1 Introduction, Objectives, Motivation

Skin cancer can be life-threatening if not detected early, yet many communities lack access to specialized dermatologic care. To bridge this gap, algorithms designed for use in primary care or non-clinical settings must excel at analyzing lower-quality images [1].

This project tries to address those problems and goals mentioned above. In our work, we aim to create an algorithm that differentiates histologically confirmed malignant skin lesions from benign lesions on a patient. The quality of the images used for the algorithm resembles close-up smartphone photos, which are regularly submitted for telehealth purposes. Installing an algorithm like this on their cell phones could help people directly who do not have easy access to medical professionals. It could also reduce the cost of healthcare, helping those who cannot afford expensive examinations.

2 Previous Solutions, Their Advantages, Disadvantages

One solution was presented in *Nature* in 2017 [4]. The researchers used a deep CNN (Google's Inception v3) and 129,450 clinical images of varying quality. Here are the advantages and disadvantages of the approach:

2.1 Advantages

- The algorithm achieved diagnostic performance comparable to 21 board-certified dermatologists.
- The model demonstrates robustness against variations in photographic quality, including differences in zoom, angle, and lighting, which are common in real-world images. It also does not require image preprocessing.

2.2 Disadvantages

- The algorithm does not take patient history into account.

- Heavily reliant on the available training data (true for all deep learning algorithms), therefore regions with limited data or diverse skin types might see reduced performance.

Besides the algorithms mentioned above, there are systematic reviews about deep CNNs used for skin cancer detection [2, 3].

3 System Design

We use transfer learning for our project. Initially, we import the image data and the metadata and match every image with its corresponding label. After that, we balance the dataset using techniques like data augmentation and oversampling and convert it into a format so the CNN can work with it well. The network consists of Google's Inception v3 CNN and a fully connected neural network. Finally, we evaluate the model and plot the accuracy, the loss, and also visualize the confusion matrix.

4 Database

We have used a dataset provided for the ISIC 2024 - Skin Cancer Detection with 3D-TBP challenge [1]. The dataset includes diagnostically labeled images accompanied by metadata. The images are in JPEG format, while the corresponding .csv file provides a binary diagnostic label (target), potential input variables, and supplementary attributes such as the image source and detailed diagnosis.

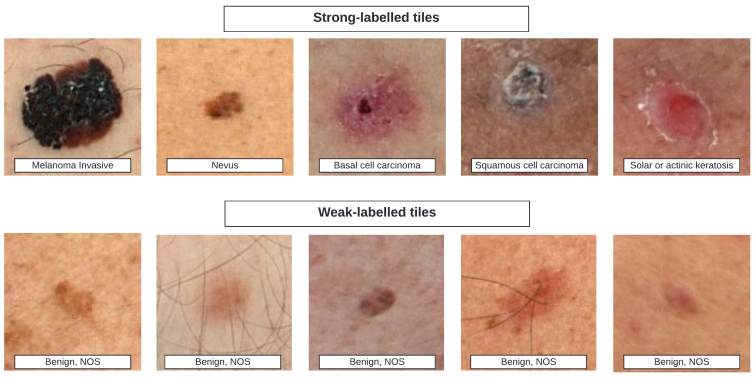

Figure 1: Example of dataset images. Replace with actual image.

The goal is to differentiate benign from malignant cases. We classify each image based on which class it belongs to. To replicate non-dermoscopic images, the dataset utilizes standardized, cropped lesion images derived from 3D Total Body Photography [1].

5 Data Interpretation

To understand the dataset, we first examined its origins and analyzed the accompanying metadata. This provided insights into the distribution of samples across various factors, including age groups, sexes, and bodily locations.

The metadata suggests certain correlations between malignancy and demographic factors. For instance, there appears to be a relationship between the bodily location of a melanoma and its likelihood of being malignant. Studies have shown that melanomas on the head and neck may have a higher malignancy rate due to increased exposure to direct sunlight [6]. However, these melanomas are more likely to be detected early due to their visibility [7]. Given this nuance, we decided not to include the bodily location data in our model to avoid potential biases.

The dataset also exhibits an age distribution skewed towards individuals between 50 and 65 years old. In contrast, the general population's age structure is significantly different, which could introduce bias and lead to

overfitting [8]. Therefore, this factor required careful consideration during the preprocessing phase.

Sex-based differences in melanoma incidence are another important factor. Studies indicate that men are generally more likely to develop malignant melanomas than women. Behavioral factors, such as a lower likelihood of wearing sunscreen, play a significant role in this disparity [5]. However, these behavioral patterns may vary across different populations, which could reduce the generalizability of the model. As a result, we chose not to include sex as a feature in our final model.

Ultimately, we decided to create a model that utilizes only the melanoma images as input, excluding the metadata. This approach ensures that the model remains focused on visual features and minimizes the risk of introducing biases from demographic variables.

6 Architecture, Training, Challenges, and Their Solutions

We utilized Google's Inception V3 model, which has been widely applied in similar projects. The model was pretrained on approximately 1.28 million images across 1,000 object categories from the 2014 ImageNet Large Scale Visual Recognition Challenge. We then fine-tuned it on our dataset using transfer learning [4].

Following this, we implemented a fully connected network composed of dense layers with varying numbers of neurons. The first fully connected layer contained 512 neurons, and the number of neurons was halved in each subsequent layer, down to 128. To prevent overfitting and stabilize the learning process, we employed batch normalization and dropout techniques across these layers. ReLU activation was used for the intermediate layers, while a sigmoid activation function was applied to the final layer to accommodate the binary classification task. This initial setup was later refined through hyperparameter optimization.

One major challenge arose during data visualization: there was a significant imbalance in the dataset, with malignant cases accounting for only

0.01% of the total while benign cases made up 99.9%. This extreme imbalance skewed evaluation metrics, leading the model to classify nearly all cases as benign, achieving apparent accuracy but failing to provide meaningful predictions.

To address this, we oversampled the minority class and simultaneously downsampled the majority class. Despite these efforts, a slight imbalance remained, with the minority-to-majority ratio adjusted to approximately 30%–70%. To mitigate the remaining imbalance, we applied class weights, helping to balance the influence of each class during model training.

Figure 2: Final preprocessing results. Replace with actual image.

We still have our work cut out for us to perfect the model, optimize the hyperparameters, and work on the preprocessing of the dataset.

7 Model Optimization

7.1 Hyper parameter optimization

For hyperparameter optimization, we utilized the built-in Keras tuning framework. This tool allows for automated and efficient exploration of the optimal parameters for our model by leveraging search algorithms, such as random search and Bayesian optimization. By using this approach, we minimized personal biases in the selection process and ensured a more systematic fine-tuning of parameters.

The hyperparameter tuning process involved adjusting various settings, including the learning rate, number of neurons in dense layers, batch size, and dropout rates. Each combination of parameters was evaluated based on validation accuracy to identify the best configuration. The use of automated search methods significantly improved the model's performance compared to manual tuning.

7.2 Threshold optimization

In threshold optimization, various threshold values are tested, and the corresponding V1 score is evaluated for each threshold. The goal is to find the

threshold that results in the lowest V1 score within the specified interval. This process helps in determining the optimal threshold for binary classification, balancing false positives and false negatives effectively.c

8 Results and Their Evaluation

Here are the initial loss and accuracy diagrams, and also the confusion matrix:

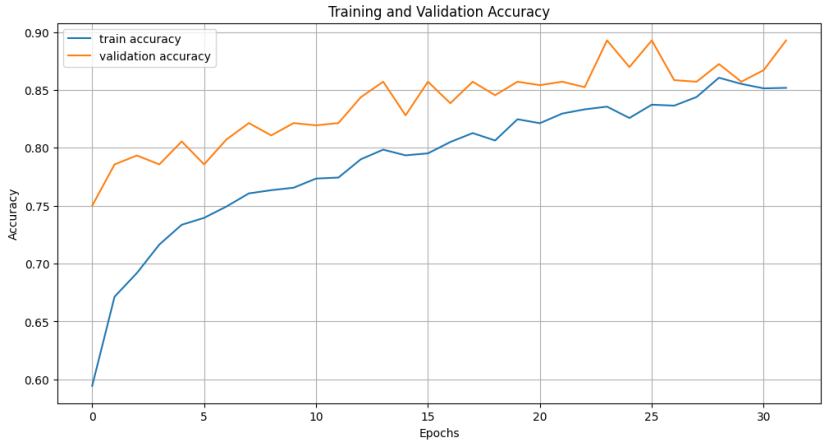

Figure 3: Accuracy graph. Replace with actual image.

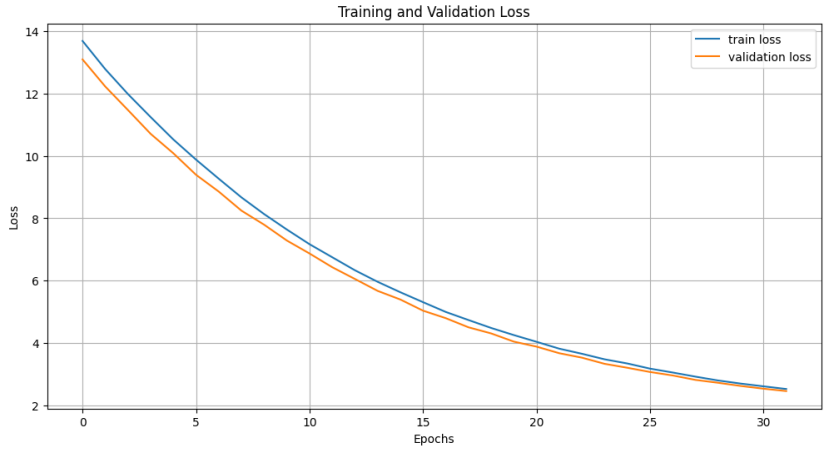

Figure 4: Loss graph. Replace with actual image.

9 Summary

In conclusion, the model performed well, with satisfactory results on the test dataset set aside for final evaluation. However, during the evaluation process, challenges arose due to insufficient validation data for all epochs. While we repeated the training data to mitigate this issue, repeating the validation data would have introduced bias, so this approach was not viable. Unfortunately, these difficulties could not be resolved. Nevertheless, the model achieved an accuracy of 84

10 Appendix

In our work, we utilized assistance from AI models, primarily ChatGPT. We frequently relied on the model when encountering challenges with syntax or comments. It also played a role in the creation of this documentation, mainly for grammar checks. Overall, ChatGPT was an invaluable resource throughout this project, and its contributions deserve recognition.

References

[1] Isic 2024 challenge: Skin cancer detection with 3d-tbp. Available at: `https://www.kaggle.com/competitions/isic-2024-challenge/data`.

[2] A. Author and B. Author. Deep learning approaches for skin cancer detection: A systematic review. *Frontiers in Medicine*, 2023.

[3] C. Author and D. Author. A comprehensive review of cnns for skin lesion detection. *Cancers*, 16(3):629, 2023.

[4] Andre Esteva, Brett Kuprel, Roberto A. Novoa, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639):115–118, 2017.

[5] Centers for Disease Control and Prevention. Melanoma incidence and mortality trends and disparities — united states, 1999–2019. `https://www.cdc.gov/mmwr/volumes/71/wr/mm7122a5.htm`, 2022. Accessed: January 5, 2025.

[6] Y Hemo, M Gutman, and JM Klausner. Anatomic site of primary melanoma is associated with depth of invasion. *Archives of Surgery*, 134(2):148–150, 1999.

[7] R Laskar, A Ferreiro-Iglesias, DT Bishop, MM Iles, PA Kanetsky, BK Armstrong, MH Law, AM Goldstein, JF Aitken, GG Giles, Australian Melanoma Family Study Investigators, Leeds Case-Control Study Investigators, HA Robbins, and AE Cust. Risk factors for melanoma by anatomical site: an evaluation of aetiological heterogeneity. *British Journal of Dermatology*, 184(6):1085–1093, June 2021.

[8] Max Roser, Esteban Ortiz-Ospina, and Hannah Ritchie. Age structure - our world in data. `https://ourworldindata.org/age-structure`, 2024. Accessed: January 5, 2025.