50.039 Theory and Practice of Deep Learning

Group 15

Project Report

Name	Student ID	Contribution
Venkatakrishnan Logganesh	1006050	Model, report
Sherin Karuvallil Saji	1005228	Preprocessing, model
Tang Heng	1006102	Preprocessing, report

Introduction

This paper proposes a deep learning approach for the early detection of skin cancer, leveraging the potential of machine learning models to identify skin lesions indicative of cancer from image data. Skin cancer poses a significant public health challenge globally, with a high survival rate of 99% if detected early. However, traditional screening methods suffer from slow turnaround times, underscoring the urgency of developing more efficient detection techniques.

Our study aims to utilize the HAM10000 dataset to train a neural network model, considering various factors such as pigmentation, skin asymmetry, and demographic risks, to perform binary classification of skin lesions. We plan to employ pre-trained CNN architectures, namely MobileNetV2 and DenseNet201, to extract features from images and utilize a combination of data preprocessing techniques, including contrast stretching and image augmentation. The output will be a binary label indicating the presence or absence of skin cancer, aiming to enhance early detection and contribute significantly to public health efforts against skin cancer. This report will detail our model and our findings regarding the training and testing of our model.

Running Locally

To run the code locally, we would require an anaconda 3.12 environment setup with all the necessary packages and libraries specified in the *requirements.txt* file. For more information, you can refer to the section “Running Locally” under *README.md* file under the main branch of our GitHub Repository (*Sherinksaji/Skincancerpredictionmodel* 2024).

Data Preparation

Dataset

As per the literature review [2], skin cancer is an agglutination of abnormally growing skin cells which vary widely from one individual to another in terms of prominent features like pigmentation, skin, asymmetry, shape and size which can be effectively captured at a more granular depth by the machine learning model through the RGB pixels of the respective image inputs. We are also aware that certain demographic groups (age & gender) could have higher levels of risk exposures to specific kinds of cancer. Moreover, certain cancers can develop more in more localized body regions.

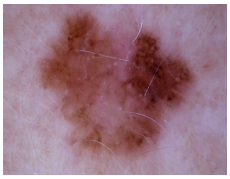
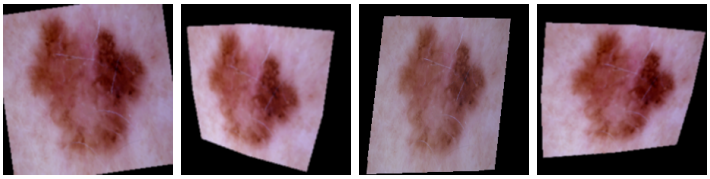
Given these considerations, we intend to use the HAM10000 dataset to train our neural network model. We have reviewed that while both datasets might be a good benchmark, the HAM10000 offers higher generalizability and better relevance as base models which can be improved at a later stage due to its diversity in the extent of skin lesions covered versus ISIC2018 specifically tailored for melanoma detection. In addition, there are structured annotations in the metadata for HAM10000 datasets allowing for asserting ground truths.

Data Loader Object

The targets of the dataset assume that label==1 means that melanoma is present while label==0 means that melanoma is not present. However, the DataLoader class assigns digits to classes based on the alphabetical order of the folder names in the train,test and valid directories which assumes that a person affected by melanoma is in groundtruth label class zero while a person not affected by melanoma is in class 1.

Data Pre-processing

Since HAM10000 is procured from photographs and dermatoscopic images taken by professionals in controlled and clinical settings, the training data may not be able to generalize well to smartphone photographs taken by laymen in highly varied environments. Thus, we apply a series of transformations to the original dataset to help it to generalize better for our application.

Original image	Randomized transformations applied to image
	

The transformations and our justifications:

1. Contrast stretching the range of pixel values from their original values to the full 8-bit range from 0 to 255, allowing image features to be better distinguishable.
2. Random adjustments in brightness and contrast simulate variations in lighting conditions, where the user may take the photograph in a dimly lit room or directly under the sun.
3. Introducing random rotations up to 20 degrees simulates different camera angles.
4. Simulate variations in camera angles and distances with random perspective transforms. Some cameras have different lens dimensions and the user may take the image at varying distances from the skin area of concern.
5. Simulate loss of focus with gaussian blur, a common occurrence in real-world images.

Model

Inputs and outputs

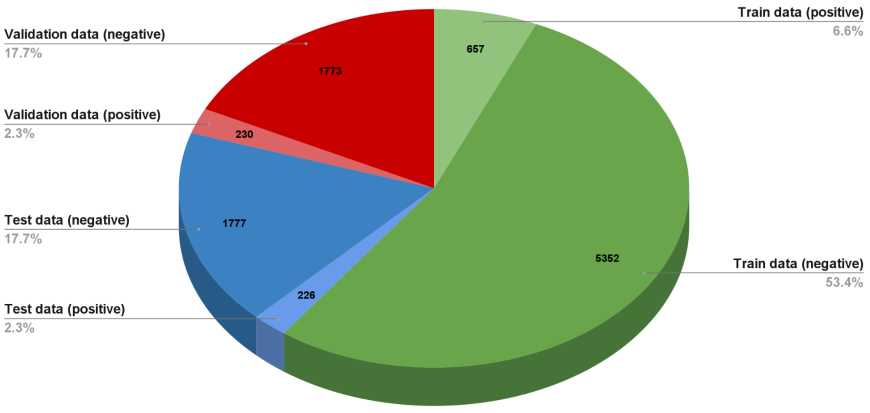
The model's inputs will be a jpeg image of the skin lesion of concern and user input for the lesion's age, sex, and location. We understand that in our literature sources (Sensors | Free Full-Text | Multiclass Skin Lesion Classification Using Hybrid Deep Features Selection and Extreme Learning Machine (mdpi.com)), the ISIC2018 has been specifically used.

As we intend to perform a binary classification task, the expected output will be a label of 0 or 1, indicating whether a person has healthy skin or is at the onset of skin cancer.

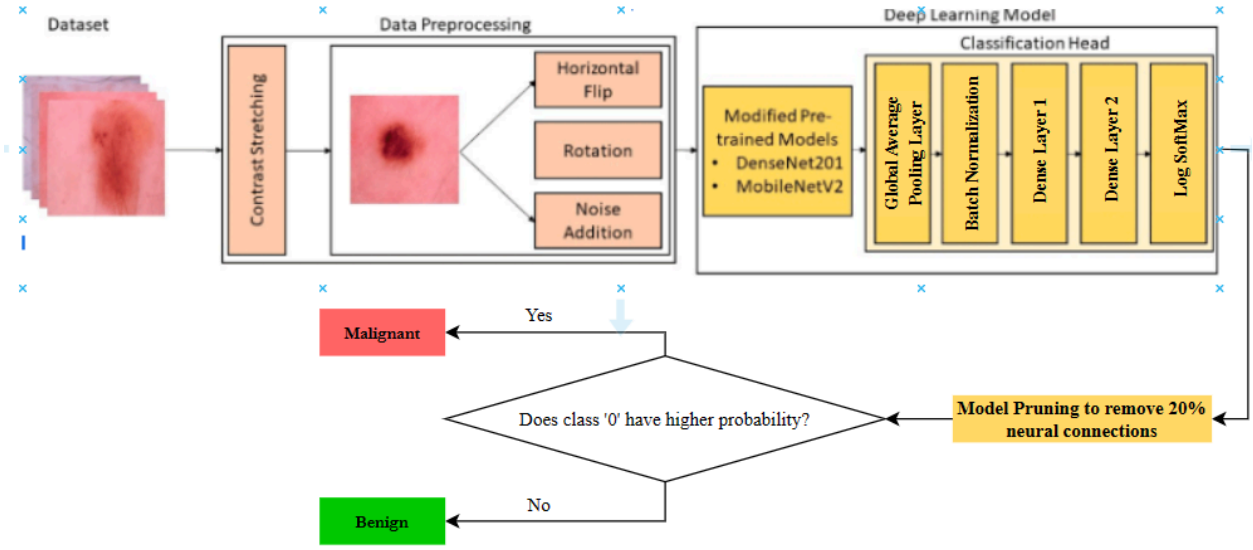
To establish the relevant ground truth labels for the training dataset, we will classify the types of skin cancers into 1 (higher likelihood of being cancerous) or 0 (higher likelihood of being benign) based on the mapping below:

Actinic keratoses, Bowen's disease, basal cell carcinoma, and melanoma are cancerous.
(ground truth label: 1)Benign keratosis-like lesions, dermatofibroma, melanocytic nevi, and vascular lesions are typically not cancerous, though some may have potential risks or associations with malignancy. (ground truth label: 0)

Dataset partitioning



Model Architecture



The input is passed into a pre-trained CNN (Convolutional Neural Network) to extract the features from the input images. These layers will extract features from the input image and will be responsible for identifying prominent features such as pigmentation, size, and asymmetry.

Modified pre-trained models

The CNN models pre-trained are the MobileNetV2 and DenseNet201 models. The former is a lightweight model that is fast and efficient, making it suitable to run on lightweight devices while the latter is a more complex model that is capable of capturing more complex features in the images. By running both the pre-trained models in parallel, the deep-learning model leverages the strengths of both networks.

Classification head

Finally, we have a classification head consisting of Global Average Pooling Layers (GAP Layer), which reduce the dimensionality of the feature maps generated; batch normalization to improve the model's ability to generalize to new data and two fully connected dense layers whose outputs are passed to a log-softmax layer that generates a probability between 0 and 1, which will be compared against a threshold of 0.5 to make predictions on whether an input is cancerous (1 if probability>0.5) or benign (probability<0.5)

Training

Training of the model was conducted on images which were resized to [224,224,3]. As shown in the model architecture, image augmentations were added to increase the pool of images by adding color jitter, contrast stretching, randomized flips and rotations as well as addition of gaussian noise. Such image augmentations allowed the model to become robust and perform better when making predictions on real-time images.

Hyperparameters

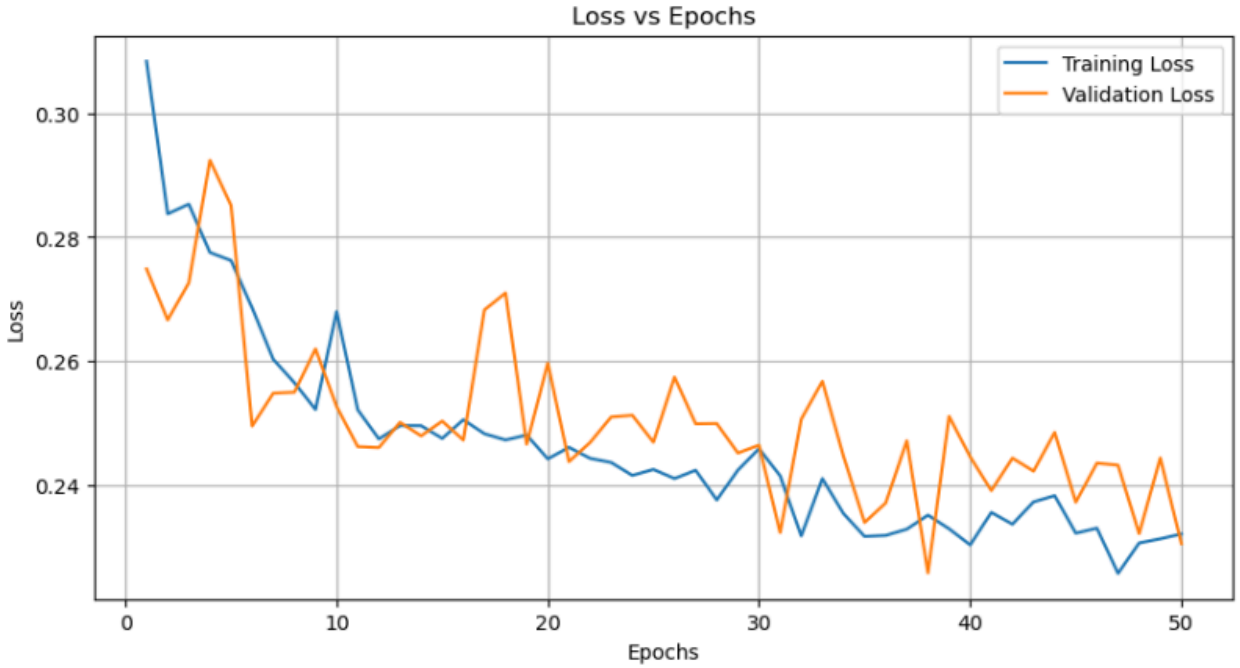
The training data was split and ingested into a data loader in mini-batches of 12, which allowed for mini-batch gradient descent to allow for the model to converge quickly to a minima while allowing it to generalize better. The gradient descent algorithm was implemented via the use of an AdamW Optimizer, a form of Adam Optimizer that treats the weight decay term independently of the gradient update by applying the weight decay term to the parameter weights after the optimizer has performed parameter update based on the gradients, instead of factoring it directly into the gradients. The optimizer implemented allowed for faster convergence of the model on training data while preventing overfitting to the training data. For the optimizer, an optimal learning rate of 0.001 with a weight decay of 0.01 has been used.

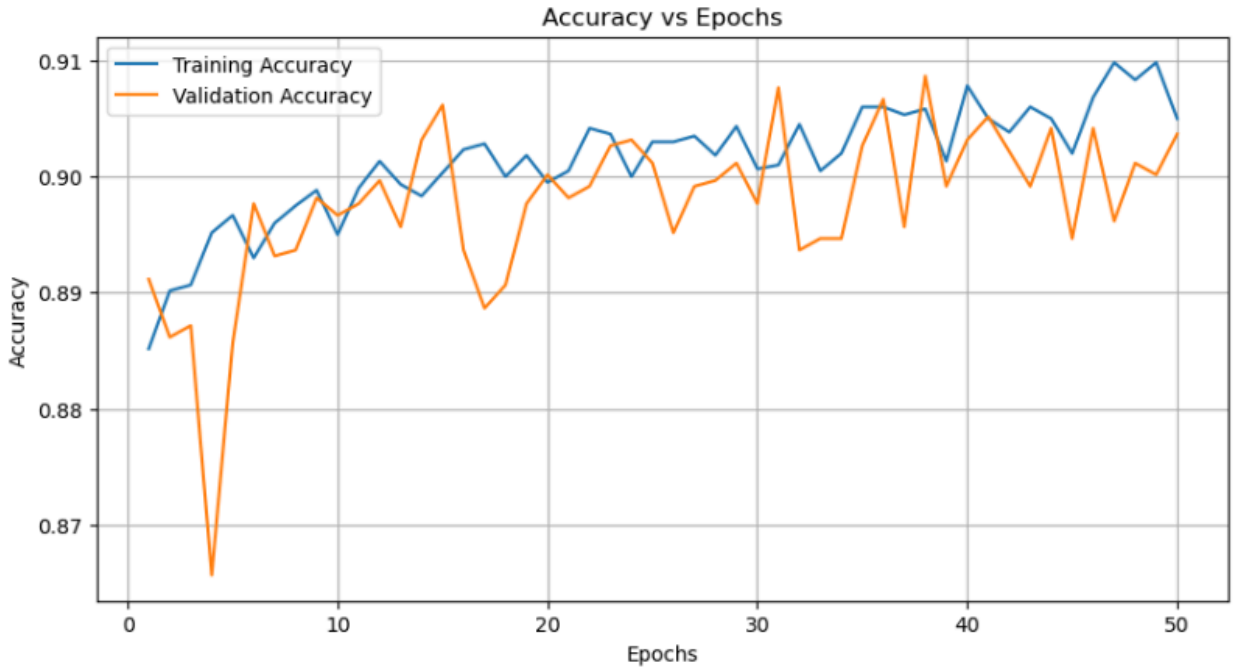
Loss

We use the negative log-likelihood loss / cross-entropy loss for this binary classification task. It calculates the negative logarithm of the predicted probability assigned to the ground truth class.

Model Evaluation

The model was trained for 50 epochs and the graphs of loss as well as accuracy have been plotted below:





As illustrated by the graphs, the loss decreases and the accuracy of the model increases as the number of epochs increases. We observe that beyond 50 epochs, the accuracy as well as the loss plateaus, indicating that the losses and accuracies do not show any appreciable improvement beyond this point. The high average accuracy on validation and test datasets of ~90% beyond 50 epochs is indicative of how the fine-tuned hyperparameters utilized had a positive effect on the model outcome. If we were to train beyond 50 epochs, the validation accuracy might drop and loss might increase resulting in overfitting of model, lack of generalization and eventually, a high variance error..

Results

To assess the performance of our model, we passed the test dataset into our model and assessed it based on the following metrics.

Actual \ **Predicted**	**No melanoma**	**Melanoma**
No melanoma	TN = 55	FP = 171
Melanoma	**FN = 23**	TP = 1754

- Test Accuracy: 0.9031
- Test Precision: 0.9112
- Test Recall: 0.9871
- Test F1 Score: 0.9476
- **Miss rate (FN rate): 0.0129 or 1.29%**

Take note that the miss rate is particularly important in the context of melanoma detection because of the risk of undiagnosed cancers.

Overall, the model achieves high accuracy, precision, recall and F1 score on the test dataset. The low miss rate indicates the model's effectiveness in minimizing false negative predictions, which is directly tied to its reliability.

Model Benchmark

According to (Azeem et al., *SkinLesNet: Classification of skin lesions and detection of melanoma cancer using a novel multi-layer deep convolutional Neural Network* 2023), the SkinLesNet model architecture which implemented a four-layer CNN model with convolution and max pooling layers and categorical classifier head achieved an accuracy of 96%; precision of 97%; recall of 92% and a F1-score of 92% with an objective that involved a multi-categorical classification of whether a lesion is a melanoma, nevus or seborrheic keratosis.

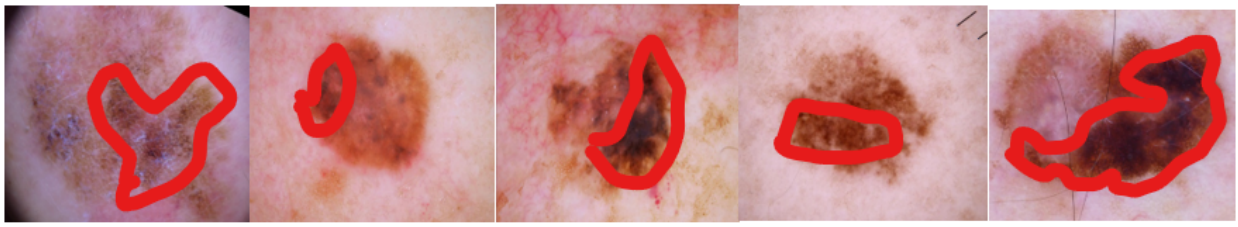
For our objective, a binary classification objective was used to predict if a lesion is melanoma or not. The F1-score and recall obtained outperform the benchmark's metrics, suggestive of a lower overall rate of False Negatives in our model. According to our definition, we classify '0' as the presence of melanoma while '1' is the absence of melanoma. Given the definitions, a lower rate of False Negatives implies a case where the model has a lower probability of predicting patients with non-melanoma lesions as melanoma.

On the other hand, our model underperforms the benchmark in terms of accuracy, revealing a greater tendency for false positive predictions from our model where melanoma lesions are misdiagnosed as non-melanoma.

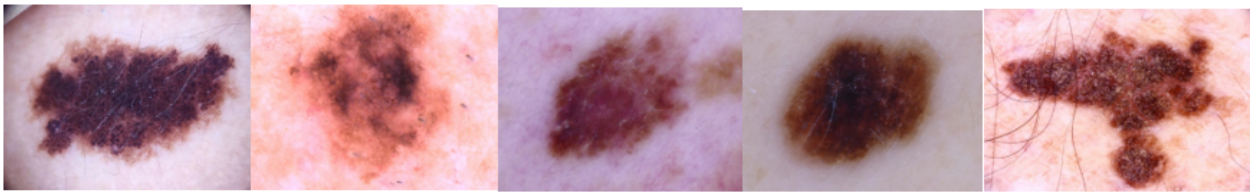
Evaluation

Based on our confusion matrices, we took a few samples of the false positive, true positive, true negative and false negative images. We have displayed the images for your viewing:

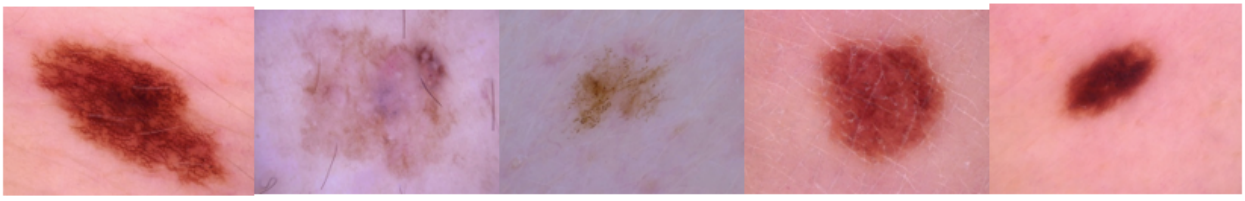
People not affected by melanoma, but model predicted otherwise. (False Negatives)



People affected by Melanoma and model predicts correctly. (True Negatives)



People affected by melanoma and model predicts otherwise (False Positives)



People not affected by melanoma and model predicts correctly. (True Positive)

The top 2 rows of samples (Group 1) are indicative of images which have been predicted as “False” - images present with melanoma while the bottom 2 rows of samples (Group 2) are predicted as “True” - images absent with melanoma. One discriminating factor between Group 1 and Group 2 is the uniformity in pigmentation. While in group 1, the images have high variance in terms of pigmentation intensity, the group 2 images have uniform pigmentation intensity. In particular, we observe that within the melanoma moles of the top 2 rows, there are regions of high intensity of dark pigmentation (shown by the red bounding regions of the 2nd row).

The variance of pigmentation does not serve as a single source of truth for the presence of melanoma. For example, if we look at the third row (False Positive cases), we observe that one characteristic feature defining these lesions in particular is a uniformly dark pigmentation throughout. Another defining factor of the False Positive images is that they seem to be asymmetrical. If we want to improve the model further, we could leverage a segmentation model to extract the shape of the lesion (regular/irregular) in addition to the size (diameter of lesion) and average intensity of the hue of pigmentations as a feature before performing further image augmentations. These extracted features can be passed in parallel through a dense 2-D neural network along with the CNN (Convolutional Neural Network layers) processing the images.

References

1. *Skin cancer*. American Academy of Dermatology. (n.d.). https://www.aad.org/media/stats-skin-cancer
2. Mayo Foundation for Medical Education and Research. (2022, December 6). *Skin cancer*. Mayo Clinic. https://www.mayoclinic.org/diseases-conditions/skin-cancer/symptoms-causes/syc-20377605