SKINSCAN AI: SKIN LESION CLASSIFICATION TOOL

Keshav S
Student# 1006277903
keshav.jindal@mail.utoronto.ca

Abhishek Arora
Student# 0997398427
abhishek.arora@mail.utoronto.ca

Pasut Aranchaiya (Sean)
Student# 1011835935
pasut.aranchaiya@mail.utoronto.ca

ABSTRACT

Skin cancer is the most commonly diagnosed cancer in the United States, with over 5.4 million cases diagnosed annually in US alone. Melanoma, the deadliest form of skin cancer, accounted for 287,723 new cases and 60,712 deaths worldwide in 2018. Early detection significantly increases survival rates, with a 5-year survival rate of over 98.4%. This project aims to develop a hybrid deep learning model that leverages Convolutional neural networks (CNN's) using transfer learning and Artifical neural networks (ANN's), to classify dermoscopic images from the HAM10000 dataset into seven distinct categories of skin lesions. The dataset has 10015 images of skin lesions that range from benign (safe) conditions like melanocytic nevi to malignant ones like melanoma and basal cell carcinoma. Not all skin lesions are cancerous; for example, Melanocytic Nevi are commonly known as "moles". Patient metadata such as age and medical history will be taken into consideration. Prior research demonstrates that CNNs outperform classical models in image classification, achieving 89% accuracy compared to the 86% accuracy of Decision Trees and other traditional approaches. Our model architecture integrates CNNs for visual analysis and ANN's for processing patient metadata to enhance classification accuracy. Ethical concerns, such as bias in datasets and data privacy, are also addressed to ensure fairness and transparency in predictions. This project has the potential to assist dermatologists in accurate skin lesion classification, reducing miss-classification rates and false negatives in malignant case detection. A dangerous false negative is when a Melanoma skin lesion is classified as benign by the dermatologist. —-Total Pages: 9

1 INTRODUCTION

Melanoma is a subset of skin cancer that occurs when pigment producing cells, melanocytes undergo adverse genetic mutation from UV radiation exposure (Bhattacharya et al., 2017). Over the last two decades, the number of patients diagnosed with melanoma has risen steadily. Currently, In 2015, there were 73000 new estimated cases, with over 9000 deaths (Bhattacharya et al., 2017). Although early states have high survival rates (Stage 1A at 97%), without early diagnosis and preventive care, the cancer can quickly spread and become fatal (Stage 4 survival of 10-20%) (Bhattacharya et al., 2017). Poor diagnostic precision adds around 673 million dollars in overall cost for managing the disease (Bhattacharya et al., 2017). The core goal of this project is to classify skin lesions into 7 classes with a focus on malignant lesions: melanoma and BCC (Basal cell carcinoma).

1.1 CURRENT APPROACHES TO IDENTIFYING MALIGNANT SKIN LESIONS

Clinical diagnosis and classification of Melanoma and other skin lesions present significant challenges. The ABCDE method is currently employed for visually classifying pigmented skin lesions for malignancy. Factors such as asymmetry, border irreguality, color variegation, lesion diameter and evolution are taken into consideration (Bhattacharya et al., 2017). The sensitivity results are outlined below. As we can see, general practioners score only 62 % (Menzies, 2005).

Table 1: Melanoma Detection Sensitivity by Different Groups (Menzies, 2005)

Group	**Sensitivity (%)**
Dermoscopy Experts	90
Dermatologists	81
Trainee Dermatologists	85
General Practitioners	62

Note: Sensitivity is the ratio between given by (true positives)/ (true positives + false negatives)

1.2 The relevance of deep learning approaches for skin legion classification

As of 2014, there are only 1.62 per 100,000 dermatologists practicing in Ontario (R., n.d.). This shortage poses a challenge for timely and wide-spread skin cancer diagnostics. An automated diagnosis tool powered by deep learning can bridge this gap by offering a scalable, reliable and cost effective alternative for patients.

Furthermore, deep learning models achieve lower error rates compared to humans. In ImageNet Large Scale Visual Recognition Challenge (ILSVRC) 2015, ResNet got an error rate of only 3.57%. Conversely, the human baseline for image classification is 5.1%. Hence, with sufficient data size, computation powers and advancing algorithms; deep learning models can outperform humans in image classification (APS360: Lecture 5, 2024).

1.3 The HAM10000 dataset and Project Scope

The deep learning model will classify skin lesions in the HAM10000 dataset (10015 images) (Tschandl et al., 2018) into one of 7 categories listed below:

Table 2: Types of Skin Lesions and Their Descriptions (Tschandl et al., 2018)

Lesion Type	**Description**
Melanocytic nevi (nv)	Common benign moles
Melanoma (mel)	A highly dangerous and aggressive form of skin cancer
Basal cell carcinoma (bcc)	A malignant but slow-growing form of skin cancer
Actinic keratoses (akiec)	benign Pre-cancerous lesions that may progress to squamous cell carcinoma
Benign keratosis-like lesions (bkl)	Harmless and benign skin growths
Vascular lesions (vasc)	Benign growths made of blood vessels
Dermatofibroma (df)	Common benign skin growths

We will process the HAM10000 dataset by using two parallel networks - a pre-trained CNN using transfer learning for handling the image data after preprocessing, and an ANN model for processing the metadata features. The CNN branch uses pre-trained weights (transfer learning) and remains frozen, while the ANN branch will be trained to process the metadata. The final features from both networks are concatenated and fed through additional dense layers to leverage both visual and metadata patterns for classifying skin lesions. Patient metadata, such as age, lesion location, and sex, will be utilized by the ANNs, where as the HAM10000 images will be fed into CNN.

1.4 Input and Output data within the model

The HAM10000 dataset is extremely imbalanced with majority lesions being benign (8388 v 1627). Hence, we'll address these concerns in the data processing and primary model sections, such as changing weights to penalize miss-classification of minority classes, data augmentation to increase minority size etc.

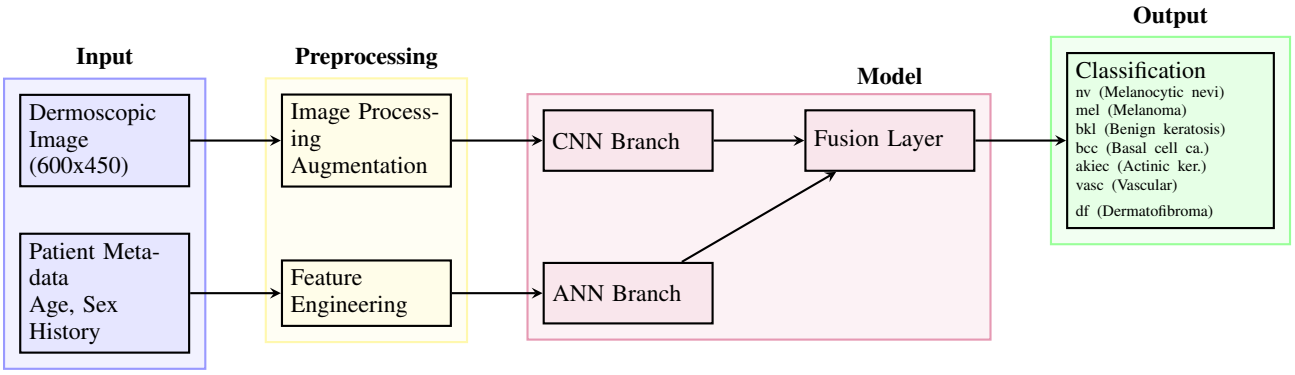


Figure 1: Model architecture: input-output diagram showing data flow within the model

2 Individual Contributions and Responsibilities

2.1 Team Communication and Version Control Tools

Team members communicate with each other via Whatsapp. We are using google Colab for version control, and to code our model. Google Colab has a revision history feature allowing us to keep track of each member's contributions over time. The team meets regularly every week (4-5 pm) to delegate tasks and provide updates. Moving forward, we are going to be using a trello board for managing tasks.

The approach can be broken up into 4 major components. These are data processing, CNN architecture and transfer learning, ANN model, and hyper-parameter tuning. Abhishek will be responsible for data processing, Sean will work on CNN model, and Keshav will work on ANN model and hyper-parameter tuning. It is expected for all team members to contribute to hyper-parameter tuning and data processing.

Table 3: Project plan and Individual contributions

Team Member	**Deliverable Responsibilities**
Keshav Responsible for Hyper-parameter tuning and implementing ANN's in primary model, team coordination	**Completed:** 1) Organizing Team meetings (ongoing) 2) Writing Abstract, Introduction, Ethical considerations, Project plan, Risk register and document formatting/submission in latex for project proposal 3) Writing Abstract, Introduction and Contributions tables in project progress report 4) Working on strategies for data pre-processing to mitigate imbalanced dataset) **Future Work/ In Progress:** 1) Data preprocessing - helping Abhishek with necessary data augmentations to mitigate data set imbalance (Due November 10th 2024) 2) Implementing ANN architecture for Primary model (Due November 15th 2024) 3) Hyper-parameter tuning - Finding necessary hyperparameters like learning rate, batch size etc for optimal performance (Due November 18th 2024) 4) Writing relevant sections in final report, and making video presentation (Due 25th November 2024)

Team Member	Deliverable Responsibilities
Abhishek Responsible for data processing and analysis, research	**Completed:** 1) Finishing data processing section in Project Proposal and Progress report 2) Finding a dataset other than HAM10000 to test our model on (BCN20000 Dataset) 3) Researching approaches used by dermatologists to classify melanoma **Future Work/ In progress:** 1) In charge of Implementing all data processing techniques (Due November 10th, 2024) 2) Helping rest of team in hyper-parameter tuning (Due November 18th, 2024) 3) Writing relevant sections in final report, and making video presentation (Due 25th November 2024) 4) Performing actively in team meetings (Ongoing)
Sean Responsible for implementing transfer learning in CNN, Overseeing model design and optimization	**Completed/In Progress:** 1) Finishing background related work and baseline model in Project proposal 2) Implementing all machine learning models for the baseline, and finishing writeup (project progress report) 3) Brain-storming data processing strategies for primary model to mitigate unbalanced dataset 4) Designing CNN and ANN architecture for primary model with Keshav & Abhishek **Future Work:** 1) Implementing different transfer learning approaches for CNN architecture in primary model (Due November 15th, 2024) 2) Brainstorming different approaches to hyper-parameter tuning (Due November 16th, 2024) 3) Performing actively in team meetings (Ongoing) 4) Writing relevant sections in final report, and making video presentation (Due 25th November 2024)

2.2 PROJECT STATUS

Currently, we have a model that can classify the HAM10000 data into 7 skin lesions with the sensitivity scores and F1 scores shown below. The primary model has an F1 score of 0.67. We want to have our model to have a sensitivity score exceeding 80% when specifically classifying melanoma and BCC (Basal cell carcinoma) (along with being able to accurately classify other lesions). Currently General practitioners can only achieve 62% sensitivity in classifying melanoma (Menzies, 2005) . Hence, an accurate deep learning model will help them immensely. Note: the table is used later in the document as well.

Table 4: Model Performance Metrics Comparison

MODEL	F1	SENSITIVITY
Logistic Regression	0.75	0.77
Decision Tree	0.71	0.72
Random Forest	0.72	0.78
Simple CNN	0.73	0.705
Primary Model	0.67	0.62

2.3 Project Risks

1) Computational power: As the model complexity rises, the computation time may be excessively long. To mitigate this, one of the team members will taken a google colab pro subscription to have access to more powerful gpu's like Nvidia A100. This will substantially reduce computation time.

2) Team member dropping course: One of our team members, Youssef abruptly dropped the course the day that progress report was due. His part was not finished. To mitigate against this scenario in the future, we will enforce a strict internal deadline so that we do not have unfinished sections on the due date.

3 Data processing

3.1 Data Sourcing

Our data will be sourced from the popular data science platform Kaggle. Specifically, "Skin Cancer MNIST: HAM10000" [10]. The HAM10000 dataset is a large collection of labeled dermatoscopic images aimed at supporting skin lesion classification research. It contains 10,015 images of pigmented skin lesions from 7 classes(Mader, 2018). Throughout the course of this project, we have found out that our model suffered from the quantity of HAM10000 dataset, leading to unoptimized performance; therefore, we have planned to include the BCN20000 dataset to increase the amount of data and boost our model's performance.

3.1.1 HAM10000 Dataset

The providers of the HAM10000 dataset found out that neural network models have been suffering through the lack of available data and its diversity. Therefore, HAM10000 ("Human Against Machine with 10000 training images") has been created from gathering dermoscopic images from various populations, accumulating up to 10015 dermatoscopic images(Tschandl et al., 2018). As images are gathered with various methods, multiple data preprocessing methods have been applied by the provider, resulting in pictures centered on the skin cancer without any black border. In addition to the images, the dataset includes 4 features: type of diagnosis (dx_type), age, sex, and localization. These features correspond to a label which is one class of seven important pigmented lesions types, included by the provider, including Actinic keratoses and intraepithelial carcinoma / Bowen's disease (akiec), basal cell carcinoma (bcc), benign keratosis-like lesions (bkl), dermatofibroma (df), melanoma (mel), melanocytic nevi (nv), and vascular lesions (vasc)(Mader, 2018).

3.1.2 BCN20000 Dataset

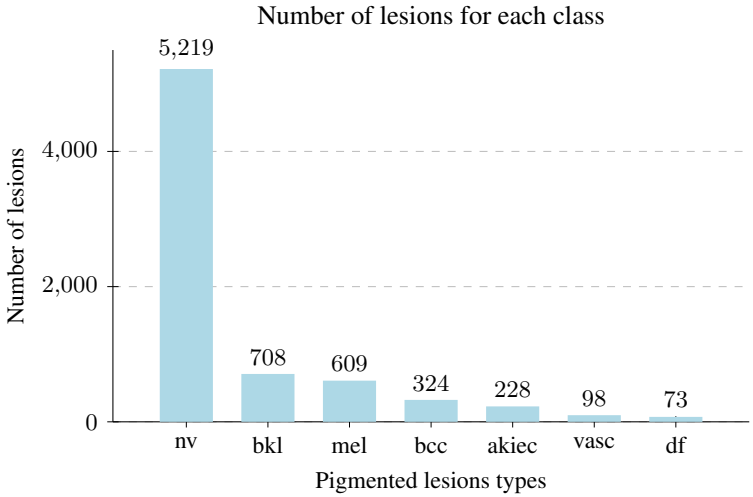
The purpose of the BCN20000 dataset followed the HAM10000 dataset; however, it was also aimed to be used in classification of unconstrained dermoscopic images of pigmented lesions in hard-to-diagnose locations(Hernández-Pérez et al., 2024). Due to this and the fact that BCN20000 is not preprocessed, images from this dataset suffer from a lot of diversity like black borders, uncentered images, and distinct patterns of images from various locations. The BCN20000 consists of 18,946 dermoscopic images gathered from the Hospital Clínic in Barcelona, Spain from 2010 to 2016, and it includes 3 features: age, sex, and localization(Hernández-Pérez et al., 2024). Each image and corresponding features belongs to one of eight classes, which are seven classes of HAM10000 and an additional class Squamous carcinoma (SCC).

3.2 Data Cleaning

The dataset contains certain rows with values of unknown for "sex", "localization" as well as "age". These values need to be removed by manner of dropping said rows from the dataset. Overall, there are 57 rows with null values for the age column, 234 rows with unknown values for the localization column, and 57 rows with unknown values for the sex column.

The dataset also has multiple duplicates. We conducted a test to find out if duplicates are from the rows with the same values in "lesion_id" or the rows with the same values in all the columns. Then, we found that rows with the same values in "lesion_id" but different values in other columns

are different data points; therefore, we removed the rows with the same values in all the columns. Next, we remove the irrelevant columns, specifically "lesion_id" and "dx_type". The "sex" and "localization" columns are then encoded using one-hot encoding, while the "age" column is retained as numerical values. Finally, a 70%-15%-15% train-validation-test split is done on the cleaned dataset.



3.3 Data Processing (Taken from ISIC 2018 winner)

As images from the HAM10000 dataset are fairly preprocessed, we apply minor image processing steps with the purpose to include some randomness and eliminate unnecessary features by cropping. For each image, we resize so the height is 1.25x the input height, while maintaining the aspect ratio. Then, a random square center crop with size between 0.8 and 1 of the resized image is implemented. However, the validation and test set take a fixed crop with size 0.9. After that, the image is resized again to our desired input size (Nozdryn-Plotnicki et al., 2018). Finally, the image is normalized and ready to be passed into our neural network models. This data processing step will be our starting point, and adjustment will be done in the future to improve the model.

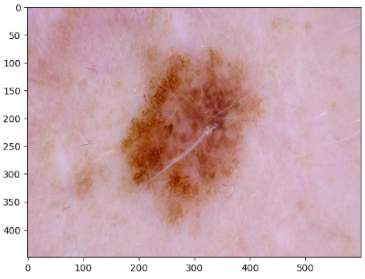


Figure 2: Unprocessed image

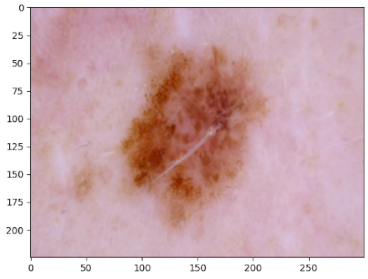


Figure 3: Processed image

3.4 Data Augmentation(Taken from ISIC 2018 winner)

Our approach on data augmentation has two purposes which are creating randomness to increase the model's generalization ability and increasing the number of data for minority classes by adding random copies of the original image to solve data imbalance. Therefore, data augmentation is only done on classes of "akiec", "df", and "vasc" in the training set. Each image in those classes has a probability of 50% in creating a copy of its horizontal flipped image and a probability of 50% for creating a copy in each of 90 degrees, 180 degrees, and 270 degrees rotations. In addition, a copy of the original image with brightness, saturation, and contrasts augmented randomly by a factor

between 0.9 and 1.1 will be included (Nozdryn-Plotnicki et al., 2018). This random augmentation will be done on every epoch of the model's training to generate even more randomness and noise.

3.5 Challenges

3.5.1 Small amount of data

The total amount of data after preprocessing accumulated to 7259 data points with only 73 data points in the smallest class. As our primary model is fairly complex, we need more data in the minority classes in order to solve overfitting and achieve good prediction on all the classes. Therefore, we have considered using the BCN20000 dataset.

3.5.2 Imbalance dataset

The second largest class accounts for only 13.6 % of the largest class, which shows a big imbalance in our dataset. We have not added BCN20000 to our approach yet, so we solved this issue by adding data augmented copies on minority classes to increase their quantity, using F1 score and sensitivity as evaluation metrics, and implemented weighted cross entropy loss which increases loss for minority classes.

3.6 Future Testing datasets

To obtain additional testing data we found datasets from Kaggle specifically the dredgen743 skin cancer Dataset (Dredgen743, 2024). This dataset contains a lot of duplicates from the HAM10000 dataset. Fortunately, this dataset comes with a metadata file that contains the image id of all the images in the dataset. We can easily compare the image_id of this new dataset with those of the HAM10000 to isolate images not in both, which can be used for testing our model. The metadata file also includes features which are age, sex, and localization needed for the input of our model.

4 Baseline Model

We implemented four simple baseline models which are Logistic Regression, Decision Tree, Random Forest, and a Simple CNN model to evaluate our primary model's performance. As our data suffers from data imbalance, we have used F1 score and sensitivity as our evaluation metrics.

4.1 Logistic Regression

Based on a research named "Towards Skin Cancer Classification Using Machine Learning and Deep Learning Algorithms: A Comparison", the performance of Logistic Regression on skin cancer classification task is slightly worse than CNN (Khan & Jr, 2021). This means that Logistic Regression will be a suitable competitor to our primary model. From this and the fact that Logistic Regression is a simple machine learning model which requires minimal tuning, we have evaluated that Logistic Regression is a suitable choice. To prepare inputs for the model, unprocessed images are resized to a width of 24 pixels and a height of 18 pixels. These resized images are then flattened and concatenated with the metadata to form a comprehensive input vector. The model takes the standardization of these input vectors as inputs. Due to its low complexity, the model demonstrated remarkable robustness against over fitting, enabling it to perform well even on a small dataset with a large number of features. This led to an impressive F1 score of 0.75 and sensitivity of 0.77 on the validation set.

4.2 Decision Tree & Random Forest

Decision Tree, a model which performs well on tabular data, has been selected as a comparison to our primary model with the expectation that our primary model will perform better on our data, which is images. However, our choice faced a significant problem that Decision Tree and Random Forest are vulnerable to over fitting, meaning that a fair amount of tuning is necessary. The preparation of the inputs of these models followed the methods in Logistic Regression with standardization removed. Through multiple experimentation with various parameters, we found that Decision Tree and Random Forest are unable to perform well on our dataset, which has high dimensional features

and small amounts of samples. The reason is that Decision Trees and Random Forests are prone to over fitting our data, and applying extensive tuning to mitigate this issue would overly constrain the models, preventing them from fitting the data effectively. As a result, the Decision Tree is able to achieve a validation F1 score of 0.71 and a validation sensitivity of 0.72, and the Random Forest is able to achieve a validation F1 score of 0.72 and a validation sensitivity of 0.78.

4.3 Simple CNN model

As CNNs models have been prominent in skin cancer classification tasks, we expect a simple CNN to perform quite well. To increase the performance even further, we added complexity to CNN and fully connected layers on the metadata. Considering the added value of complexity and large size of data, our primary model should achieve greater performance than these state-of-the-art models. For our simple CNN, we have used a CNN with two convolution layers, two pooling layers, and two fully connected layers on processed images with no augmentation. The model is able to converge, achieving the best validation F1 score of 0.73 and the best validation sensitivity of 0.705.

Table 5: Model Performance Metrics Comparison

MODEL	**F1**	**SENSITIVITY**
Logistic Regression	0.75	0.77
Decision Tree	0.71	0.72
Random Forest	0.72	0.78
Simple CNN	0.73	0.705
Primary Model	0.67	0.62

4.4 Challenges

Our primary model achieved the worst F1 score and sensitivity. As complex neural network architecture needs large amounts of data, we assumed the reason for this result is lack of available data, leading to better performance in simpler models. We will solve this issue by adding BCN20000, and explore more on our approach to improve the primary model's performance.

5 Primary Model

5.1 Architecture

The architecture chosen for our primary model combines a Convolutional Neural Network (CNN) for image processing and a fully connected Artificial Neural Network (ANN) for metadata processing. This hybrid approach leverages both visual and non-visual information, which is particularly valuable in complex, multi-faceted tasks like skin cancer classification.

The CNN portion of our model is built on ResNet18, a pre-trained architecture with a proven record in image analysis tasks. We utilize transfer learning here, freezing the pre-trained weights to capitalize on the extensive feature extraction capabilities learned by ResNet18 without incurring the computational costs of fine-tuning over 11 million parameters. This decision to freeze the weights also minimizes the risk of overfitting given our limited training data and available resources, while still allowing the model to capture essential visual patterns in skin lesion images.

For metadata processing, we include a dedicated ANN. This ANN consists of two layers with dropout on the first layer to help prevent overfitting on the smaller metadata input. The ANN accepts a vector of 17 features (metadata specific to each patient, such as age or lesion location), transforms this through an intermediate representation of 64 and 32 dimensions with ReLU, and outputs a compact embedding of size 32 before being concatenated with the CNN's flattened feature map. The model consists of 3232 trainable parameters.

To achieve the final classification, we concatenate the feature map from the ResNet18 CNN with the output from the metadata-processing ANN. This combined feature vector is then fed into the final two fully connected layers with 70,663 trainable parameters. The first layer maps the concatenated

feature vector to a vector of size 128 with ReLU activation and dropout for robustness. Then, the final layer produces a 7-class output vector corresponding to the different skin lesion types in the HAM10000 dataset, providing a multiclass classification output.

5.2 DIAGRAM

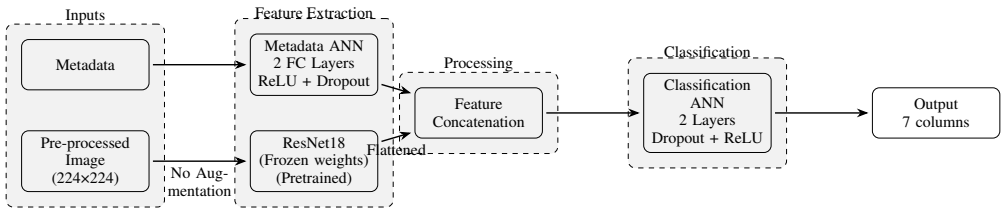


Figure 4: Network architecture overview

5.3 MODEL JUSTIFICATION

This hybrid CNN-ANN approach was chosen to leverage both the image content and patient metadata, as the HAM10000 dataset offers additional features beyond just image data. Skin lesion classification can benefit from contextual information, and by integrating metadata, our model can learn subtle patterns that might correlate lesion characteristics with patient demographics, potentially boosting classification accuracy. For instance, certain lesions may appear more frequently in specific age groups or body locations.

The CNN component, designed with ResNet18, focuses on capturing visual features such as texture, color, and shape from lesion images, while the ANN component focuses on patient-specific features. By combining these two sources of information, we enhance the model's ability to differentiate between lesion types.

5.4 TRAINING AND EVALUATION

At first, our primary model is trained on preprocessed and augmented data; however, the result showed that the model is unable to converge, resulting in the removal of image augmentation. To solve the problem of data imbalance, we implemented weighted cross entropy loss where the loss of each class is multiplied by `N/Ni` (N is the total samples in the training set and `Ni` is the number of samples in each class), allowing the model to not overlook the minority classes. We implemented this knowledge and trained the model using adam optimizer with a learning rate of 0.001 for 10 epochs and a processed input size of 224x224. To optimize our model, we save its parameter every epoch and the model with the highest F1 score is chosen as our final model. Our best model is able to achieve a validation F1 score of 0.67 and a validation sensitivity of 0.62. As our model is very complex, we believed the biggest reason for the unimpressive result is insufficient amount of data.

5.5 CHALLENGES

Throughout this task, we faced challenges in balancing computational efficiency with model complexity. Freezing the CNN weights was a necessary decision to ensure feasible training times; however, we believe training the pretrained CNN is also necessary to achieve better performance. Another challenge has been on optimizing the model on a small dataset, which we have decided that the best solution is to include more data. We would also need to adjust data augmentation to be suitable to our model.

6 GOOGLE COLAB LINK

`https://colab.research.google.com/drive/1KCFxgy9dRUaq4mT5ytLKJ6Fz9WFWkvL8?usp=sharing`

REFERENCES

APS360: Lecture 5. Week 5: Convolutional neural networks - part ii, 2024. URL https://q.utoronto.ca/courses/363324/files/33584930?module_item_id=6191305. Lecture slides, University of Toronto.

Arjun Bhattacharya, Andrew Young, Alex Wong, Seth Stalling, Michelle Wei, and Drew Hadley. Precision diagnosis of melanoma and other skin lesions from digital images. In *AMIA Joint Summits on Translational Science Proceedings*. AMIA, July 26 2017. https://pmc.ncbi.nlm.nih.gov/articles/PMC5543387/#r1-2612795.

Dredgen743. Skin cancer. Kaggle, May 1 2024. URL https://www.kaggle.com/datasets/dredgen743/skin-cancer.

Carlos Hernández-Pérez, Marta Combalia, Sergi Podlipnik, Noel C. F. Codella, Veronika Rotemberg, Allan C. Halpern, Ondrej Reiter, Cristina Carrera, Antonio Barreiro, Brian Helba, Susana Puig, Veronica Vilaplana, and Josep Malvehy. Bcn20000: Dermoscopic lesions in the wild. *Scientific Data*, June 17 2024. doi: 10.1038/s41597-024-03387-w. https://www.nature.com/articles/s41597-024-03387-w.

Iqra Khan and Ijist Jr. Towards skin cancer classification using machine learning and deep learning algorithms: A comparison. 3:110–118, 12 2021. doi: 10.33411/IJIST/2021030508.

Kai S. Mader. Skin cancer mnist: Ham10000. Kaggle, September 20 2018. URL https://www.kaggle.com/datasets/kmader/skin-cancer-mnist-ham10000.

Scott W. Menzies. The performance of solarscan. *Archives of Dermatology*, November 1 2005. URL https://jamanetwork.com/journals/jamadermatology/fullarticle/400665.

Aleksey Nozdryn-Plotnicki, Jordan Yap, and William Yolland. Ensembling convolutional neural networks for skin cancer classification. In *International Skin Imaging Collaboration (ISIC) Challenge on Skin Image Analysis for Melanoma Detection. MICCAI*, 2018.

M. R. R. Changes in the practice patterns and demographics of ontario dermatologists. *Journal of Cutaneous Medicine and Surgery*, n.d. URL https://pubmed.ncbi.nlm.nih.gov/29519145/.

Philipp Tschandl, Cliff Rosendahl, and Harald Kittler. The ham10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Scientific Data*, 5, 2018. doi: 10.1038/sdata.2018.161. https://www.nature.com/articles/sdata2018161.