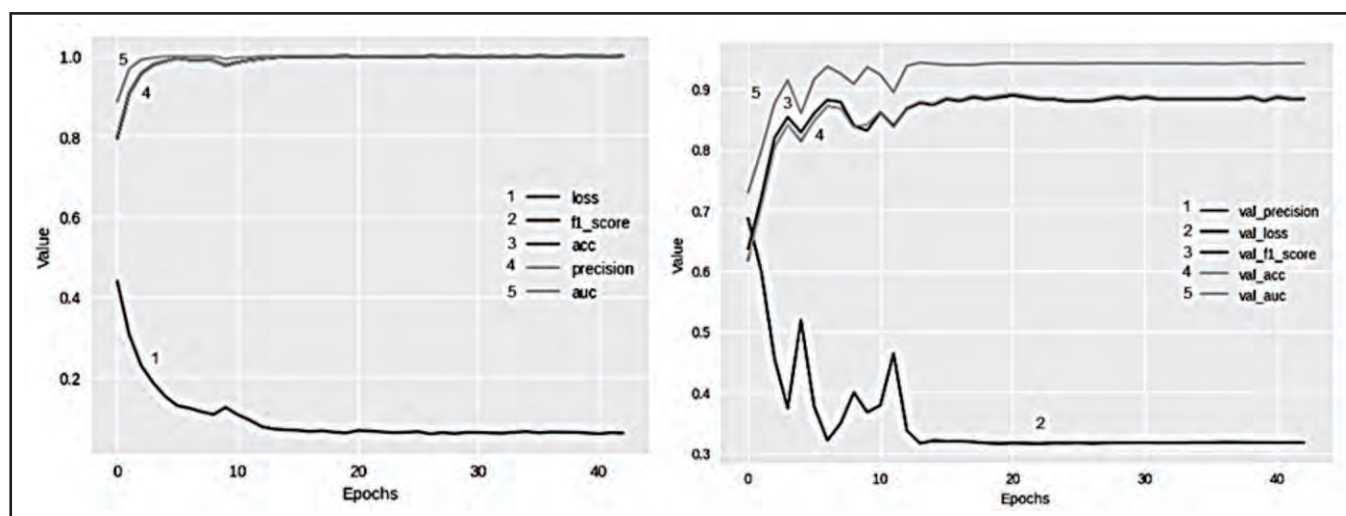

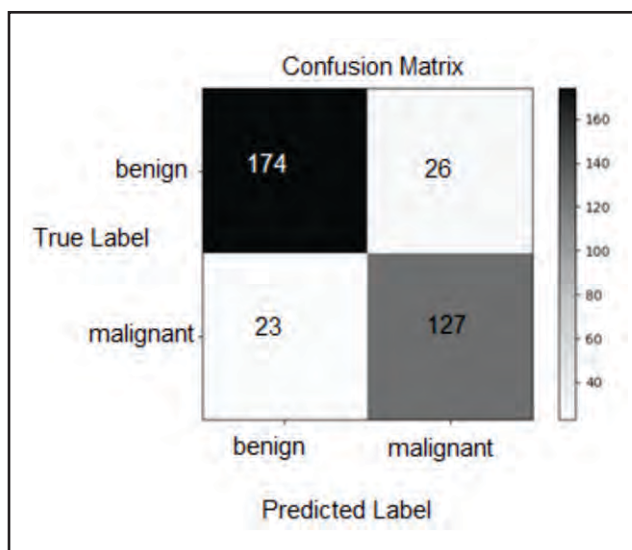**Fig. 7. Graphical Representation of Training and Validation Metrics - DenseNet201**

Fig. 8. Confusion Matrix - DenseNet201

B. ResNet50

The ResNet50 architecture is equipped with skip connections, that is, the input to a layer can be directly passed to some other layer, thereby improving the performance of this neural network over others. Along with this, it also helps in solving the problem of vanishing gradients through identity mapping. The confusion matrix along with the metrics used during the training and validations steps were plotted to draw comparisons with the other models used in this study. The model gave an accuracy of over 86.57% on the test data. Fig. 10 shows graphical representation of training and validation metrics - ResNet50. Fig. 11 shows Confusion Matrix - ResNet50.
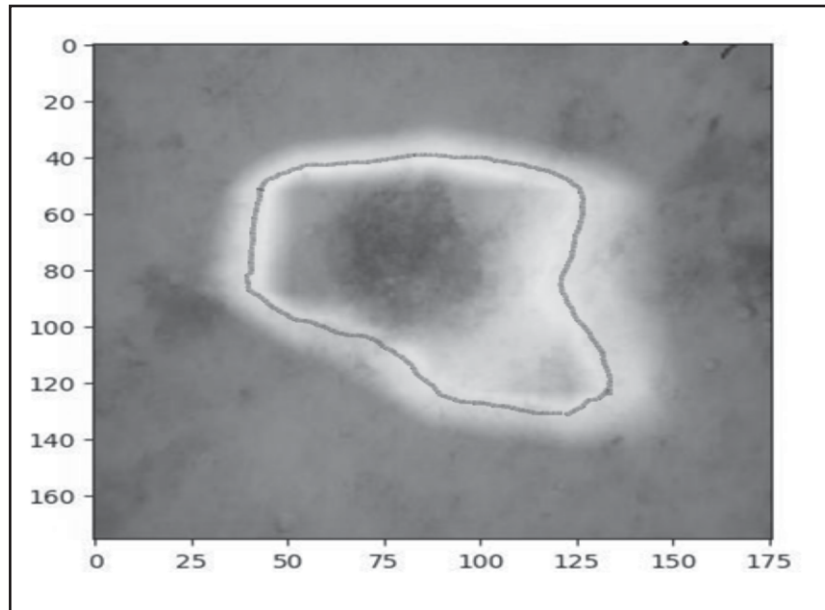
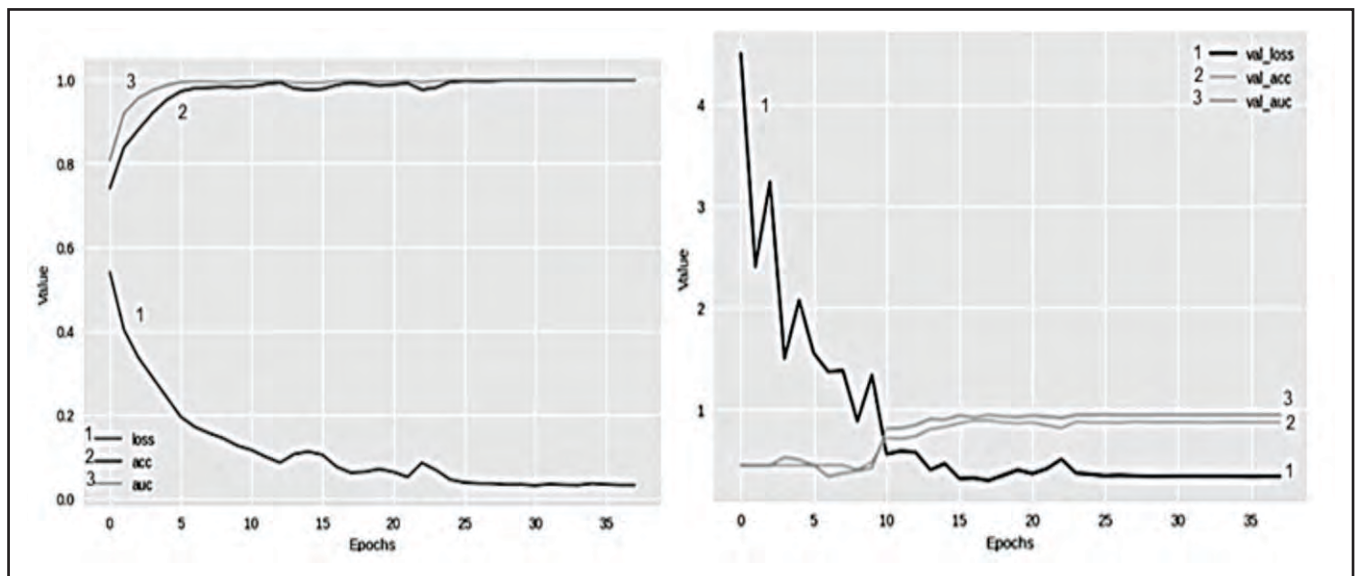
Fig. 9. Grad-CAM - DenseNet201

Fig. 10. Graphical Representation of Training and Validation Metrics - ResNet50
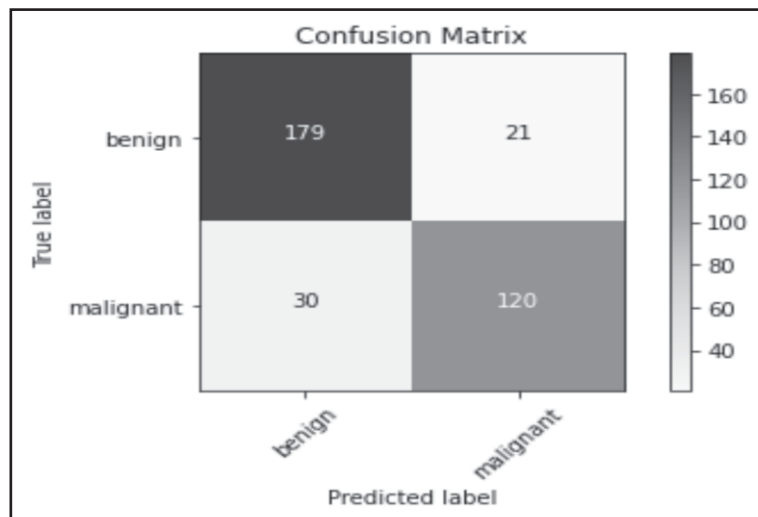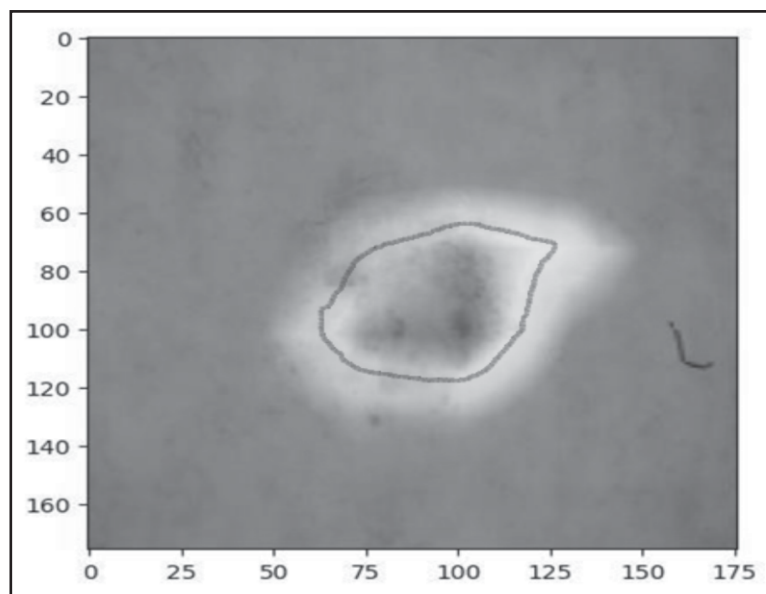
Fig. 11. Confusion Matrix- ResNet50

Fig. 12. Grad-CAM - ResNet50

The class activation map for this model was prepared where the activation provided by the last layer is represented as an overlayed heat map. The marked area depicts the region of interest while the rest represents area of low activation. Fig. 12 shows Grad-CAM - ResNet50.

C. XceptionNet

The XceptionNet architecture has 36 convolutional layers forming the feature extraction base of the network. The 36 convolutional layers are compiled into 14 different modules in linear geometry. This arrangement helped us gain immense accuracy on the succeeding organization of the dense and batch normalization layers.

The model was trained on the same set of training dataset and the performance of the Model was analyzed consistently with the same set of validation images. The validation graph and training graph were plotted and the confusion matrix was analyzed for the test dataset to get a better understanding of the results and for the analysis of the performance of the model. The model gave an accuracy of over 82.5% as is depicted in Fig. 13 and Fig. 14.
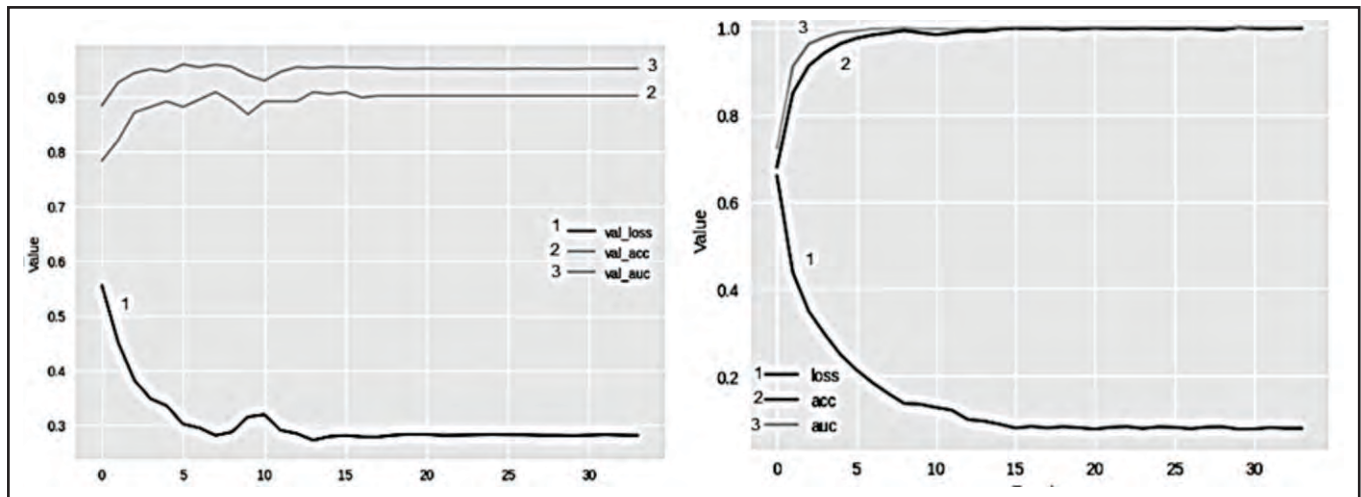
Fig. 13. Graphical Representation of Training and Validation Metrics - XceptionNet
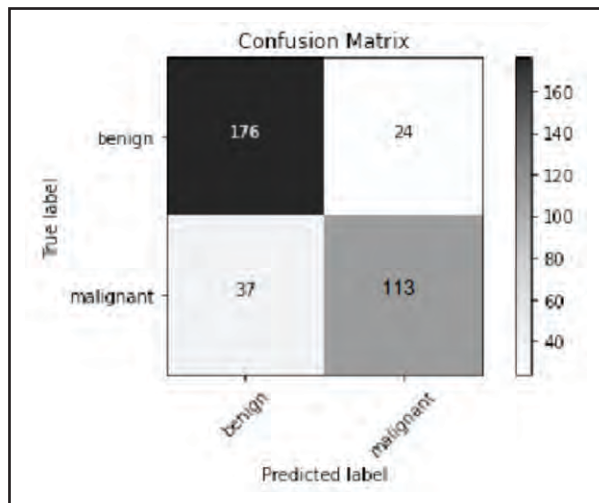

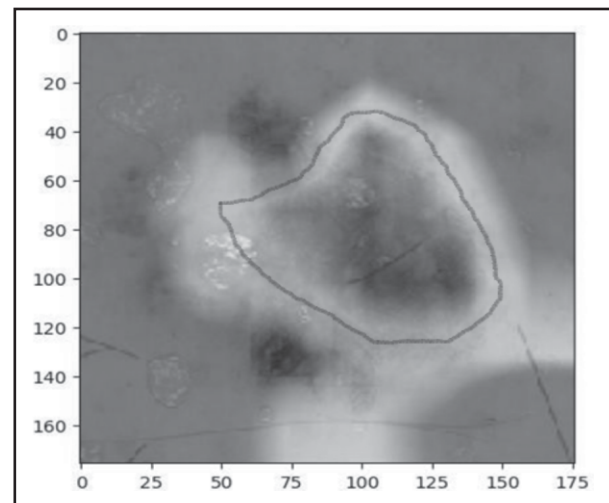**Fig. 14. Confusion Matrix – XceptionNet**

Fig. 15. Grad-CAM - XceptionNet

The class activation map for this model was prepared where the activation provided by the last layer is represented as an overlayed heat map. The marked area depicts the region of interest (Fig. 15) while the rest of the area represents area of low activation.

D. MobileNet

MobileNet is a simpler and lighter model, usually trained for handheld lower computing powered devices. The contemplation behind the usage of MobileNet was to enable the scanning for malignant or benign lesions using globally accessible devices, for example, smartphones. This model was introduced to bring the power of this model to the common man. The model performed very decently considering its more efficiency focused nature.

This architecture too was trained over the entirety of the dataset. The model gave an accuracy of over 80.8% as documented in the tabulated data. Fig. 16 shows graphical representation of training and validation metrics – MobileNet.

The class activation map for this model was prepared where the activation provided by the last layer is represented as an overlayed heat map. The marked area depicts the region of interest while the rest represents area of low activation. Fig. 18 shows Grad-CAM – MobileNet.

E. Comparisons

The results from all the models were compiled and the metrics plotted. Some metrics were tabulated based on the results obtained from the prediction of the test set. The
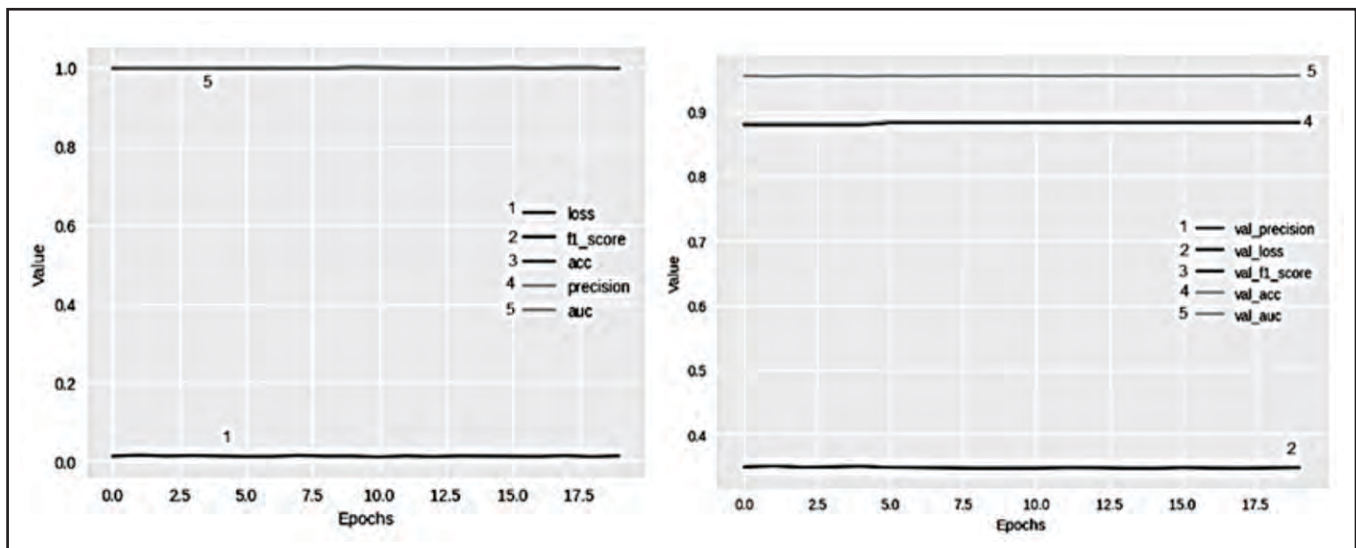
Fig. 16. Graphical Representation of Training and Validation Metrics - MobileNet
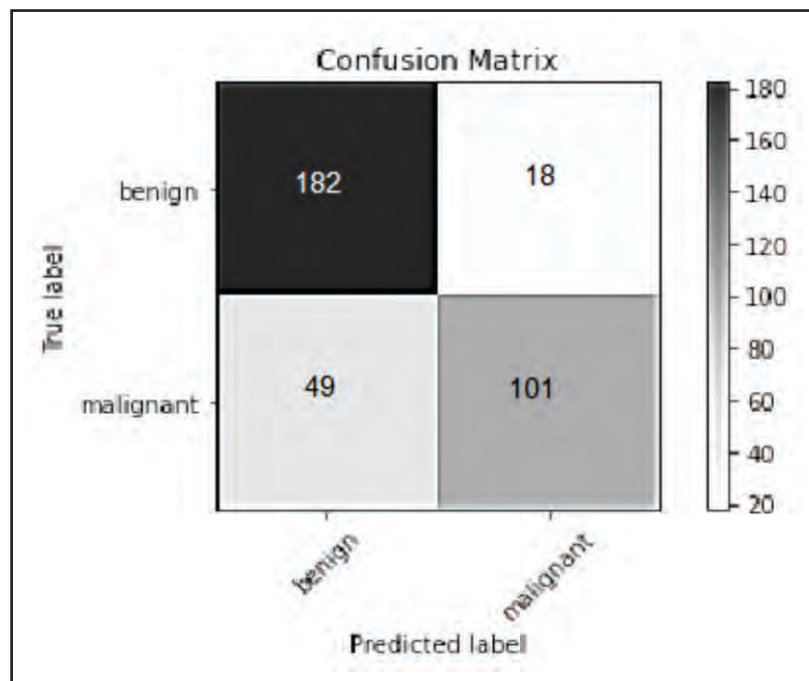

Fig. 17. Confusion Matrix – MobileNet

metrics Accuracy, Precision, Recall, and $F1$-Score were used to compare the model performance among all the chosen architectures (Table I, Fig. 19).

IV. CONCLUSION

Machine Learning and Artificial Intelligence based classification can help in instantaneous diagnosis and classification of Skin Cancer and like diseases. The models developed by us are proof of the concept that cost-
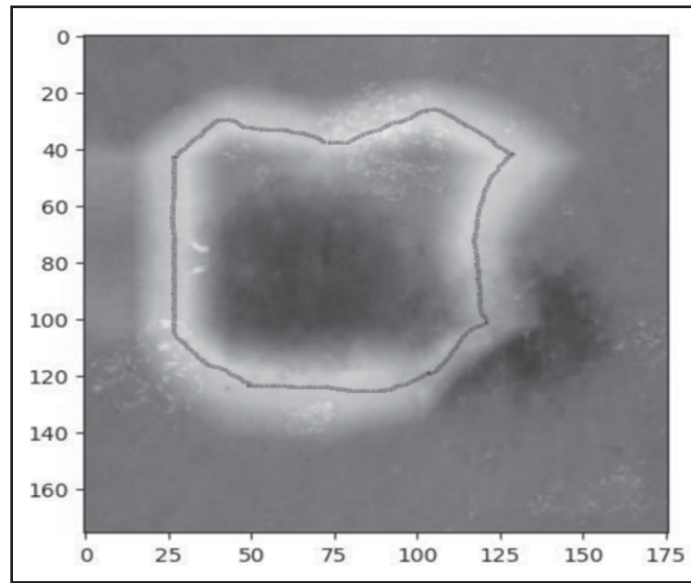
Fig. 18. Grad-CAM – MobileNet

TABLE I.
COMPARISON OF RESULTS FOR THE MODELSUSED

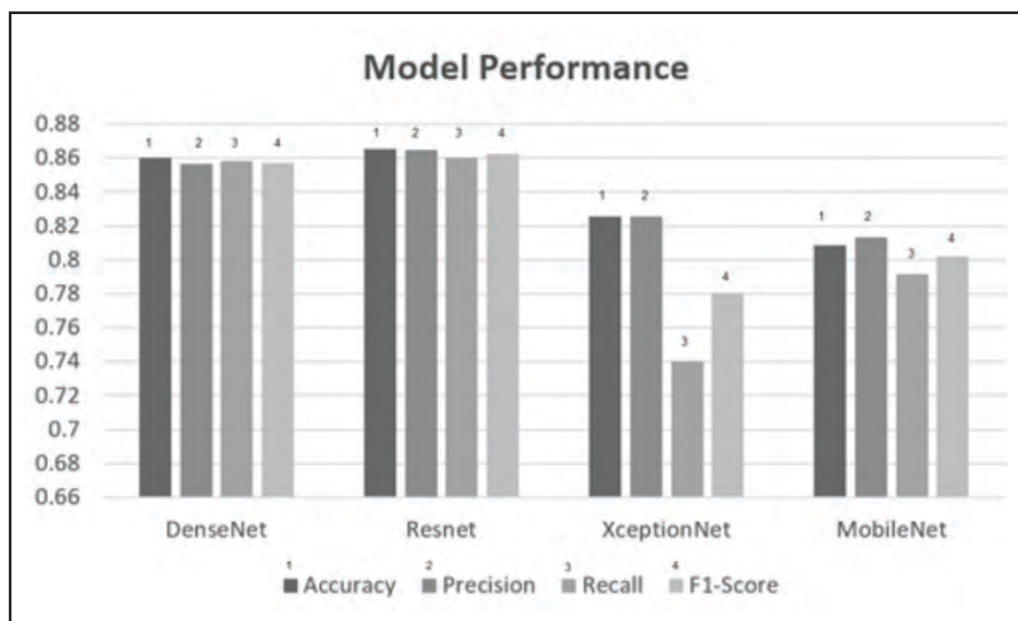|	DenseNet	Resnet	XceptionNet	MobileNet
Accuracy	0.86	0.86571	0.82571	0.80857
Precision	0.8566	0.8648	0.82555	0.81309
Recall	0.85833	0.86	0.73991	0.79166
F1-Score	0.85746	0.86239	0.7803	0.802232

Fig. 19. Graphical Representation of Overall Results

effective, user-friendly, and non-invasive Machine Learning and Artificial Intelligence based methods can be developed for the first step towards battling Skin Cancer and other diseases. During the research process, each model we trained was a successor of the last one, every new model we created was based on the faults of the previous one. All the insights gained from the precursing models were employed to improve the accuracy of upcoming new models.

In the study presented, the final models were trained and analyzed to improve the efficiency and reliability of the methods used for the detection of skin cancer. Comparative analysis between various models helped in determining the best model which can have real world applications in the early detection of the disease. The models used in the study presented different fronts in detecting the disease, with each model equipped with better techniques and features. The results obtained from this study have been presented both in written and visual form (vide the help of graphs and confusion matrices) for the perusal of the scientific community.

Further, we hope that our findings will help with the early detection of Cancer and be a useful contribution in fighting this disease. However, human judgement is indispensable. Hence, with our model we are not trying to replace a Dermatologist but merely aid with the predictions.

V. FUTURE SCOPE

This study can be understood and referred to as the basis for further analysis in the future. The elementary results explored here may just be the beginning of an exploratory analysis of the power of Machine Learning in the field of Cancer study. Such studies may one day help in general to discover the presence of any malignant tumour easily and get treated in time.

AUTHORS' CONTRIBUTION

Kartikeya Agarwal and Tismeet Singh are the authors and have performed the entirety of the research work as mentioned in the paper. Both the authors conceived of the presented idea, developed the theory, and trained the models. All authors discussed the results and contributed to the final manuscript.

CONFLICT OF INTEREST

The authors proclaim that there is no irreconcilable situation with respect to the distribution of this paper. The authors declare that they have no financial interests or personal affiliations that could have influenced the study and findings.

FUNDING ACKNOWLEDGEMENT

REFERENCES

[1] "Cancer Facts & Figures 2016," Amer. Cancer Soc., 2016. [Online]. Available: http://www.cancer.org/acs/groups/content/@research/documents/document/acspc-047079.pdf

[2] "How can I tell if I have skin cancer?" Amer. Academy of Dermatology Assoc. [Online]. Available: https://www.aad.org/public/diseases/skin-cancer/find/know-how

[3] U. Ansari and T. Sarode, "Skin cancer detection using image processing," *Int. Res. J. Eng. Technol.*, vol. 4, no. 4, pp. 2875–2881, 2017. [Online]. Available: https://www.irjet.net/archives/V4/i4/IRJET-V4I4702.pdf

[4] M. A. Ottom, "Convolutional Neural Network for diagnosing skin cancer," *Int. J. Adv. Comput. Sci. Appl,,* vol. 10, no. 7, doi: 10.14569/IJACSA.2019.0100746.

[5] I. Giotis, N. Molders, S. Land, M. Biehl, M. F. Jonkman, and N. Petkov, "MED-NODE: A computer-assisted melanoma diagnosis system using non-dermoscopic images," *Expert Syst. Appl.*," vol. 42, no. 19, pp. 6578– 6585, 2015, doi: 10.1016/j.eswa.2015.04.034.

[6] M. Hasan, S. D. Barman, S. Islam, and A. W. Reza, "Skin cancer detection using Convolutional Neural Network," in *ICCAI'19: Proc. 2019 5th Int. Conf. Comput. Artif. Intell.*, pp. 254–258, 2019, doi: 10.1145/3330482.3330525.

[7] M. M. Vijayalakshmi, "Melanoma Skin Cancer detection using image processing and Machine Learning," *Int. J. Trend Scientific Res. Develop.*, vol. 3, no. 4, pp. 780–784, 2019, doi: 10.31142/ijtsrd23936.

[8] "Skin cancer: Malignant vs. benign." [Online]. Available: https://www.kaggle.com/fanconic/skin-cancer-malignant-vs-benign

[9] "Welcome to ISIC." ISIC. https://www.isic-archive.com/#!/topWithHeader/wideContentTop/main

About the Authors

Kartikeya Agarwal is a studentof Computer Science Engineering. He is currently in the 3rd year of his Bachelor's Degree. He has great fervor for Machine Learning and Data Science and works on real-life problems with the assistance of programming.

Tismeet Singh is a Computer Science undergraduate at Netaji Subhas University of Technology. He is a Machine Learning and Artificial Intelligence enthusiast currently working in the field of Computer Vision and Natural Language Processing.