Deep Learning Skin Lesion Classification

Stavros Chatzipavlidis
Group: 39
Student Number: u774845

October 2023

1 Introduction

1.1 Sample visualization

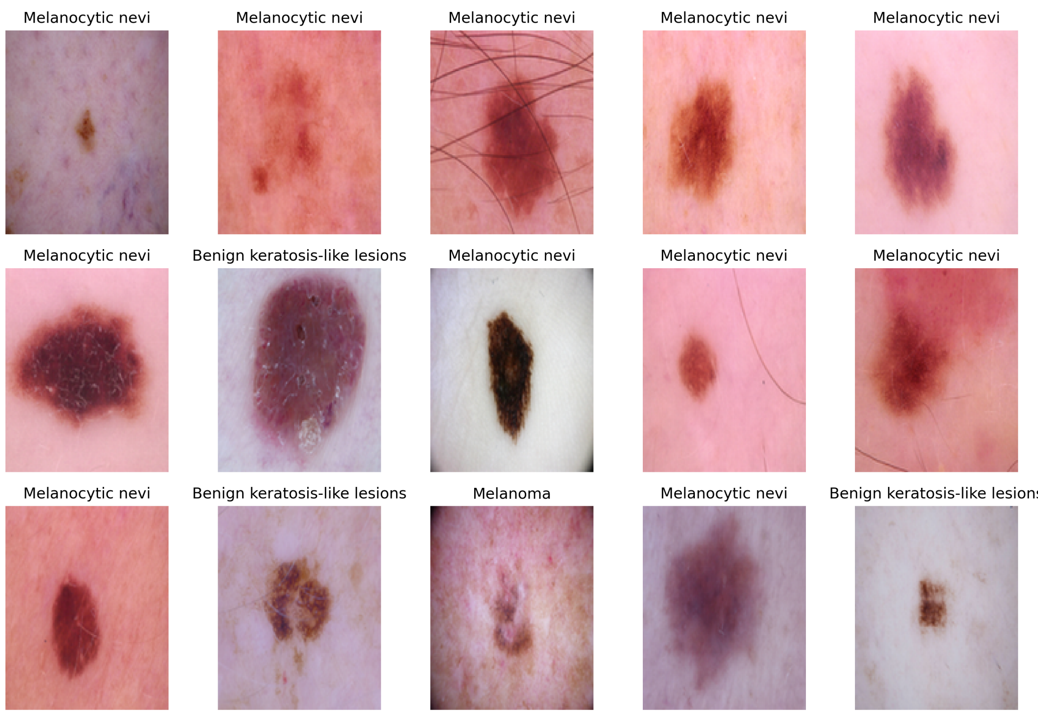


Figure 1: Random Sample Visualization

1.2 Class Distribution

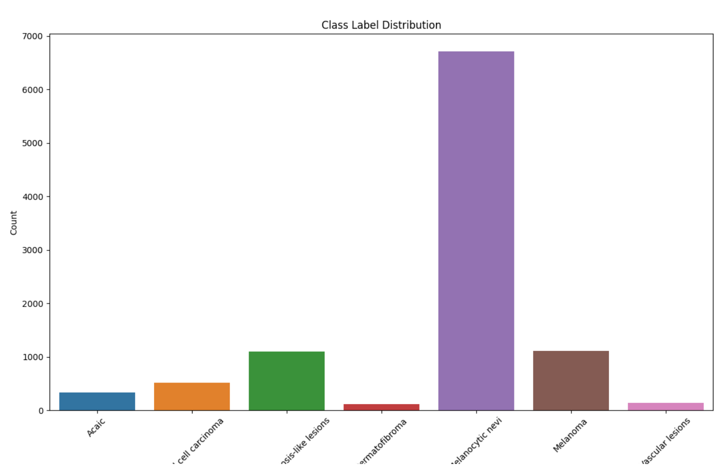


Figure 2: Class Distribution

2 Baseline

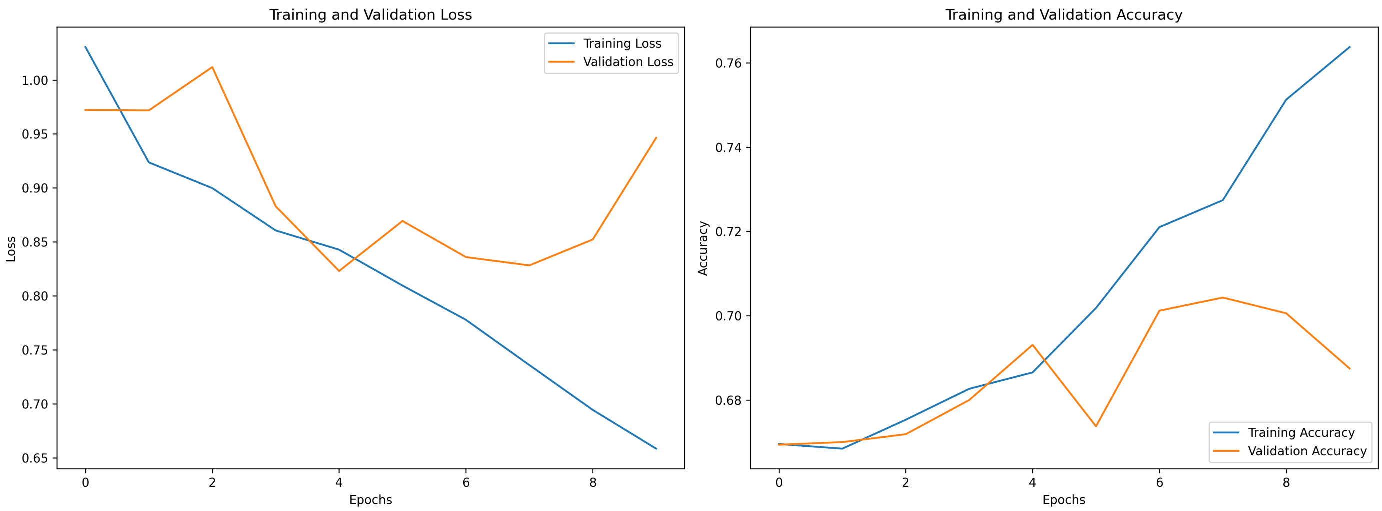
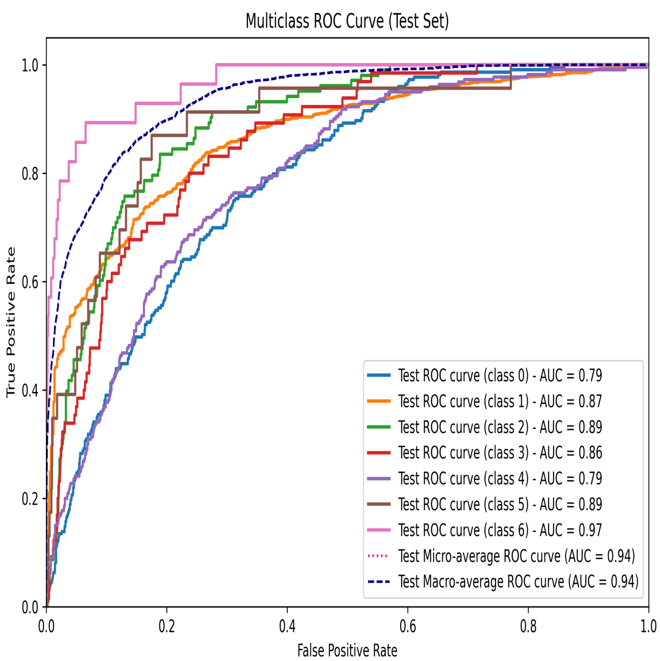
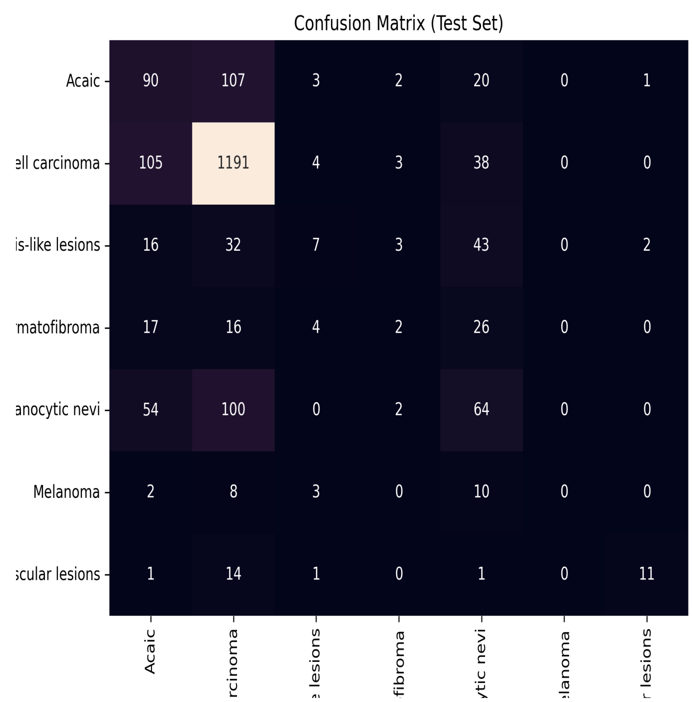


Figure 3: Model Loss and Accuracy



(a) Multi Roc-Auc Curve



(b) Confusion Matrix

Figure 4: Combined Figures

The model architecture comprises two convolutional blocks. In the first block, there are 64 filters in the initial Conv2D layer and 32 filters in the subsequent Conv2D layer, both with a (3, 3) kernel size and ReLU activation. A MaxPooling2D layer with a (2, 2) pool size follows each Conv2D layer. The second convolutional block mirrors the structure of the first, with a Conv2D layer having 64 filters followed by another Conv2D layer with 32 filters, both with ReLU activation. Each Conv2D layer is succeeded by a MaxPooling2D layer with a (2, 2) pool size. The flattened output is then connected to two dense layers, each with 32 units and ReLU activation. The output layer, with softmax activation, consists of 7 units, representing the number of classes.In terms of observable behavior, the growing training loss and decreasing validation loss signal that the baseline model is overfitting. The improvement in training accuracy supports this, showing that the model is responding too closely to the training data. However, the decline in validation accuracy indicates a lack of generalization to previously unknown data. Regularization techniques like as dropout, L2 regularization, or data augmentation could be used to reduce overfitting, and hyperparameter modification may be required to establish a better balance between model complexity and generalization.

3 Enhanced Model

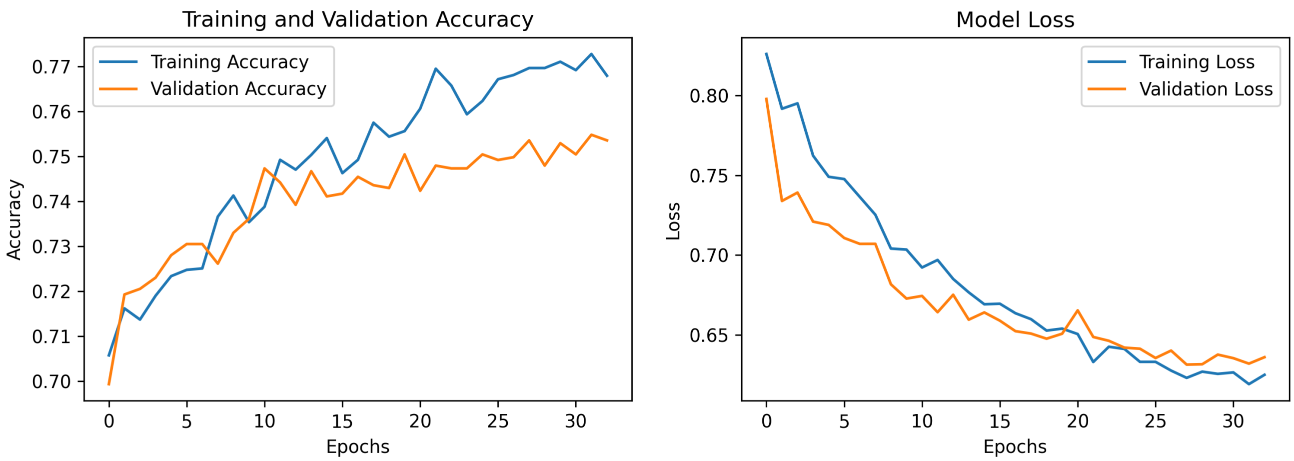


Figure 5: Model Loss and Accuracy

The improved model outperforms the baseline by achieving better accuracy in both validation and test sets. The identical trend observed between training and validation loss, as well as training and validation accuracy, indicates effective generalization and mitigated overfitting. Dropout layers introduce a random aspect during training, preventing dependency on specific neurons and therefore improving model robustness. L2 regularization reduces overfitting by punishing large weights, promoting a more generalized model. The personalized learning rate schedule ensures dynamic modifications during training, optimizing convergence while avoiding local minima. Finally, data augmentation adds variety to the training set, improving the model's capacity to handle a wide range of real-world circumstances.

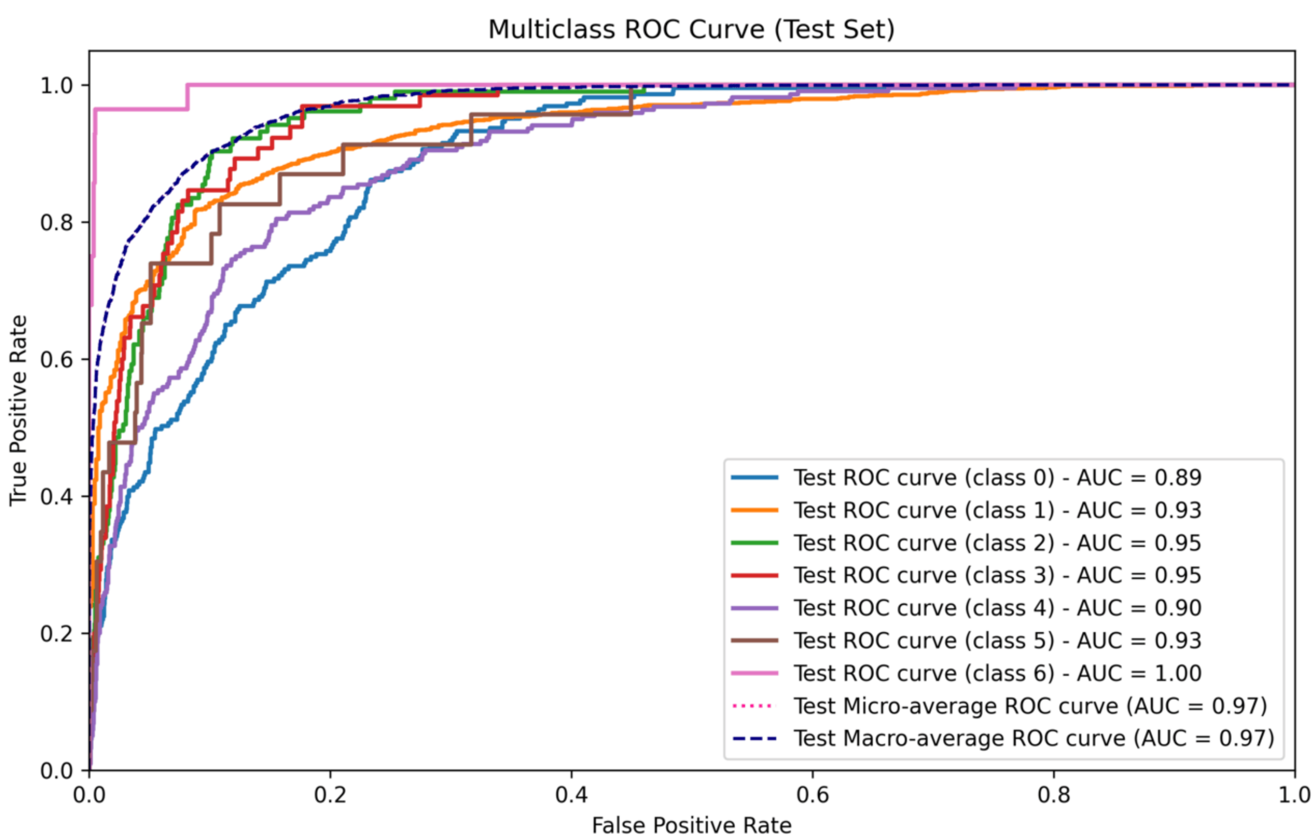


Figure 6: Multiclass Receiver Operating Characteristic (ROC)

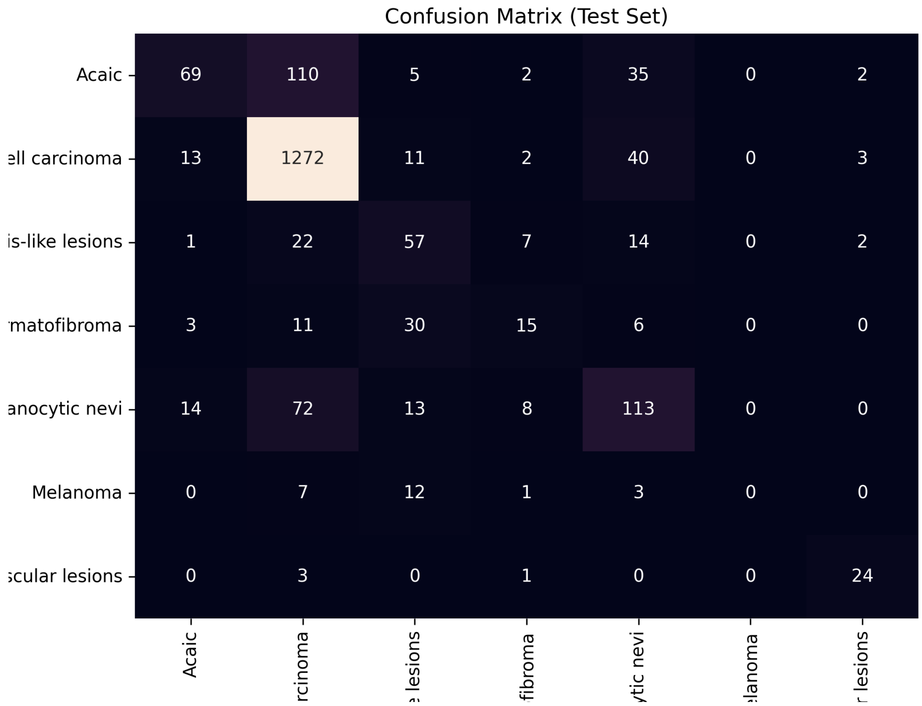


Figure 7: Enchanced Model Confusion Matrix

4 Transfer Learning

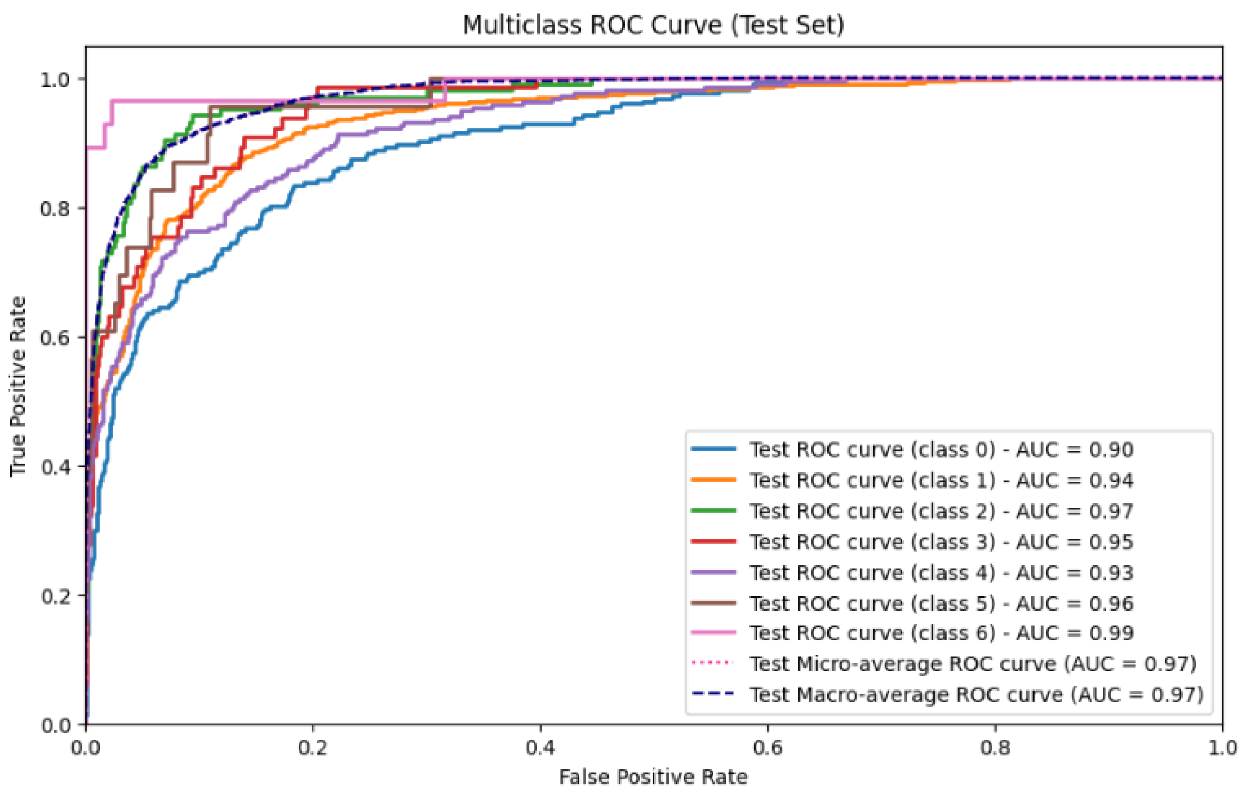


Figure 8: Multiclass Receiver Operating Characteristic (ROC)

The Xception architecture is used as a base. The model is trained for 40 epochs. The custom model architecture includes a Global Average Pooling layer followed by two Dense layers with 256 and 128 filters accordingly with ReLU activation and dropout of 0.5 for regularization. The output layer has 7 units with softmax activation for multi-class classification. The training process uses the Adam optimizer with a learning rate of 0.001 and a learning rate scheduler that decreases the learning rate by 10% every epoch. Early stopping is implemented to monitor the validation loss and restore the best weights when patience equal to 5 runs out.

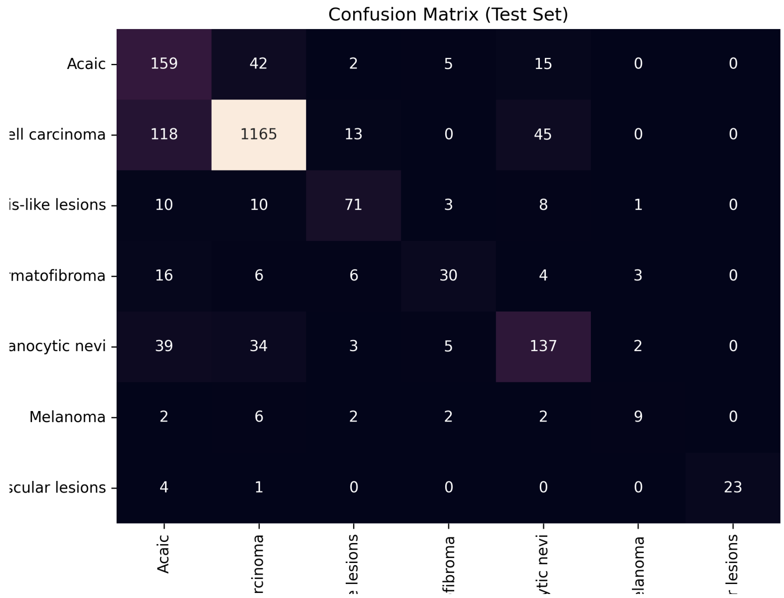


Figure 9: Enchanced Model Confusion Matrix

5 Conclusion

Metric	Baseline (Test)	Enhanced (Test)	Xception (Test)
Total Accuracy	0.6815	0.7738	0.7958
Total Precision	0.39	0.51	0.72
Total Recall	0.30	0.49	0.65
Total F1-Score	0.31	0.49	0.67

Table 1: Skin Lesion Classification Metrics (Macro Average)

We thoroughly examined several setups while investigating hyperparameters for image multiclass classification. Initially, the model encountered issues with overfitting, prompting the introduction of dropout, which showed improved results. To further address overfitting, L2 regularization was implemented, yielding even better outcomes. Subsequently, a customized learning rate schedule, decreasing with each epoch, was incorporated into the model. Additionally, data augmentation was applied to optimize performance. These successive steps collectively contributed to mitigating overfitting and enhancing the model's overall effectiveness. In particular variations in the number of convolutional layers, ranging from one to three, with one being the best, were investigated alongside different kernels for each layer, including 32, 64, and 128. Three convolutional blocks were constructed with one layer and convolutional filters were fine-tuned with values of 32, 64, and 128. Systematic testing of kernel sizes, such as $(3, 3)$ and $(4, 4)$, provided insights into their impact on feature extraction. Activation functions, such as Rectified Linear Unit ($ReLU$) and Hyperbolic Tangent ($tanh$), were evaluated for their effects on introducing non-linearity. The downsampling components, represented by pooling layers, were assessed with different sizes like $(2, 2)$ and $(3, 3)$ where $(2, 2)$ was found optimal. Other critical parameters, such as dropout rates ranging from 0.2 to 0.8 and 0.7, batch sizes ranging from 12 to 64 with 32 offering the highest performance, and learning rates, were meticulously changed to assess their impact on model performance. Incorporating early halting mechanisms, with patience levels ranging from 3 to 10 and epochs ranging from 10 to 100, ensured convergence in local minima without entanglement. Dropout rates, batch sizes, and learning rates were evaluated at intervals within the prescribed ranges. A custom learning rate schedule, particularly Adam optimizer with a learning rate of 0.001, coupled with a learning rate scheduler that decreases the learning rate by 10% every epoch, was identified as optimal. Values in proximity, like 0.002×0.9^{epoch}, were investigated in order to fine-tune the learning rate dynamics. L2 regularization strengths ranging from 0.001 to 0.01 were investigated, with 0.002 appearing as the best regularization strength. Furthermore, Stochastic Gradient Descent (SGD) was explored as an alternative to Adam but produced poorer results, validating the Adam optimizer's efficacy in this setting. Finally Data augmentation after experimenting with many values, involved applying diverse transformations to input images, such as rotation (up to 20 degrees), horizontal and vertical shifts (0.1), shear deformation (0.2), zooming (up to 0.2), and horizontal flipping. These adjustments aim to enhance model robustness by exposing it to a wider range of input variations, contributing to improved generalization capabilities. The 'nearest' fill mode is employed to fill empty spaces resulting from augmentations with the nearest available pixel values.The Enhanced model outperforms the Baseline model in terms of accuracy, precision, recall, and F1-score. When comparing the Enhanced and Xception models, the latter is clearly superior. Across all metrics, Xception's advanced architecture, which is specifically designed for image tasks, outperforms the Enhanced model. Xception, in particular, has higher accuracy, precision, recall, and F1-score, demonstrating its ability to capture intricate features critical for accurate dermatological diagnostics. Other techniques that could improve performance include Ensemble Learning Attention Mechanisms, Hyperparameter Search Techniques, Batch Normalization, Advanced Activation Functions, Self-Supervised Learning, Capsule Networks, and AutoML Techniques.