EEEM068: Applied Machine Learning
Project: Melanoma Classification

Group Members (URN):
Taman Bachani (6846172)
Azizbek Fatkhullaev (6850354)
Aptha Sara Mohan (6847445)
Yasar Efe Pekgoz (6855188)

Abstract

Melanoma is one of the most prominent types of cancer in the world. This report introduces our approach to Melanoma classification, with the help of Convolutional Neural Networks (CNNs). Our work includes a meticulous analysis of five different model architectures, and an exploration of various augmentation techniques, optimisers, and hyperparameter tuning. This was performed on a publicly available Melanoma dataset with 10,605 images classified as either 'benign' or 'malignant'. As a result of our experimentations, 'EfficientNetb0' showcased exceptional performance. Hence, the model solution, with optimal tunings is proposed through this project.

Introduction

It is known that skin cancer is widespread and every year, the damage it causes to our population, grows. Melanoma is one of the most common forms of skin malignancies [1]. Melanoma's visual features are one of the biggest differentiators from the other types of cancer. Making a model see and learn those features creates an opportunity to perform an early diagnosis for it. We as a group, acknowledge the importance of precision in predicting cases like these. Hence, we work towards that through our project.

The Image Classification accuracy of state-of-the-art machine learning models has proven successful in many instances. This shows that emerging technologies like machine learning algorithms offer possibilities for improving the classification of Melanoma [2]. By using CNNs, early diagnosis can be achieved for countless cases of skin cancer. We observed that successful Melanoma Classification requires a profound understanding of Image Classification and Cancer Biology.

Our aim in this project is to implement the best technologies for Melanoma Classification cases and fine-tune them to achieve the most optimal results possible. This points us to neural networks, an emerging technology for making a machine intelligent for solving real-life object categorisation problems [3]. Using different Convolutional Neural Network models like ResNet, MobileNet, and EfficientNet, we plan to achieve high accuracy in detecting Melanoma cases.

Literature Review

The literature reviewed underscores the significant role of advanced computational techniques in overall melanoma research. Foundational insights into melanoma biology [1] are critical for the application of targeted therapies and diagnostic tools, that use AI.

Studies on data augmentation [4][5] emphasise the importance of sophisticated image manipulation techniques to train robust models, addressing challenges related to limited data diversity, in problems similar to ours. Additionally, discussions on the effectiveness of binary cross entropy in multi-class classification scenarios [6] are commonly employed to handle imbalanced datasets prevalent in medical imaging studies.

The systematic approach to model optimisation through hyperparameter tuning using grid search [7] is essential for maximising clinical application performance. Collectively, these studies highlight the convergence of technology, biology, and methodological rigor necessary in the fight against melanoma.

Dataset

Brief analysis of the original dataset:

Our original pick for the dataset consisted of 33,126 records, each representing medical image data associated with the metadata of the patient. It had eight columns and notable features including patient demographic information, anatomical site information, and a binary classification of benign or malignant conditions.

The dataset primarily covered dermatological images with a focus on benign conditions across a diverse patient demographic. The significant class imbalance between benign and malignant cases was a challenge for predictive modelling (Fig. 1.1). A challenge we attempted to address. We sought guidance and tried a plethora of methods from oversampling the malignant data, to applying a weighted loss, tremendous data augmentation, and even the generation of synthetic data (SMOTE). We realised to make the model deliver good accuracy, we needed to train it for longer. The results we achieved with 10 epochs were underwhelming. Hence, the lack of resources and results pushed us to opt for a different dataset so we could focus on the key task of Melanoma classification.

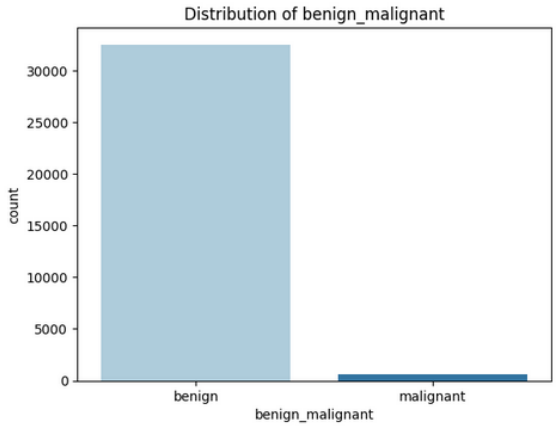

Fig. 1.1: Class Distribution (old)

Our **new melanoma dataset** comprised of 10,605 entries. There were 5,500 cases classified as non-melanoma. This represented around 51.87% of the dataset. There were 5,105 cases identified as melanoma accounting for about 48.13% of the dataset.

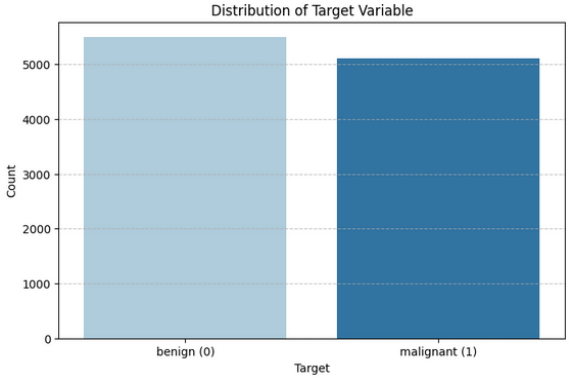

Fig 1.2: Class Distribution (new)

Overall, the dataset had a nearly balanced distribution between benign(0) and malignant(1) which was more suitable for conducting our experiments. The dataset was also entirely non-null with no missing values. He, itnce was more reliable for our analysis. This helped us centre our focus towards the underlying computer vision task of classification.

Methodology

Data Pre-Processing and Augmentation

Data pre-processing and augmentation are one of the industry standards to improve machine learning performance. It is also a way to extend training data to allow models to see more data while training and to achieve good results in situations where the data is limited [4].

Our Kaggle dataset was a balanced dataset populated with 10,605 images. These images were split into an 80-10-10 split for training, cross-validation, and testing. The data was loaded with the given methods in the PyTorch library.

For data augmentations, we applied a combination of augmentations (Fig. 2.1). We resized the images to 224x224, converted them into a tensor and normalised them according to the standard deviation and mean values derived from ImageNet.

Data augmentation allows the model to see more data, and hence learn more, and improve its accuracy [5]. Considering that we applied multiple data augmentations to our model to improve our results. We applied random horizontal flip, random rotations of 15°, and color jitter with the values of 0.2 brightness, contrast, saturation and 0.1 hues. And at last, random resized crop. This allows the model to see more data and therefore generalise well to unseen data. A detailed example of an image from our dataset can be seen below (Fig. 2.2).

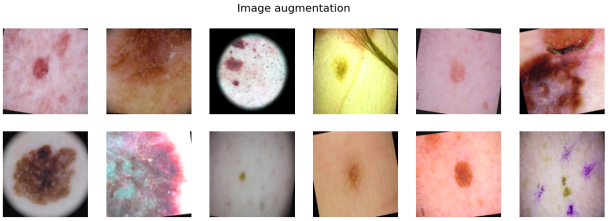

Fig. 2.1: Augmented Views of an Image

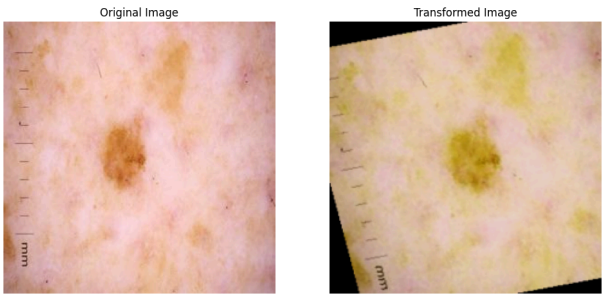

Fig. 2.2: Combined Augmentation over Image

Models

Upon extensive research for our project, we established that the model solution should be based on the convolutional neural networks (CNNs) architecture. We experimented with several models to identify the most effective approach for accurately classifying skin lesions. Let's discuss these in detail.

ResNet34 and ResNext101 are part of the Residual Network family that use skip connections to allow deeper networks by addressing the vanishing gradient problem. ResNet34, with a relatively simpler architecture, showed an accuracy of 92.46% and an f1 score of 0.9243, on the test set with around 1100 images. ResNext101, which introduces grouped convolutions for increased model complexity and capacity, slightly improved upon this with an accuracy of 92.55% and an f1 score of 0.9253.

MobileNetV2 is designed for mobile and edge devices, focusing on efficiency and speed. It uses depthwise separable convolutions, which reduces the number of parameters and saves on computational cost, resulting in an accuracy of 90.94% and an f1 score of 0.9093 in our project.

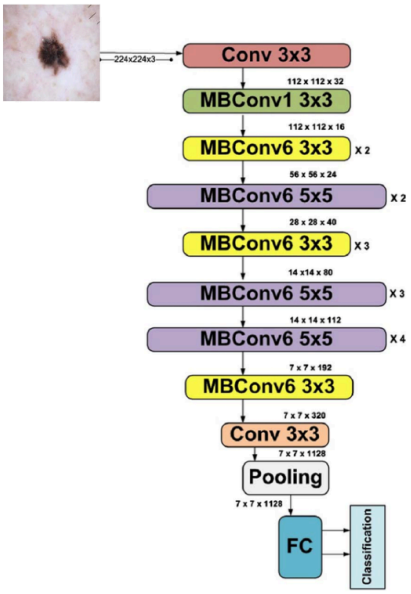

Fig. 3: EfficientNetb0 Architecture

Finally, EfficientNetB0, which scales network width, depth, and resolution through a compound scaling method, the rate of scaling depends on the model chosen from the ensemble (B0 to B7). Our model outperformed all the other models in our tests. It achieved the highest accuracy of 96.80% and an f1 score of 0.9679. EfficientNet's architecture leverages a balance between different dimensions of the network, leading to superior performance without excessive computational demands.

This decision to settle with EfficientNet was driven by the need to maintain a balance between computational efficiency and predictive performance (Table 1), making it an optimal choice for our academic project setting.

Model	Accuracy	f1 score
MobileNetV2	90.94%	0.9093
ResNext101	92.55%	0.9253
ResNet34	92.46%	0.9243
EfficientNetb0	**96.80%**	**0.9679**

Table 1: Model Performance Comparison

Experiments

Loss Function and Optimisers

For our binary melanoma classification task we utilized Binary Cross-Entropy Loss (BCELoss). Which works extremely well with classification problems. It relies on the calculation of distances between actual and predicted values during training. Thus providing a strong metric for the model to learn from.

We explored two optimizers: Adam and Stochastic Gradient Descent (SGD). Adam, with its adaptive learning rate capabilities, combines the advantages of AdaGrad and RMSProp, making it suitable for large-scale problems. SGD, while simpler, requires careful learning rate tuning and scheduling.

In our experiments, Adam proved more effective, leading to faster convergence and requiring less manual tuning compared to SGD. Its adaptive learning rate mechanism facilitated quicker model accuracy improvements.

The combination of BCELoss and Adam optimiser enhanced our models' predictive performance,

ensuring effective learning from training data and generalization to unseen data.

Hyperparameter Optimisation

We used a structured approach to hyperparameter optimisation to enhance the performance of our model. Acknowledging the pivotal influence of learning rate and batch size on the training dynamics of convolutional neural networks, we conducted a grid search [7] to identify the most optimal settings for these parameters.

The optimisation process involved experimenting with different learning rates and batch sizes to identify their impact on the model's F1 score (Table 2). The experimentation revealed distinct patterns: a lower learning rate of 0.001 gradually improved the F1-score as batch size increased, reaching optimal performance with a batch size of 64. This configuration suggested that a slower, more deliberate learning process combined with larger batches provided the most stable and effective learning environment.

An intermediate learning rate of 0.01 showed strong performance with a 16-batch size. However, when the batch size was increased to 32, the performance dipped slightly. This could be attributed to the learning rate being too high for the larger batch size, resulting in noisy updates and impeding convergence. Moreover, as the batch size was further increased to 64, the performance rebounded, indicating that the larger batch size helped in stabilising the training process and mitigating the adverse effects of the higher learning rate. Higher learning rates of 0.1 generally led to inferior performance across all batch sizes, suggesting that such rates might be too aggressive, hindering effective convergence.

From these findings, the optimal hyperparameter combination for our model was a learning rate of 0.001 and a batch size of 64. (Table 2) The insights gained from this hyperparameter tuning process were instrumental in fine-tuning our models to achieve exceptional accuracy and reliability.

F1 score for different parameters on CV set

LR \ BS	16	32	64
0.001	0.9241	0.9252	**0.9443**
0.01	0.8763	0.8649	0.8976
0.1	0.7977	0.8725	0.8763

Table 2: Grid Search for Hyperparameter Tuning

Observations and Outcome

Acting upon our observations, the project's outcome underscored the efficacy of the EfficientNetB0 model in melanoma classification, demonstrating superior performance metrics as compared to other tested architectures. We successfully achieved a remarkable accuracy and F1 score due to the model's balanced approach to scaling the depth, width, and resolution of the network.

The loss per epoch graph (Fig. 4.1) exhibits a descending trend in both training and validation loss, indicating effective learning and generalization over epochs. While we only ran the model over 10 epochs, we are certain of its capability to deliver more performance over more epochs (provided the resources). The initial sharp decline followed by a gradual decrease suggests that major learning occurs in the early epochs, with fine-tuning in subsequent passes. This trajectory reflects well-implemented learning rate adjustments and an adequate balance of bias-variance, ensuring the model learns detailed features without overfitting, hence we see the jagged nature of the cross-validation loss curve.

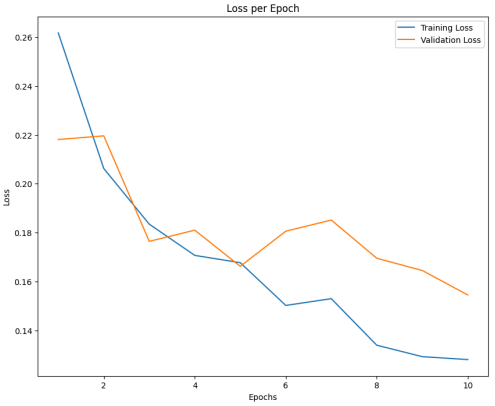

Fig. 4.1: Loss over 10 epochs for EfficientNetb0

The confusion matrix (Fig. 4.2) further highlights the model's robustness, displaying a high number of true positives and true negatives, while minimising the false positives and false negatives. This achievement emphasizes the model's precision in

classifying both classes accurately is an essential factor in medical diagnostics where the cost of errors can be high.

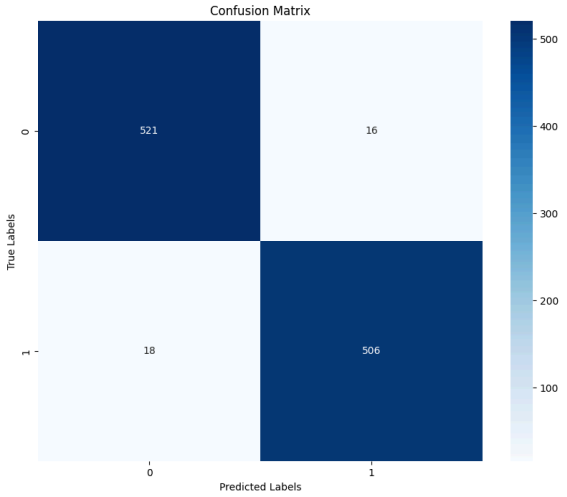

Fig. 4.2: Confusion Matrix for our best model (tuned)

Visualization of the model's internal workings through activation maps (Fig. 4.3) offers significant insight into how the CNN processes melanoma images (The visualisations are colour-corrected to showcase contrast). From the original image to successive layers, the model focuses increasingly on vital features relevant to melanoma classification, such as asymmetry, border irregularities in skin lesions, and the detection of hair on the skin. The activation intensifies around these key areas as the layers progress, demonstrating the network's ability to abstract and emphasize features critical for accurate classification.

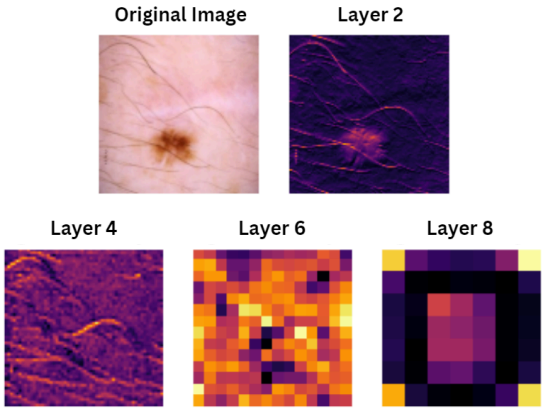

Fig. 4.3: Activation Maps of our model (visualised)

This analysis not only confirms the model's effectiveness in identifying and classifying melanoma but also showcases its potential to provide explainable AI outputs, vital for further research into AI-driven diagnosis.

Conclusion and Future Work

This study presents a comprehensive exploration into melanoma classification using CNNs, highlighting the superior performance of EfficientNetB0 in achieving unparalleled accuracy. Through meticulous experimentation with various models, augmentations, and hyperparameters, we have paved the way for advanced, real-time diagnostic tools in dermatology.

Looking ahead, our focus will shift towards integrating multimodal data sources, including clinical records and genetic profiles, to enhance the model's diagnostic precision. Exploring unsupervised learning techniques to identify unseen patterns in melanoma and employing explainable AI methods will further demystify the decision-making process, providing deeper insights for healthcare professionals.

References

[1] V. Gray-Schopfer, C. Wellbrock, and R. Marais, "Melanoma biology and new targeted therapy," *Nature*, vol. 445, no. 7130, pp. 851–857, Feb. 2007, doi: https://doi.org/10.1038/nature05661.

[2] P. R. Esfahani *et al.*, "Leveraging Machine Learning for Accurate Detection and Diagnosis of Melanoma and Nevi: An Interdisciplinary Study in Dermatology," *Cureus*, vol. 15, no. 8, Aug. 2023, doi: https://doi.org/10.7759/cureus.44120.

[3] N. Sharma, V. Jain, and A. Mishra, "An Analysis Of Convolutional Neural Networks For Image Classification," *Procedia Computer Science*, vol. 132, pp. 377–384, 2018, doi: https://doi.org/10.1016/j.procs.2018.05.198.

[4] A. Mumuni and F. Mumuni, "Data augmentation: A comprehensive survey of modern approaches," *Array*, vol. 16, p. 100258, Nov. 2022, doi: https://doi.org/10.1016/j.array.2022.100258.

[5] L. Perez and J. Wang, "The Effectiveness of Data Augmentation in Image Classification using Deep Learning," *arXiv:1712.04621 [cs]*, Dec. 2017, Available: https://arxiv.org/abs/1712.04621

[6] U. Ruby and V. Yendapalli, 'Binary cross entropy with deep learning technique for Image classification', *International Journal of Advanced Trends in Computer Science and Engineering,* vol. 9, 10 2020.

[7] B. H. Shekar and G. Dagnew, "Grid Search-Based Hyperparameter Tuning and Classification of Microarray Cancer Data," 2019 Second International Conference on Advanced Computational and Communication Paradigms (ICACCP), Gangtok, India, 2019, pp. 1-8, doi: 10.1109/ICACCP.2019.8882943.