

# Detection and Classification of Skin Cancer Using Deep Learning Techniques

Meer Aram Omer  
MMU ID: 22451299  
Department of Computing and  
Mathematics  
Manchester Metropolitan University  
22451299@stu.mmu.ac.uk

**Abstract**— Skin cancer is the most common disease in the US (United States). Health care providers and patients alike are tasked with identifying suspicious complexion lesions to diagnose skin cancers in an early stage and treat them in a short time. Recently, many innovative skin cancer detection technologies and techniques have been developed to increase diagnostic accuracy for skin cancers. This research paper presents a comprehensive study of using three popular deep learning models DenseNet121, ResNet18, and AlexNet to classify 25,331 labeled images in the ISIC-2019 dataset. The objective is to evaluate the performance and accuracy of each model and compare the results obtained in categorizing two classes which are Benign and Malignant. This study also highlights the key differences between the metrics of each model. The research found that all models achieved good accuracy in the prediction results. However, DenseNet121 demonstrated better performance in classifying the targets.

**Keywords**— Skin Cancer Detection, Deep Learning, ResNet18, AlexNet, DenseNet121.

## I. INTRODUCTION

Skin cancer represent a major health problem and it is one of the most dangerous and active types of cancers [1]. Early diagnosis is dependent on patient attention and accurate assessment by a medical practitioner. Therefore, if it is detected at early stages, it is more likely to be cured. Generally, skin cancer is categorized into two classes: melanoma and non-melanoma. Although melanoma is rare, it is the most hazardous type of skin cancer, and it has the highest rate of death globally. To beat this problem and improve the reliability of the diagnosis process, it is very important to develop computational tools and system for automated diagnosis that operate on quantitative measures. In this paper the performance of three types of deep convolutional neural network architecture namely ResNet18, DenseNet121, and AlexNet will be compared and discussed. Regarding the dataset, ISIC-2019, which is composed of 25,331 dermatoscopic images of common pigmented skin lesions, have been utilized to train all the models and predict the type of skin lesions. Furthermore, Recall, Precision, and Accuracy is also used to evaluate the performance of the models and compare them to explore which of them perform better.

The motivation behind this research is to develop a prototype capable of extracting features and classifying skin cancer types from ISIC dataset. Additionally, ISIC is imbalanced which can impact the performance of the models and make them biased. This issue will be also addressed in this paper through applying some techniques along with striving to improve the accuracy of the models by experimenting different hyperparameter optimization, data augmentation etc.

## II. DATASET

Small size of dataset and lack of diversity is always a huge problem which all neural networks suffer from. ISIC tackled this issue by providing a big and diverse collection of dermatoscopic images from different population [2]. The dataset consists of 25,331 images distributed across 2 different classes, including Benign and Malignant. The data has been divided into train/test/validation, with %70 percentage for training and %15 percentage each for test and validation set. The following preprocessing and transformation have been applied to the data:

- 1) The images have been resized to dimensions of 224 by 224.
- 2) Some data augmentation techniques are also applied to the images such as RandomRotation, RandomVerticalFlip, ColorJitter and etc. In addition, normalization is also applied to improve convergence and facilitate models training.
- 3) Eventually, the images are organized into 32 batches for AlexNet and DenseNet121, and ResNet18. This decision is based on the observation that ResNet18 demonstrated higher precision and accuracy when trained with this batch size.

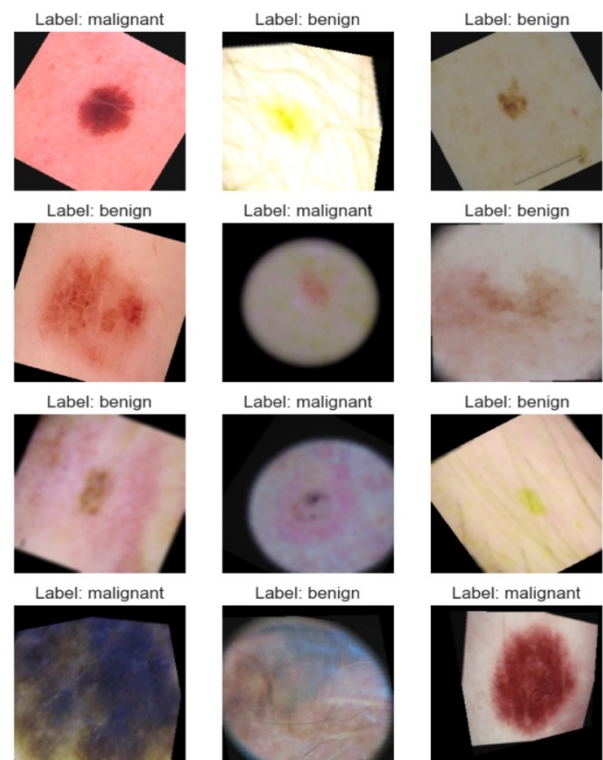


Figure 1 Sample images from ISIC dataset

### III. RELATED WORK

Research conducted by [3] focuses on classifying ISIC, MED-NODE, Derm (IS & Quest) images using Deep Convolutional Neural Networks. The main objective is to implement a highly accurate model by utilizing transfer learning and data augmentation techniques. The study develops a pre-trained Alex-Net and by fine-tuning the network weights the researchers improved the accuracy of the model. Their proposed method has outperformed existing models for classifying skin cancer and achieved accuracy of 96.86%, 97.70%, and 95.91% for MED-NODE, Derm (IS & Quest) and ISIC datasets respectively. Kassem *et al.* [4] utilized pre-trained GoogleNet to classify eight types of skin cancer including melanoma, melanocytic nevus, basal cell carcinoma, actinic keratosis, benign keratosis, dermatofibroma, vascular lesion, and Squamous cell carcinoma. ISIC 2019 dataset has been used to train the model which is composed of 25331 dermatoscopic images. Furthermore, they applied transfer learning in three steps. Firstly, the model is selected which is GoogleNet that is trained with ImageNet. Secondly, the last layer of the pre-trained model is replaced. Lastly, they reused the model's layer for new tasks by using fine-tuned layers. To evaluate their method they used accuracy, sensitivity, specificity, and precision with 94.92%, 79.8%, 97%, and 80.36%, respectively. Data augmentation strategies have been proposed by Srinivasu *et al.* [5] to balance various forms of lesions to the same range of images. The proposed model, which is predicated on the LSTM and MobileNet V2 approaches, was found to be effective in classifying and detecting skin diseases with little effort and computational resources. When using CNN transfer learning, Mahbod *et al.* [6] confirmed that image size affects skin lesion categorization performance. They also showed that image cropping outperforms image resizing in terms of performance. Hosny *et al.* [7] used pre-trained AlexNet with transfer learning. As initial values, the parameters from the original model were used. and the weights of the last three replaced layers were randomly initialized. ISIC 2018, the most recent public dataset, was used to test the suggested technique.

Many researchers have endeavored and contributed to improve the performance of models which has been built for skin lesion classification by utilizing different machine learning and deep learning techniques. Moreover, they have worked on different datasets incorporating MED-NODE, Derm (IS & Quest), ISIC, HAM10000 etc.

### IV. METHODOLOGY

In this paper three convolutional neural network architectures have been utilized to predict tumor malignancy incorporating AlexNet, ResNet18, and our proposed method which is pre-trained DenseNet121.

#### 1) AlexNet Convolutional Neural Network

AlexNet is a famous CNN architecture which has been developed by [8]. It consists of eight layers, five convolutional layers and three fully connected layers. Because of its deep architecture, it can learn complex features from the dermatoscopic images which make it one of the most efficient architectures for our task. AlexNet can yield 4096-dimensional feature vector for each image, which contains the activations of the hidden layer immediately prior the output layer. AlexNet uses ReLU activation function which is defined as  $f(x) = \max(0, x)$ . The function returns zero for

negative input and return the input itself when it is positive. This activation function introduces non-linearity to the model so it can learn complex pattern from the input images. Another crucial component in AlexNet architecture is Batch Normalization. This component is applied after every activation function, and it helps stabilize and speed up the training process.

#### 2) DenseNet121 Convolutional Neural Network (our proposed method)

DenseNet is another convolutional neural network architecture which has been developed by [9]. DenseNet is a great architecture to increase the depth of the neural network. DenseNet allows a dense block to concatenate feature maps from every earlier layer and link each layer to every other. By utilizing this mechanism, the network's learning ability significantly amplified, and the gradient flow is improved in training phase. There is different version of DenseNet namely DenseNet-121, DenseNet-169, DenseNet-201 etc. Due to limited computational power only DenseNet-121 was used. Another technique which has been used with our proposed method is transfer learning. Transfer learning is a technique, which means a model like DenseNet121 is reused for skin lesion classification task which it has been trained previously with a similar task. Due to insufficient number of images in the available datasets and the massive resources required for deep learning algorithms, these datasets are not suitable to train a deep neural network from the beginning. Therefore, Transfer learning overcame this issue for us.

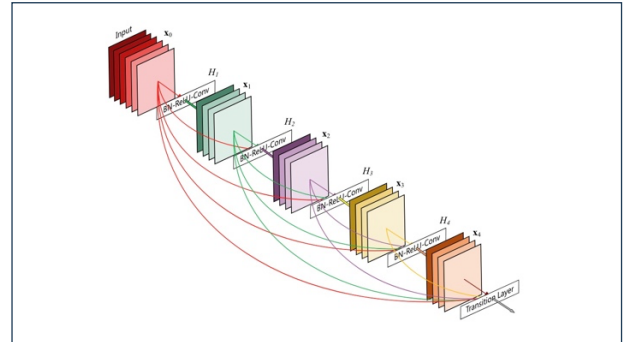


Figure 2 5-layer dense block with a growth rate of  $k=4$

#### 3) ResNet18 Convolutional Neural Network

ResNet18 is a deep learning model in which each layer compares its results to the weighted results of the previous layers through residual functions. This helps the model to pass information more easily (skip connection) between the layers while preserving the original outputs, enabling more precise predictions [10]. As shown in Figure 2 below, the identity of the input  $x$  is passed and added to the output of another weighted layer  $F(x) + x$ :

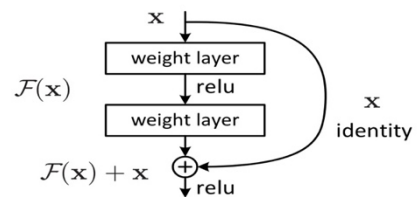


Figure 3 Illustration of Resnet Residual Functions

This ensures that the vanishing gradient problem is prevented while improving the performance of the model. There are different types of ResNet networks, which are determined by the number of layers used, such as ResNet 18, 34, 101, 152, and 1202. For this project, ResNet18 is considered as the optimal model to classify the ISIC images. As shown in Figure 4, the model consists of 18 layers, and each layer has residual blocks that contain a convolutional layer, batch normalization, ReLU, and a fully connected layer that performs the classification of the classes.

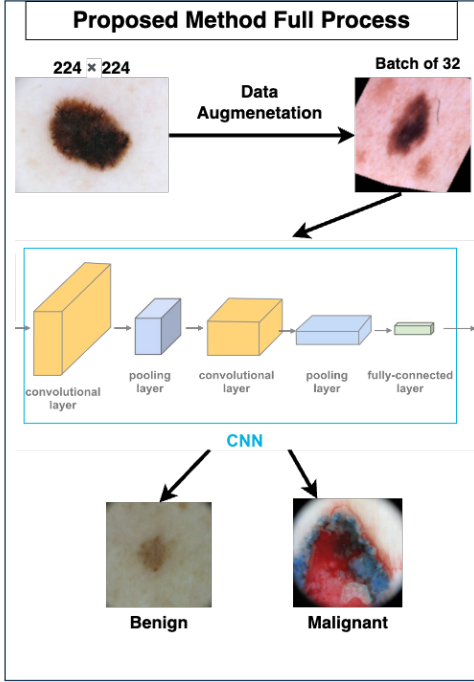


Figure 4 pre-trained DenseNet-121 Process

ISIC-2019 is imported via PIL (Python Imaging Library) and combined. As mentioned in section II, the data has been divided into train/test/validation. Then, As shown in figure 3, some preprocessing and data augmentation was applied to the images namely resizing the images to dimensions of 224 by 224, RandomHorizontalFlip, RandomRotation etc. Once the sets are split, the data is loaded, and batches are prepared using a shuffled batch size of 32, resulting in 1278 train, 274 test, and 275 validation batches for the AlexNet, ResNet18 and our proposed model (pre-trained DenseNet121). After creating the batches, the train and test step functions are developed. The train step function trains the model with each batch of images provided. First, the train step function makes a prediction with an image, calculates the loss and accuracy. Then, backpropagation is performed to compute the gradients based on the loss function. The gradients are computed using the chain rule to determine the magnitude and direction of adjustment in the model parameters. Once the gradients are calculated, optimization is performed to update the model parameters based on the computed gradients. The train step function also computes the loss and accuracy ratio to track the model's performance. This process is repeated for all the batches in the train, test, and validation sets.

## A. Parameters

There are various parameters that are adjusted to enhance the performance and accuracy of the models. These parameters are fine-tuned through multiple experiments until the optimal balance is achieved between the actual values and predictions in the train and validation sets. The parameters for each model are determined based on the test set as outlined below:

### 1) Batch Size:

For all models, the default batch size of 32 is used. This decision is based on the observation that ResNet18 demonstrated higher precision and accuracy when trained with this batch size.

### 2) Loss function:

The Cross-Entropy-Loss function is utilized for all models. This loss function is commonly used for classification tasks. By using Cross-Entropy-Loss the difference between discovered probabilities distribution of our model and the predictive values is computed. By adjusting the weights of our model Cross-Entropy-Loss tries to find the optimal solution during training. The lower the loss error, the better the model performs. The mathematical notation of Cross-Entropy-Loss is:

$$CE(y, \hat{y}) = - \sum_{i=1}^K y_i \cdot \log(\hat{y}_i)$$

As shown above,  $y_i$  represents ground truth labels and  $\hat{y}_i$  represents predicted probabilities for each class. For each class, the loss is computed based on the logarithm of the predicted probability of that class. And lastly, the losses for all classes are summed up.

### 3) Optimizer:

Our proposed method (pre-trained DenseNet121) model exhibits substantial performance improvement when Stochastic Gradient Descent (SGD) is utilized as the optimizer. Other optimizers such as Adam was tested for all models, but did not outperform SGD. For all models, a fixed learning rate of 0.00001 is used with SGD, however, it did not illustrate a promising result. Several experiments have been conducted with different learning rate and learning rate of 0.001 showed a significant performance improvement.

## B. Metrics used for model evaluation

To evaluate the overall performance of the model, various evaluation metrics can be used, including accuracy, precision, and recall. These metrics can be computed using scikit-learn's standard metrics library, as shown below:

### 1) Accuracy

Accuracy is a proportion of correctly predicted images to the whole number of images.

### 2) Precision

Precision measures the ratio of true positive predictions to all positive predictions.

### 3) Recall/Sensitivity

Recall measures the proportion of true positive predictions to all the actual positives.



#### 4) F-1 Score

F-Score is a harmonic mean of the precision and recall.

### V. RESULTS AND DISCUSSION

All models achieved high accuracy and low loss ratio with a good performance, however, our proposed model which is DenseNet-121, outperformed AlexNet and ResNet-18. Figure 5 below demonstrates the result of our proposed method (pre-trained DenseNet-121), which was trained for 50 epochs. The loss ratio decreased for both the train and validation datasets. Additionally, the model's accuracy showed a remarkable increase.

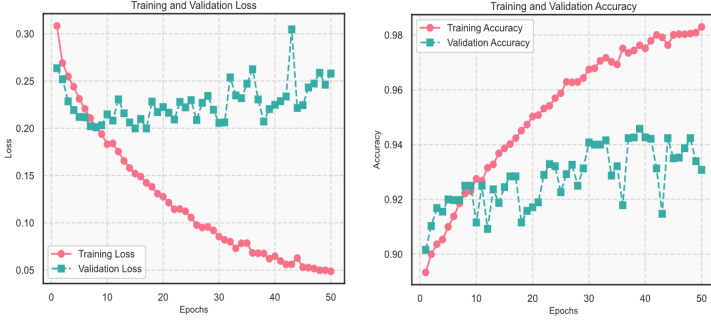


Figure 5 pre-trained DenseNet-121 loss and accuracy

On the contrary, the loss function of DenseNet-121 exhibited a significant decrease for the train set but only a slight decrease for the validation set. As depicted in Figure 5, the accuracy ratio experienced a substantial increase for the train set, whereas it showed a slight improvement for the validation data. Figures 6 depicts the confusion matrix computed on the test data, which comprises 3,776 records. In Figure 6, the DenseNet model accurately predicted 281 benign images and made 111 incorrect predictions. Additionally, it correctly identified 3,248 malignant cases but had 136 false predictions.

#### A. Discussion

Overall, all models achieved a high accuracy score. However, as stated in Section IV, additional metrics are necessary to assess and comprehend the behavior of each model. To further analyze the models, accuracy, precision, recall, and F1 score are computed on the test data for each model. As depicted in Table 1, DenseNet-121 exhibited slightly better performance and obtained higher scores in terms of prediction accuracy.

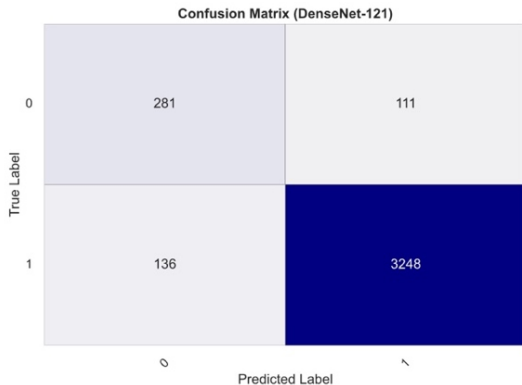


Figure 5 DenseNet-121 Confusion Matrix

Models					
	Accuracy	Precision	Recall	Sensitivity	Specificity
AlexNet	88.53%	87%	89%	96%	27%
ResNet	93.56%	93%	<b>94%</b>	<b>99%</b>	49%
DenseNet	93.46%	<b>94%</b>	93%	96%	72%
Kassem <i>et al.</i>	<b>94.92%</b>	80.36	-----	79.8%	<b>97%</b>

Table 1 Model Scores Metrics

Looking at Table 1, the method proposed by Kassem *et al.* Scored highest accuracy While our proposed method recorded highest precision. Overall, DenseNet-121 have performed better than ResNet-18 and AlexNet as shown in Figure 7.

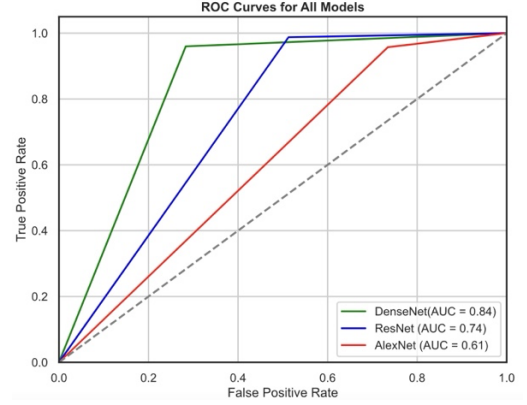


Figure 7 ROC Curve

#### B. Research Limitations

All experiments were carried out using Mac book air which has limited computing resources. Due to these constraints, it was not possible to conduct the research with a complete hyperparameter tuning process and test more complex and deep models. Instead, a trial-and-error approach was employed, testing random parameters on the model using the validation set until an optimal configuration was identified. However, there is potential for further enhancing the model's performance by exploring a wider range of hyperparameters.

### VI. CONCLUSION

In conclusion, this research paper presents a comparative performance analysis of benign and malignant classification using AlexNet, ResNet-18 and DenseNet-121 as three deep learning models. Additionally, the performance metrics of each model in classifying ISIC-2019 images are evaluated. While all models demonstrate remarkable performance and accuracy, DenseNet-121 outperforms AlexNet and ResNet-18 in all metrics. Despite some fluctuations in the loss function of the test set, the model reaches an accuracy of 93.46% in 50 epochs. Image augmentation is also recommended to mitigate overfitting issues and enhance the model's ability to make generalized predictions. Consequently, DenseNet-121 proves to be an ideal model for classifying benign and malignant images. To further improve the model's performance, future considerations may involve increasing the sample size and variations, conducting full-scale hyperparameter tuning, and implementing additional CNN techniques and deeper architectures.

## VII.

Links to the code for all models are provided below:

- 1- [AlexNet](#)
- 2- [ResNet-18](#)
- 3- [DenseNet-121](#)
- 4- [Model-Comparison-ROC](#)

## VIII. REFERENCES

- [1] S. A. Mehwish Dildar, "Skin Cancer Detection: A Review Using Deep Learning Techniques," International Journal of Environmental Research and Public Health, 2021.
- [2] B. Cassidy, "Analysis of the ISIC image datasets: Usage, benchmarks and recommendations," Medical Image Analysis , vol. 75, 2022.
- [3] K. M. Hosny, "Classification of skin lesions using transfer learning and augmentation with Alex-net," PLOS ONE, 2019.
- [4] M. A. KASSEM, "Skin Lesions Classification Into Eight Classes for ISIC 2019 Using Deep Convolutional Neural Network and Transfer Learning," IEEE Access, vol. 8, 2020.
- [5] P. N. Srinivasu, "Classification of Skin Disease Using Deep Learning Neural Networks with MobileNet V2 and LSTM," Sensors, 2021.
- [6] A. Mahbod, "Transfer learning using a multi-scale and multi-network ensemble for skin lesion classification," Computer Methods and Programs in Biomedicine, vol. 193, 2020.
- [7] K. M. Hosny, "Classification of Skin Lesions into Seven Classes Using Transfer Learning with AlexNet," Journal of Digital Imaging , vol. 33, p. 1325–1334, 2020.
- [8] A. Krizhevsky, "ImageNet Classification with Deep Convolutional Neural Networks," ACM Digital Library, vol. 60, no. 6, 2017.
- [9] Y. LeCun, "GradientBased Learning Applied to Document Recognition," PROC OF THE IEEE, 1998.
- [10] K. He, "Deep Residual Learning for Image Recognition," CVPR, 2015.