Skin Cancer Detection

Kislay Kislay
Department of Mathematics
Stevens Institute of Technology
Hoboken, United States
kkislay@stevens.edu

Mohammed Salman Taqi
Department of Mathematics
Stevens Institute of Technology
Hoboken, United States
m4@stevens.edu

***Abstract*—The most common type of cancer in the country is skin cancer. In the United States, more than 4 million cases of skin cancer are found each year. In order to reduce effort, time, and human life, an accurate automated system for skin lesion recognition is absolutely necessary for early detection. The goal of this project is to develop a deep learning model that can classify skin lesions at an early stage in order to accurately diagnose the illness and aid in clinical decision-making, increasing the likelihood that the condition can be treated before it spreads. We would be uploading all the code on github in the link mentioned below:**

https://github.com/kislay09/Deep_Learning_CS583

I. Introduction

Skin cancer is one of the most frequent cancers not only in the United States, but also worldwide, with about 10,000 people diagnosed with it in the United States every day. De- spite the fact that the number of Melanoma deaths is expected to rise by 22% in the coming year, early identification of the disease can lead to a 99% survival rate. Early detection of skin cancer is critical and can prevent further spread in some cases, such as melanoma and focal cell carcinoma. In any case, there are several factors that have a negative impact on detection accuracy. In recent years, the use of image processing and computer vision in healthcare and medical applications has increased significantly. In this study, we use deep learning models to detect and categorize cancer based on dermoscopic images of pigmented lesions.

II. Related Work

Skin cancer classification has been a focal point in medical image analysis, leveraging advancements in machine learning and computer vision techniques. Various studies have explored different methodologies and datasets to enhance classification accuracy and diagnostic capabilities.

One prominent study by Esteva et al. (2017) utilized a deep convolutional neural network (CNN) architecture trained on a large dataset of dermoscopic images. Their work demonstrated significant progress in classifying skin lesions into multiple categories, achieving performance on par with dermatologists in identifying melanoma and other common skin diseases. Furthermore, Tschandl et al. (2018) presented an extensive evaluation of different CNN architectures and data augmentation techniques using the International Skin Imaging Collaboration (ISIC) dataset. Their findings emphasized the importance of data preprocessing and augmentation in improving classification accuracy across various skin lesion types. Another notable contribution by Haenssle et al. (2018) highlighted the effectiveness of a machine learning-based algorithm in supporting dermatologists during clinical practice. Their AI system showed high sensitivity in detecting melanoma, aiding medical professionals in making accurate diagnostic decisions. Moreover, recent research by Codella et al. (2019) introduced a dataset focusing on challenging aspects of skin lesion analysis, emphasizing the

need for robust models capable of handling diverse clinical scenarios. Their dataset incorporated different skin types, ages, and environmental conditions, posing significant challenges for classification algorithms. In summary, recent advancements in skin cancer classification have primarily revolved around deep learning techniques, dataset curation, and the integration of machine learning systems into clinical workflows. These studies showcase the potential of AI-driven approaches in assisting dermatologists and improving diagnostic accuracy in skin cancer detection.

III. Description of Dataset

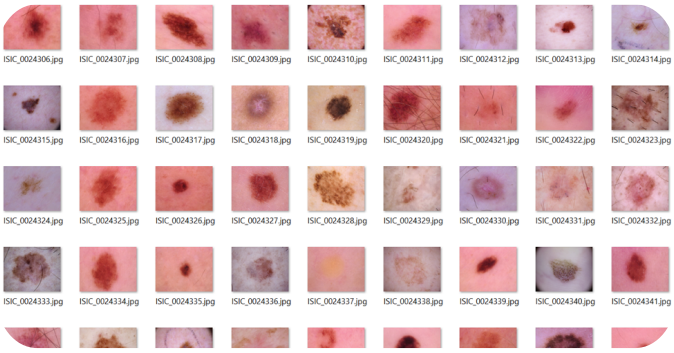
The HAM10000 ("Human Against Machine with 10000 training images") dataset is a significant resource in dermatology, providing a vast array of dermatoscopic images aimed at advancing the machine learning applications in the field. Released by Harvard University in June 2018, the dataset comprises 10,015 dermatoscopic images of skin lesions, which are instrumental for both educational purposes and algorithmic development in the automated diagnosis of skin cancer. This comprehensive collection includes a diverse set of images representing various demographics. Accompanying these images is a metadata file that contains detailed demographic information pertaining to each lesion. Validation of over half of the lesions is conducted through histopathology (histo), with the remaining confirmed by follow-up examinations, expert consensus, or confocal microscopy. This multifaceted approach to validation underscores the reliability and diagnostic value of the dataset.

The accompanying file, HAM10000_metadata.csv, is particularly valuable, encapsulating critical information such as the type of skin lesion depicted in each image. Analysis of the metadata reveals key features like age, gender, body localization, and cell type of the lesions. Understanding and leveraging this metadata is crucial for researchers to effectively use the dataset in machine learning models, ensuring that the features extracted are relevant and predictive. The dataset is publicly available

	lesion_id	image_id	dx	dx_type	age	sex	localization
2	HAM_0000118	ISIC_0027419	bkl	histo	80	male	scalp
3	HAM_0000118	ISIC_0025030	bkl	histo	80	male	scalp
4	HAM_0002730	ISIC_0026769	bkl	histo	80	male	scalp
5	HAM_0002730	ISIC_0025661	bkl	histo	80	male	scalp
6	HAM_0001466	ISIC_0031633	bkl	histo	75	male	ear
7	HAM_0001466	ISIC_0027850	bkl	histo	75	male	ear
8	HAM_0002761	ISIC_0029176	bkl	histo	60	male	face
9	HAM_0002761	ISIC_0029068	bkl	histo	60	male	face
10	HAM_0005132	ISIC_0025837	bkl	histo	70	female	back
11	HAM_0005132	ISIC_0025209	bkl	histo	70	female	back
	HAM_0001396	ISIC_0025276	bkl	histo	55	female	trunk

for research and educational purposes, reflecting an ongoing commitment to open science and the promotion of AI in medicine. For access to the HAM10000 dataset and to explore the possibilities it presents, the dataset can be found at the following URL:

https://www.kaggle.com/datasets/kmader/skin- cancer-mnist-ham10000



IV. Data Preparation/Preprocessing

Image Data Merging: To compile a comprehensive dataset, we integrated image data from two distinct folders. This process involved the creation of a dictionary that linked image IDs with their corresponding file paths and lesion types. The integration not only consolidates the data but also simplifies access and manipulation.

Metadata Handling: A CSV file containing essential metadata for each image was imported into a pandas DataFrame. This metadata encompasses various attributes,

with 'dx' being a pivotal one, as it provides a concise code indicating the type of skin lesion depicted in each image.

Lesion Type Dictionary: We constructed a dictionary (lesion_type_dict) to map the diagnostic codes ('dx') to their full descriptive names. This mapping facilitates a clear understanding of the lesion types, making the data more accessible and interpretable.

- Melanocytic nevi (nv)
- Melanoma (mel)
- Benign keratosis-like lesions (bkl)
- Basal cell carcinoma (bcc)
- Actinic keratoses (akiec)
- Vascular lesions (vas)
- Dermatofibroma (df)

Label Mapping: For the purpose of training machine learning models, we created two dictionaries: label_mapping and reverse_label_mapping. The former converts textual labels into numerical values suitable for algorithmic processing, while the latter allows for the reversion of numerical labels back to their original text format.

Resizing: Each image in the dataset was resized to a uniform shape to ensure compatibility with the neural network's input layer. Standardizing the image size, commonly to 28x28 pixels, is essential for consistent processing by the model.

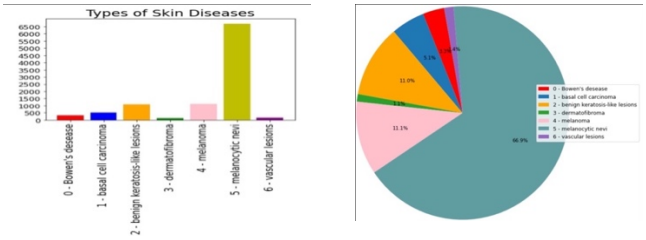
Normalization: The pixel values of the images, originally ranging from 0 to 255, were normalized to fall between 0 and 1. This normalization, achieved by dividing the pixel values by 255, optimizes the network's learning efficiency by scaling the inputs to a more appropriate range.

Train-Test Split: The dataset was partitioned into a training set and a test set. The former is utilized to train the model, while the latter assesses the model's performance, ensuring its ability to generalize to new, unseen data.

Data Augmentation: To bolster the model's generalization capability and mitigate overfitting, we employed data augmentation techniques. This strategy generates variations of the training images through transformations such as rotation, shifting, and flipping. Data augmentation enables the model to recognize features under varying conditions, thereby enhancing its robustness.

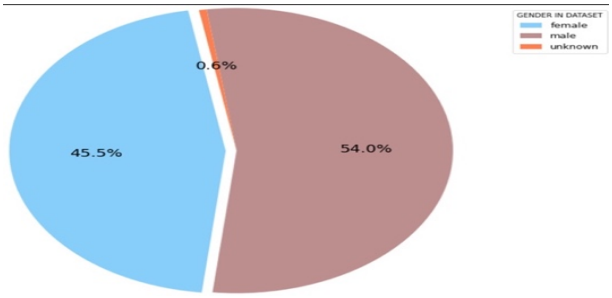
V. Data Visualization

1) Types of Skin Diseases and their Counts



The following graphs show that the dataset is imbalanced, and the highest number of cases are that of 'melanocytic nevi' which are approximately 6500 in total. The lowest number of cases are that of 'vascular lesions' which are approximately close to 200 in total.

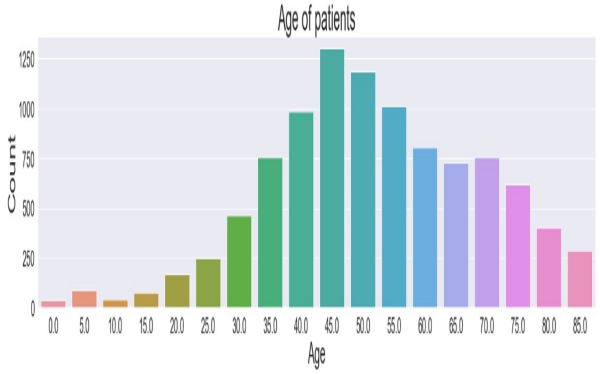
2) Distribution of skin cancer cases between male and female



The above pie-chart shows that 54% of the skin cancer patients are male and 45.5% of the skin cancer patients are female.

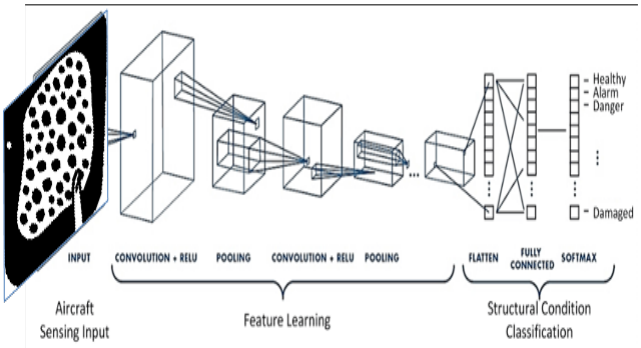
3) Distribution of age

The above plot shows the distribution of patients' age spread over the entire dataset. The distribution is roughly a normal distribution with the maximum number of patients of age 45 and minimum number of patients with age 10.



VI. Model Architecture

Our neural network model for classifying dermatoscopic images of skin lesions is built using a sequential architecture, which is a linear stack of layers. The model is designed to



recognize patterns and features indicative of various types of skin lesions, facilitating accurate classification. Below is an outline of the model's architecture:

We implemented multiple convolutional layers (Conv2D) with varying filter sizes (16, 32, 64, and 128), which are the building blocks of Convolutional Neural Networks (CNNs). These layers are responsible for extracting features from the images These layers apply a number of filters to the input to create a feature map that summarizes the presence of detected features in the input. Each convolutional layer is followed by a max pooling layer (MaxPool2D) that reduces the spatial dimensions of the feature map, effectively downsampling the input representation and reducing the number of parameters. Before the fully connected layers, a flattening layer (Flatten) is used to transform the two-dimensional feature maps into a one-dimensional vector. This vector serves as the input to the fully connected layers. The architecture includes one or more dense layers (fully connected layers) where every neuron in the layer is connected to every neuron in the preceding layer. These layers are used to classify the features extracted by the convolutional layers into the respective categories. The model includes fully connected layers (Dense), with 64 and 32 neurons, which perform classification based on the features learned by the convolutional and pooling layers. The last fully connected layer has a number of neurons corresponding to the number of classes in the dataset and uses a softmax activation function to output probabilities for each class. The final layer is a dense layer with 7 neurons, corresponding to the seven skin cancer categories. It typically uses the softmax activation function for multi-class classification to provide the probabilities of the input being in each target class.

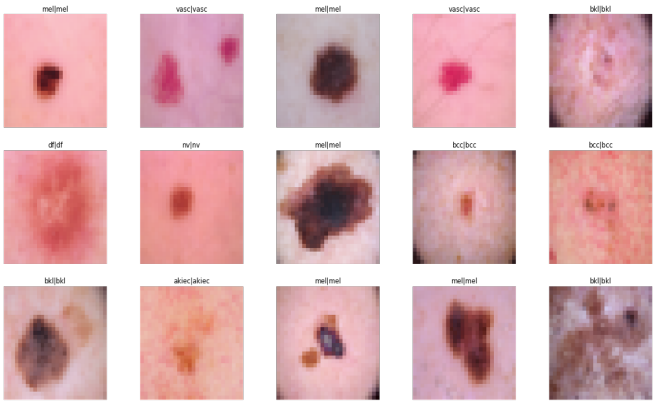
VII. Model Implementation

The CNN was trained using augmented image data to improve the model's robustness and ability to generalize. This augmentation involved rotating, resizing, shifting the images' height and width, and flipping them both horizontally and vertically. After augmentation, the pixel values were normalized to ensure they all fell within the range of 0 to 1, which is crucial for efficient model training. The model consisted of four sets

of convolutional and max pooling layers. Each convolutional layer applies a filter to the input image, transforming it to extract features while reducing the computational load and the number of learnable parameters. The ReLU activation function was applied to introduce nonlinearity and help mitigate the vanishing gradient problem. Padding was set to 'SAME' to retain the input size after convolution. After the convolutional layers, three dense layers were used to classify the images into seven types of skin cancer, with a decreasing number of neurons from 64 to 32 to 7 in the final output layer. ReLU was used for the first two dense layers to provide nonlinearity, and the Softmax activation function was applied to the output layer to yield a probability distribution over the seven classes. The Adam optimizer was chosen for its effectiveness in updating network weights iteratively. EarlyStopping was employed as a training strategy to prevent overfitting. It allowed for specifying a large number of epochs and stopping the training when the model's performance ceased to improve on a validation dataset. K-fold cross-validation was utilized to ensure that the model's performance was not based on a specific train-test split. This technique divided the dataset into 'K' subsets (folds) and performed training and validation 8 times, with each fold used once as the validation set.

VIII. Model Evaluation and Result

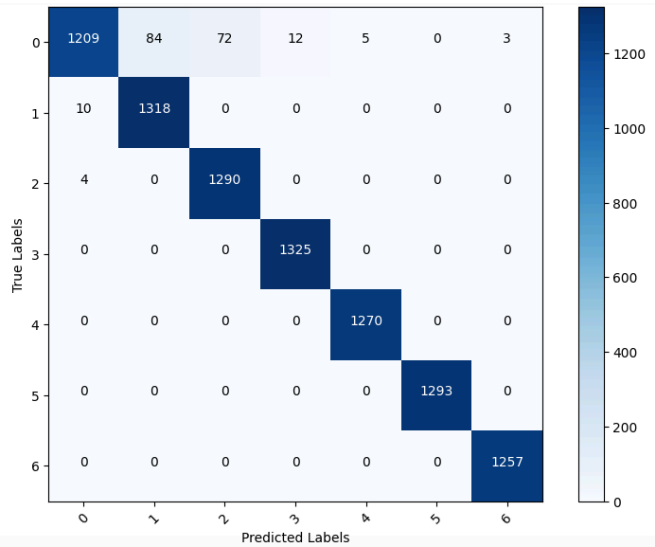
After training and evaluation, the model's predictive capability was put to the test. The prediction phase is the ultimate test of how well the model generalizes to new, unseen data. The trained model was used to predict the skin cancer classes of the images in the test set. This was done using the predict method, which outputs the probability of each class for each image. The raw probabilities obtained from the model were post-processed to yield discrete class predictions. In the case of the softmax output layer, the class with the highest probability was selected as the model's



prediction. For individual predictions, such as the case where one might want to predict the class of a single image (e.g., during deployment in a clinical setting), the model can be fed one image at a time, and it will provide the likelihood of that image belonging to each of the seven different skin cancer classes. The accuracy of these predictions was quantified by comparing the predicted classes against the ground truth labels. The model's precision is a testament to its performance, with an accuracy score of approximately 97.91%, indicating a high level of reliability.

Confusion Matrix:

The evaluation of the model's performance was conducted using a multiclass confusion matrix. This matrix helped to visualize the model's predictions against the true labels, providing



insights into the classification accuracy for each

class. The values along the matrix's diagonal (from the top left to the bottom right) represent correct predictions where the predicted classes match the true classes.

For example, we can observe that the model correctly identified 1209 instances of class '0', 1318 instances of class '1', and so forth. These high diagonal values suggest a strong predictive performance for these classes. The off-diagonal elements indicate where the model has made errors: Class 0 was most commonly confused with Class 1 (84 instances) and Class 2 (72 instances). Misclassifications for other classes are relatively low, suggesting that the model has a strong ability to discriminate between most classes. The color gradient in the confusion matrix visualizes the magnitude of the values, with darker colors representing higher numbers. This visual aid emphasizes the model's accuracy in predicting each class by highlighting the darker shades along the diagonal.

Precision: Indicates the accuracy of positive predictions for each class. The model demonstrates high precision across all classes, with perfect scores for several lesion types, which means there were few false positives.

Recall: Measures the model's ability to detect all actual positives for each class. The recall is also high for all classes, suggesting that the model successfully identified most of the positive cases.

```
Test Accuracy: 97.913%
286/286 [==============================] - 1s 4ms/step
              precision    recall  f1-score   support

          nv       0.99      0.87      0.93      1385
         mel       0.93      0.99      0.96      1328
         bkl       0.95      1.00      0.97      1294
         bcc       0.99      1.00      0.99      1325
       akiec       1.00      1.00      1.00      1270
        vasc       1.00      1.00      1.00      1293
          df       1.00      1.00      1.00      1257

    accuracy                           0.98      9152
   macro avg       0.98      0.98      0.98      9152
weighted avg       0.98      0.98      0.98      9152
```

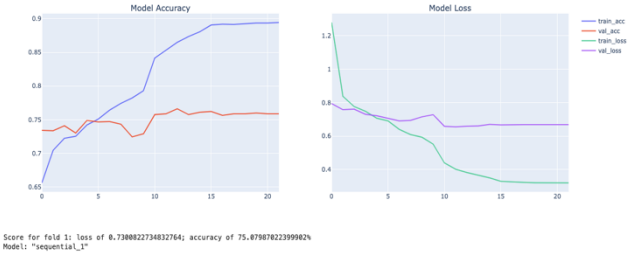
F1-Score: Harmonic mean of precision and recall, which provides a single metric for each class that considers both false positives and false negatives. The model's F1-scores are exemplary, with several classes achieving perfect scores.

Overall Accuracy: The model achieved an overall test accuracy of 97.913%, confirming its exceptional ability to classify skin lesions correctly.

Training Process Graphs:

The model's training process is depicted in two graphs showing accuracy and loss over epochs:

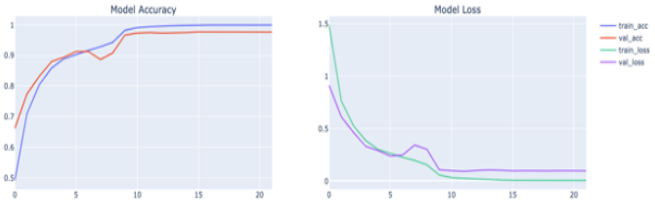
Model Accuracy Graph: Shows the training and validation accuracy over time. The model quickly reached high accuracy levels, which plateaued, indicating that additional training epochs did not lead to overfitting.



Score for fold 1: loss of 0.7300822734832764; accuracy of 75.07987022399902%
Model: "sequential_1"

Model Loss Graph: Displays the training and validation loss over time. There is a sharp decrease in loss initially, leveling off as the model converges, which is indicative of good learning dynamics.

Both graphs exhibit the characteristics of a well-



trained model that generalizes well without overfitting, as evidenced by the close performance of training and validation metrics.

The CNN model achieved a remarkable accuracy score of approximately 97.91%, which was significantly higher. This success underscores the effectiveness of deep learning, particularly CNNs, in image classification tasks such as skin cancer diagnosis from dermatoscopic images.

IX. Conclusion

Our project has successfully harnessed the power of Convolutional Neural Networks (CNNs) to tackle the complex and crucial task of skin cancer image classification. The customized model architecture, specifically designed for image analysis, has demonstrated promising capabilities in accurately identifying different types of skin cancer. A key factor in this achievement was the strategic use of data augmentation techniques, which allowed us to effectively counter the limitations posed by an imbalanced dataset. This approach not only enriched our dataset but also equipped our model with robust generalization abilities, enabling it to learn from a diverse range of image presentations. Through meticulous evaluation methodologies, including a detailed analysis with a confusion matrix and the application of K-fold cross-validation, we have established the model's dependable performance. The notable metrics—accuracy, precision, recall, and F1-score—underscore the high level of correctness in the model's predictions, reinforcing the potential of CNNs in medical imaging.

Areas for Future Improvement

Despite the model's impressive performance, the pursuit of excellence in medical diagnostics demands continual advancement. Future research could delve into exploring deeper neural network architectures to capture more nuanced patterns within the data. Investigating more advanced data augmentation techniques could further bolster the model's resilience against overfitting and improve its performance on underrepresented classes. The integration of additional data sources, such as patient history or genetic information, may provide complementary insights that enhance the model's diagnostic power. Additionally, the development of transfer learning strategies could streamline the training process and enable the adaptation of pre-trained models to this specific domain, potentially leading to improvements in accuracy and efficiency. As we look ahead, the integration of CNNs in clinical settings seems not only feasible but imminent. With continued refinement and validation, such models hold the promise of significantly aiding dermatologists in early detection and treatment planning, ultimately improving patient outcomes in the fight against skin cancer

X. References

1. https://www.kaggle.com/code/ashutoshvarma/ image-classification-using-svm-92-accuracy/notebook
2. https://www.nature.com/articles/sdata2018161
3. https://www.kaggle.com/code/hozifanasef/ ham10000-cnn-skin-cancer-type-detection
4. https://github.com/mkowalsky97/Skin_Cancer_Detection/blob/main/Main.ipynb

XI. Team Contribution

1. Kislay - Led model architecture design, contributed to the writing of the Implementation section, and managed the project repository. Managed training experiments, contributed to the Results and Experiments sections, and prepared the final presentation and Report.

2. Salman - Focused on data preprocessing, contributed to the Introduction and Data sections, and assisted with model evaluation. Led model evaluation and result, Worked on conclusion and area of improvement section and collection references.