Automated Detection of Malignant and Benign Skin Lesions Using Deep Learning

Damir Tassybayev, Zhansaya Orazbay, Dias Nursultan, Artur Kadyrzhanov
Supervised by: Prof. Daniele Pannone
University of Rome "La Sapienza"

I. Introduction

It is a prominent fact that skin cancer is one of the most fatal forms of cancer, which has been widely spread comprehensively in the world. According to statistics from the World Health Organisation, each year three million skin cancer diagnoses have a 2.37% mortality rate [1]. Two key sources of disease are the sun's damaging ultraviolet rays and the disrepair of DNA damage that causes mutations. Mutations become the head of the process that leads skin cells to rapid growth and initiate malignant or benign tumours. [2] Assuming all these factors, premature detection and timely intervention are crucial for efficient treatment and revamping patient outcomes.

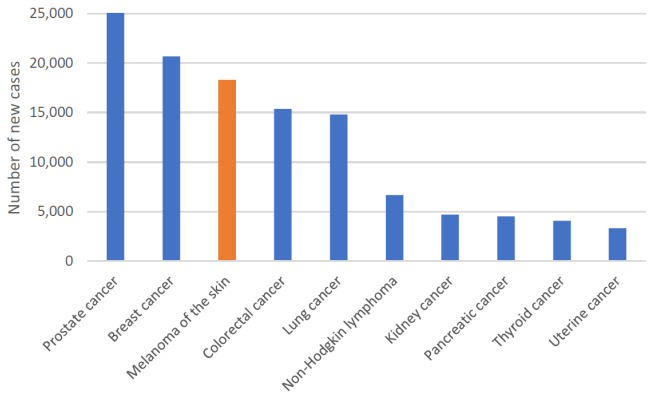

Fig. 1. Estimated Cancer Incidence in Australia, 2023. Source: Cancer Australia [3].

II. Related Work

Early research on the automated diagnosis of skin cancer began in the 1990s, with one of the earliest works being conducted by Christopher A. Borden and Scott P. Whelan. Their study, published in the 1994 IEEE paper titled *Automated Dermatology Diagnosis*, utilized computer vision techniques to analyze skin sample images for cancer detection. This research was foundational in exploring the potential of automating dermatological diagnostics, setting the stage for subsequent advancements in the field [4].

Over the following decades, the field has witnessed significant advancements across multiple disciplines, incorporating various methodologies to enhance diagnostic accuracy. A pivotal change was introduced by Abbas et al. in their 2015 paper *Automated Skin Lesion Analysis System: A Review of Classifier Selection and Performance Estimation Techniques*, which provided a comprehensive review and made significant contributions using deep learning approaches. This work emphasized the importance of selecting appropriate classifiers and robust performance evaluation metrics to improve the reliability of automated skin lesion analysis systems, marking a significant advancement in the application of deep learning in dermatology [5].

Furthermore, recent work by Alenezi, Armghan, and Polat (2023) has advanced melanoma recognition through the development of a multi-stage framework that integrates deep residual neural networks and hyperparameter optimization. Their study, published in *Expert Systems with Applications*, details a comprehensive approach to decision support in dermoscopy images, emphasizing the ongoing evolution and integration of advanced AI techniques in dermatology. This multi-stage framework substantially improves the accuracy and reliability of melanoma detection, offering a robust tool for clinical application [6].

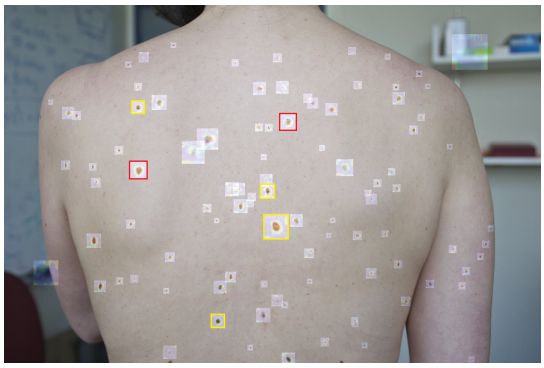

Fig. 2. Deep Learning Detection of Skin Lesions. Source: MIT News [7].

III. Goal

The primary objective of this project is to develop a robust high-accuracy classifier of skin cancer lesions into malignant or benign categories. It aims to enhance diagnostic capabilities, thereby supporting medical professionals in the early detection and intervention of malignant skin lesions.

IV. Expected Outcomes

- Improved efficiency of diagnostic precision for distinguishing between malignant and benign skin lesions, enhancing the accuracy of clinical assessments.
- Increased generalizability and robustness of the model under various imaging conditions, ensuring reliable performance in diverse clinical environments.

V. Preprocessing

A. About Dataset

The dataset was retrieved from Kaggle [8], originally taken from the ISIC dataset [9] and consists of 3263 images of skin cancer lesions with a fixed dimension of 224x224 pixels. Each image is categorized into its corresponding class by the following titles: benign and malignant. The dataset has been divided into train and test sets in an acceptable proportion. The training and testing set contains 2609 and 654 items, respectively. One of the main operations to be Color Space Transformation and Contrast Improvement.

B. Function of preprocessing

The `pre_process` function is a critical component of the data preprocessing in the project pipeline. Initially, the function reads an image and converts it from BGR to RGB format, accommodating the standard color format used in most image processing libraries. Subsequently, the image undergoes a transformation to the LAB color space, where L stands for Lightness. The Lightness channel (L) is then enhanced using Contrast Limited Adaptive Histogram Equalization (CLAHE), which significantly improves the contrast, making the features of skin lesions more distinct. This enhanced channel is merged back with the original A and B channels before being converted back to RGB. Finally, the image is resized to 224x224 pixels to match the input requirements of the deep learning model. This sequence of transformations ensures that the images are optimally prepared for effective feature extraction in the classification process.

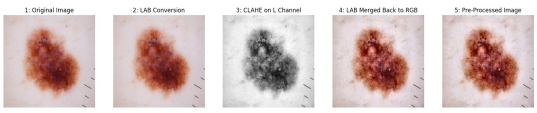

Fig. 3. Pipeline of Preprocessing

C. Data Augmentation

To enhance the robustness of the dataset, we employed an utility called ImageDataGenerator of Keras library for data augmentation during training phase. This process involved applying various transformations to increase the diversity of the training set, thereby improving model generalizability. Specific augmentation techniques were utilized included rescaling, shearing, zooming, rotating, horizontal flipping, and adjusting the width and height shifts, along with modifications in fill mode. These adjustments help ensure that the proposed model remains effective under various imaging conditions. The following list specifies the parameters used for data augmentation:

- Rescale: 1./255
- Shear range: 0.2
- Zoom range: 0.2
- Horizontal flip: True
- Rotation range: 40 degrees
- Width shift range: 0.2
- Height shift range: 0.2
- Brightness range: from 0.8 to 1.2
- Fill mode: ”reflect”
- Validation split: 20%

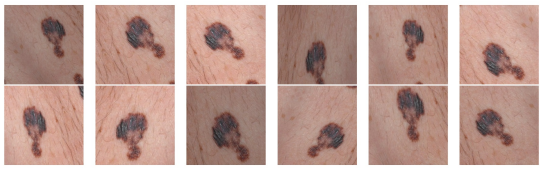

Fig. 4. Example of Data Augmentation process

VI. Model Architecture

The proposed model's structure is built to accurately identify skin cancer type using a combination of Convolutional Neural Network (CNN) and Multi-Layer Perceptron (MLP). There are three fundamental elements in the system: feature extraction, pattern recognition, and classification. The combination of the CNN and MLP components allows the model to leverage both the feature extraction capabilities of convolutional layers and the pattern recognition capabilities of fully connected layers

Fig. 5. Model Architecture Visualization

A. Feature Extraction

Base Model: VGG16 is chosen as the base model because of its effectiveness in extracting features from images. It captures hierarchical features, which is crucial for recognizing patterns in skin cancer images.

B. Pattern Recognition

1) Dense Layer: Two dense layers with 512 neurons each are added after flattening the feature maps and ReLU (Rectified Linear Unit) activation functions are used after each dense layer. These enable the model to capture and represent complex patterns in the data for accurate classification.

2) Batch Normalization: These layers are added to ensure that the outputs are normalized before being passed to the next layer. This helps in stabilizing the training process and improving the overall performance of the model by allowing for higher learning rates and reducing the dependence on weight initialization.

3) Dropout Regularization: Dropout layers with a dropout rate of 0.5 are added after each dense layer. Therefore, during training each neuron in the dense layer has a 50% probability of being deactivated, which prevents overfitting.

C. Classification

Final Classification Layer: The final dense layer with a single neuron and sigmoid activation function serves as the output layer of the model. It performs binary classification, distinguishing between benign and malignant skin lesions.

VII. Training Process

A. On Dataset

The dataset was already split into training and testing datasets in advance. The image paths were initially gathered and shuffled randomly to ensure that the data is mixed. After preprocessing of images, as described above, the original and preprocessed images were stored in separate lists. The labels—malignant and benign—were also converted into numerical values at this stage since the model cannot process data strings. The 'malignant' value is assigned as 1, while 'benign' is 0.

B. Model Compilation

The Adam optimizer with a learning rate of 0.001 is used for updating the network weights during training. A learning rate of 0.001 controls the step size in the gradient descent optimization process. For each prediction, the binary cross-entropy loss computes a value based on the predicted probability of the positive class and the actual label. It aims to minimize this loss during training.

C. Callbacks

The "EarlyStopping" callback is used, so the training will stop if the validation loss doesn't improve for 7 consecutive epochs, and the model weights will be restored to those of the epoch with the best validation loss. The "ReduceLROnPlateau" callback is also implemented to reduce the learning rate by a factor of 0.1 if the validation loss doesn't improve for 2 consecutive epochs. In the end, the best model weights are saved to restore the best-performing model whenever the validation loss improves.

D. Fit Function

The `model.fit()` function trains the model for 100 epochs on the training data and validates it on the validation data.

VIII. Results Visualization

A. Encountered Problems

As mentioned before, we started the process of training with 100 epochs. During this period, we faced typical troubles such as:

- **Overfitting**: The model showed signs of overfitting as validation accuracy stabilized or decreased while training accuracy continued to increase. This indicates potential retention of the training data.
- **Validation Loss**: Fluctuations in validation loss were perceived, indicating inconsistent implementation of the validation dataset, which could be due to the complexity of the dataset or anomalous regularization.
- **Training Time**: The extensive training process of the model, combined with the utilization of multiple epochs and intricate data augmentation strategies, led to prolonged training durations.

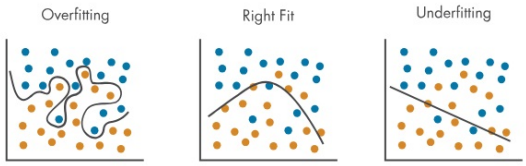

Fig. 6. Model behaviour classification. Source: MATLAB [10]

B. Outcome

Despite these challenges, during the project these results were realized:

- **Model Execution**: During different epochs, the model achieved training accuracy from approximately 73% to 85%. Validation results ranged between 63% and 86%. The final unseen test achieved an accuracy of 74%. The corresponding 99% confidence interval for the result is between 70% and 78%
- **Accuracy Trends**: Both training and validation loss decreased over epochs, indicating improvement in the model's performance. Training and validation accuracy generally increased over epochs, suggesting that the model learned from the data.

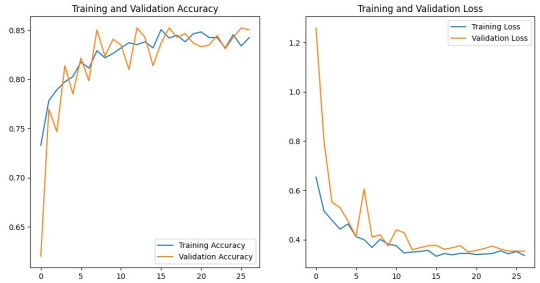

Fig. 7. Accuracy and Loss trends

IX. Practical Assessment

The presented results showcase the model's capabilities in classifying skin lesions as benign or malignant with considerable accuracy, as demonstrated in the diagnostic images. This visual representation serves as a testament to the model's effectiveness in clinical settings.

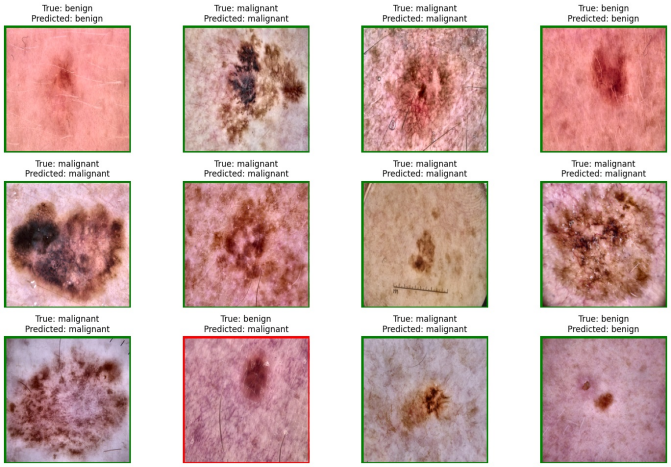

Fig. 8. Visualisation of results on random test images set

X. Epilogue and Future Work

Overall, the project highlights the potential of Machine Learning techniques and the approach of Artificial Intelligence to improve skin cancer lesion recognition/classification and emphasizes the need for ongoing research. By using advanced methods and expanding datasets, we can enhance skin cancer detection models, benefiting both patients and healthcare professionals. Future research can focus on analyzing and accomplishing a diversified data collection of substantial evidence regarding skin cancer to make the proposed model more stable and validated. Additionally, exploring other CNN models such as ResNet, DenseNet, and InceptionNet could further improve the performance of the model.

XI. References

References

[1] World Health Organization. (2024). *World health statistics 2024: monitoring health for the SDGs, sustainable development goals*. Retrieved from https://www.who.int/news-room/fact-sheets/detail/cancer

[2] A. S. Esin, M. E. Volkov, and A. V. Kubyshkin, "Advances in the Understanding of Skin Cancer: Ultraviolet Radiation, Mutations, and Antisense Oligonucleotides as Anticancer Drugs," *Molecules*, vol. 24, no. 8, pp. 1516, 2019. Available: https://doi.org/10.3390/molecules24081516

[3] Cancer Australia. (2023). Cancer in Australia statistics. Retrieved from https://www.canceraustralia.gov.au/cancer-types/lung-cancer/statistics

[4] C. A. Borden and S. P. Whelan, "Automated dermatology diagnosis," in *IEEE Proceedings*, vol. 82, no. 12, pp. 1947-1964, 1994.

[5] G. Abbas, M. E. Celebi, and I. F. Garcia, "Automated skin lesion analysis system: A review of classifier selection and performance estimation techniques," *Journal of Dermatological Science*, vol. 78, no. 2, pp. 92-106, May 2015.

[6] F. Alenezi, A. Armghan, and K. Polat, "A multi-stage melanoma recognition framework with deep residual neural networks and hyperparameter optimization," *Expert Systems with Applications*, vol. 160, Article 113790, 2023.

[7] MIT News. (2021). Artificial intelligence tool can help detect melanoma. Retrieved from https://news.mit.edu/2021/artificial-intelligence-tool-can-help-detect-melanoma-0402

[8] Kaggle. (2019). *Skin Cancer: Malignant vs. Benign*. Retrieved from https://www.kaggle.com/datasets/fanconic/skin-cancer-malignant-vs-benign

[9] ISIC. (n.d.). *ISIC 2016-2020*. Retrieved from https://www.isic-archive.com/

[10] MATLAB. (n.d.). Overfitting. Retrieved from https://www.mathworks.com/discovery/overfitting.html