SKIN CANCER CLASSIFICATION FINAL REPORT

Kiarash Alirezaei
Student# 1008887215
kiarash.alirezaei@mail.utoronto.ca

Sepehr Rajabian
Student# 1008976440
sep.rajabian@mail.utoronto.ca

Koosha Omidian
Student# 1009315304
koosha.omidian@mail.utoronto.ca

Erfan Nazarian
Student# 1009042064
erfan.nazarian@mail.utoronto.ca

1 INTRODUCTION

Skin cancer represents one in every three cancer diagnoses worldwide, making it a significant global health concern (Canadian Skin Cancer Foundation, n.d.). The disease arises from the abnormal growth of skin cells damaged by agents such as ultraviolet (UV) radiation, leading to skin lesions that differ from the surrounding skin (Cleveland Clinic, 2020). Early and accurate diagnosis is crucial, as survival rates are significantly higher when cancer is identified early and correctly classified for appropriate treatment (Medical News Today, 2023). Currently, dermatologists rely on visual examination and professional judgment to diagnose skin cancer, a method that is both subjective and error prone (Mayo Clinic, 2022). This highlights the gap for a more systematic and accurate diagnostic tool. Deep learning (DL) models, particularly convolutional neural networks (CNNs), have emerged as powerful methods for image classification tasks. CNNs have demonstrated remarkable success in medical imaging, including skin cancer classification from dermoscopic images (Mohakud & Dash, 2023). These models can learn from vast datasets and identify subtle patterns that might be missed by human observers, thereby reducing diagnostic errors and minimizing subjectivity. Unlike traditional machine learning (ML) methods that require manual feature extraction and are often tailored to specific patient data, DL models such as CNNs automatically learn and extract features from raw data. This automatic feature extraction provides a more generalizable approach, enhancing the model's robustness and effectiveness across diverse patient populations. This paper presents a CNN-based model designed to enhance the accuracy and reliability of skin cancer classification. Thus, as illustrated in Figure 1, the model's input consists of dermoscopic images of different skin cancers, and the output is a prediction of the cancer class. Implementing such a model using solely heuristics and non-ML approaches would require developing exhaustive rules to cover all possible variations in patients and skin cancer types, making it near impossible. Finding a relationship between the inputs (skin cancer images) and their respective skin cancer types (ground truth labels) is a classic ML application. Therefore, ML was chosen as a suitable approach to generate such outputs from the inputs as it can create the rules and mappings required to accurately classify skin cancer types from dermoscopic images.

2 ILLUSTRATION / FIGURE

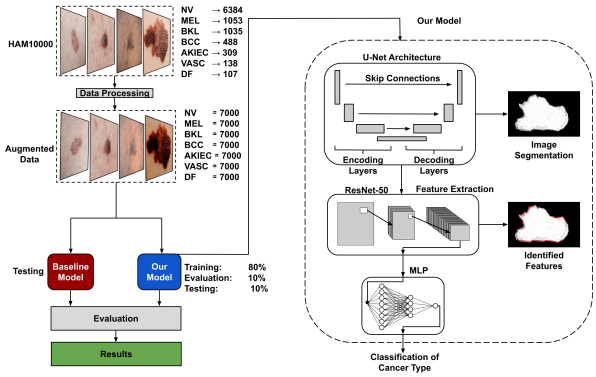


Figure 1: Overview of our model

3 Background and Related Work

In examining current industry standards for applying DL to medical imaging, particularly in skin cancer diagnosis, the team reviewed various articles detailing different approaches. These articles, which are summarized below, informed our approach by highlighting various implementations that yielded the best results.

3.1 Article One

In the study "Skin Cancer Detection using CNN with Swish Activation Function," the researchers demonstrate that a CNN model using the Swish activation function outperforms others in detecting skin cancer and its types (Fathima et al., 2020). In general, "CNN has proved to output tremendous results in ImageNet Challenge. This is considered a significant segmentation and classification challenge in the image analysis field" (Fathima et al., 2020). As stated, CNNs are very effective in medical image classification, thus researchers in this paper attempt to further enhance the CNN architecture by integrating Swish activation, which solves the "dying ReLU" problem where neurons become inactive and stop learning during training. To evaluate the effectiveness of the Swish activation function, the authors compared it with the Adaptive Piecewise Linear (APL) activation function on identical 21 layer CNN architectures. They used the Skin Cancer MNIST: HAM10000 dataset for training their models, which was augmented from 10,000 images to 38,000 images for a better training process. The models were trained for 50 epochs, and their performance was evaluated on a validation set of 1,000 images. The Swish-activated CNN had an accuracy of 76% on the validation set (they used a set of 1,000 images for validation), while the APL achieved 74.4% accuracy.

3.2 Article Two

The study "CNN-based Diagnosis System on Skin Cancer using Ensemble Method Weighted by Cubic Precision" explored various high-performance CNN models, including ResNet50, ResNeXt50, ResNeXt101, EfficientNetB4, (Jiang, 2021). The researchers first augmented the data to balance the number of samples for each skin cancer type, preventing model bias. Each model was optimized through various techniques, such as creating skip connections to address the vanishing gradient problem, applying aggregated transformations to enhance accuracy without additional hyperparameters, and using compound model scaling for uniform dimension scaling and performance improvements. Through these modifications, MobileNetV3 achieved the highest overall accuracy of 91.02%. This model utilized depthwise convolution, replaced non-linear activation functions with linear bottlenecks to prevent information loss, and balanced feature extraction with inverted residuals. Additionally, it introduced the "h-swish" activation function and an attention mechanism based on squeeze and excitation to further enhance performance.

3.3 Article Three

In the paper "A precise model for skin cancer diagnosis using hybrid U-Net and improved MobileNet-V3 with hyperparameters optimization" (Lilhore et al., 2024), the authors propose a novel approach to skin cancer diagnosis. Researchers applied various data augmentation techniques, including rotations, width and height shifting, shearing, flipping, and brightness adjustments, to enhance dataset diversity and combat overfitting. Additionally, images are resized from 450 x 600 pixels to 192x256 pixels. The model architecture combines a standard U-Net for image segmentation with an improved MobileNet-V3 for feature extraction. Bayesian optimization is utilized for hyperparameter tuning to achieve optimal performance. Transfer learning is employed by pre-training MobileNet-V3 on the ImageNet dataset, followed by fine-tuning it on the HAM10000 dataset. The proposed model achieves impressive results, with precision, sensitivity, accuracy, and specificity metrics of 97.84%, 96.35%, 98.86%, and 97.32%, respectively, outperforming existing DL-based skin cancer detection solutions.

3.4 Article Four

In the article "A multi-class skin Cancer classification using deep convolutional neural networks," (Chaturvedi et al., 2020) researchers employed a comprehensive approach that included preprocess-

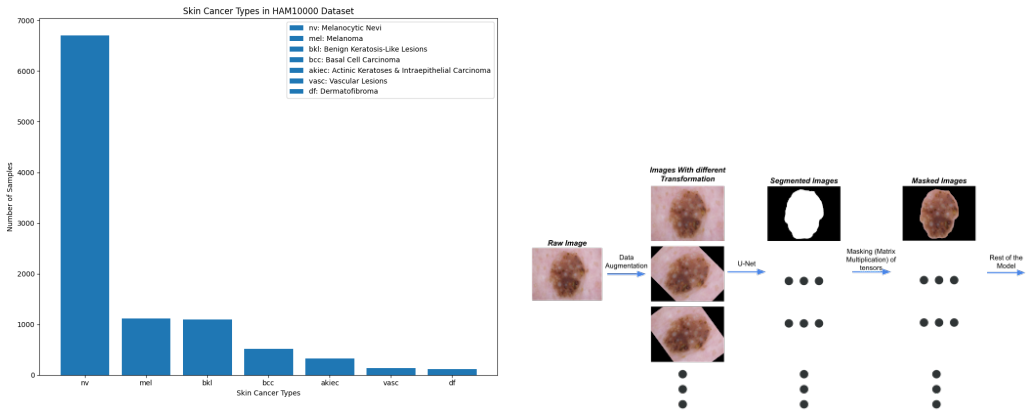
ing, classification model implementation, fine-tuning, and feature extraction. Preprocessing was minimal, utilizing the Keras ImageDataGenerator function. The models explored included Xception, InceptionV3, InceptionResNetV2, ResNeXt101, and NASNetLarge. The researchers tested both individual models and various combinations of these models. Feature extraction was performed using an Integrated Feature Extractor after fine-tuning the models. The accuracy results for the individual models were: InceptionV3 at 91.56%, ResNetXt101 at 93.20%, InceptionResNetV2 at 93.20%, Xception at 91.47%, and NASNetLarge at 91.11%. Among the combined models, the highest accuracy was achieved by 'InceptionResNetV2 + ResNeXt101' at 92.83%, followed by 'InceptionV3 + Xception' at 91.56%, and 'InceptionResNetV2 + ResNeXt101 + Xception' at 89.66%. These combination models surpassed the performance of human dermatologists and many existing DL methods for skin cancer classification. Notably, some combination models performed worse than the well-tuned standalone models, highlighting that effective tuning of individual models can sometimes be more beneficial than combining them.

3.5 Article Five

The paper "Skin Cancer Detection Using Dual Optimization Based Deep Learning Network" presents a sophisticated approach to skin cancer detection using advanced DL techniques. The authors utilize the MNIST HAM10000 dataset, applying adaptive median filtering for noise removal and a U-Net architecture for segmenting skin lesions. The key innovation of the study is the combination of Bacterial Foraging Optimization (BFO) and Particle Swarm Optimization (PSO) for feature extraction, enhancing the model's image thresholding capabilities. A deep CNN then classifies the seven types of skin cancer, achieving an impressive accuracy of 98.76%. This method outperforms traditional classifiers like SVM, KNN, and Random Forest, demonstrating the effectiveness of integrating evolutionary algorithms with DL for accurate and reliable medical image analysis. This work offers valuable benchmarks and methodological insights for our own project (Gomathi et al., 2023).

4 Data Processing

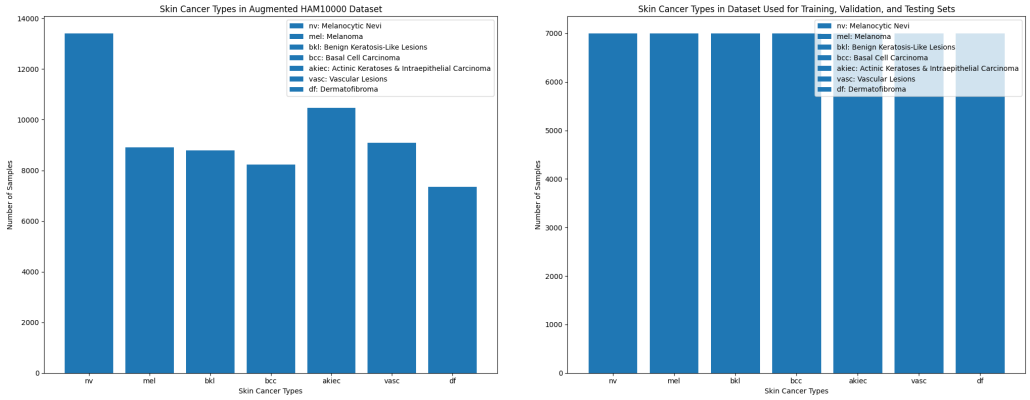
The data collected and cleaned for this project is from the HAM10000: Skin Cancer MNIST dataset (Zairza-CETB, n.d.), which comprises 10,015 dermoscopic images of seven different skin cancer types. Code was developed to extract the images and consolidate them into a main image directory. Additionally, the initial dataset contained a metadata CSV file detailing the cancer type for each image, facilitating the analysis of the dataset's composition. To visualize this distribution, the matplotlib library was utilized to create the bar graph shown in 2a. The graph reveals an imbalance in image distribution among skin cancer types, with the benign melanocytic nevi class (nv) having 6,384 images, while other classes peak at just over 1,000 images.



(a) (b)

Figure 2: Figure (a): Distribution of images across different skin cancer types in the HAM10000 dataset. Figure (b): data processing model for handling this distribution.

To address this imbalance and enhance dataset diversity, various augmentation techniques were applied. The dataset was augmented to equalize the image counts across different skin cancer types for model training, validation, and testing. Techniques such as color jittering, color affines, Gaussian blurs, and random horizontal flips were applied with multiple distinct configurations via the *torch.transforms* library. For cancer types with fewer initial samples, such as dermatofibroma (df), up to six transformations were cumulatively applied, each doubling the number of available images. Conversely, cancer types with many raw images, such as Melanocytic nevi (nv), underwent only 1-2 transformations to prevent disproportionate image counts. Data augmentation was crucial for enabling the models to learn to segment and abstract lesion types without overreliance on image-specific features such as color, rotation, or background characteristics (e.g., skin tone or hair). The resulting distribution of categories after augmentation is shown in 3a: After data augmentation, each



(a) Bar Graph 1

(b) Bar Graph 2

Figure 3: Bar graph (a) illustrating the distribution of images across various skin cancer types after performing augmentation to the HAM10000 dataset and bar graph (b) showing distribution of images across different skin cancer types in the main dataset after data augmentation.

cancer type contained over 7,000 samples, and in order to ensure dataset homogeneity and balance, only 7,000 images per cancer type were selected for the main dataset used in training, validation, and testing. The composition of this dataset is illustrated in Figure 3b. In consideration of the ethical implications of developing a skin cancer classification model, the team indexed image metadata to retrieve patient demographics. 4 displays the sex distribution among patients. This approach enabled the team to evaluate potential disparities in model performance between genders. For the

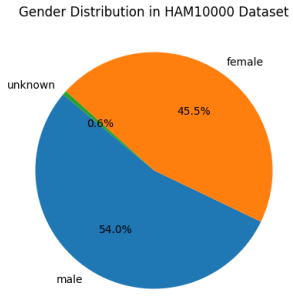


Figure 4: Distribution of gender in the HAM10000 dataset.

baseline SVM model, the dataset was split into training and testing sets via *sklearn's* built-in train test split function. In contrast, when training the U-Net model, which is used for image segmentation, the dataset was split into training (80%), validation (10%), and testing (10%) sets. This split emphasizes training to enhance the model's segmentation learning. Given the use of a pre-existing segmentation dataset, extensive validation and testing were unnecessary. The same split ratio was

applied to train, validate, and test the MLP classifier. For the final testing of the model, new data from the Skin cancer ISIC: The International Skin Imaging Collaboration dataset Katanskiy (2019) was used. This dataset contains 2,160 dermoscopic images of the same skin cancer types the model was trained on. Since the model had not previously encountered images from this dataset, this evaluation assessed the model's ability to generalize beyond the training data, simulating real-world scenarios where the model must classify unseen samples.

5 ARCHITECTURE

The team's model consists of three main components: image segmentation via a U-Net, consisting of a convolutional neural network and transposed convolutional neural network. Next, A feature extractor via transfer learning, consisting of a pure Convolutional Neural Network (CNN) from the ResNet-50 model. Finally, a MLP classifier, consisting of fully connected linear layers. For a detailed illustration of the model, layer by layer, Refer to Figure 5. The U-Net is a unique CNN autoencoder implementation typically used in image segmentation applications. This network comprises a contracting path (encoder) with three convolutional layers, a symmetric expanding path (decoder), and a bottleneck connecting the two, creating the "U" shape it is named after. The encoder is responsible for capturing various features of the image. The decoder then up-samples the features while also using skip connections from the encoder to the decoder to preserve spatial information. This helps reconstruct segmented images with precise object boundaries. Normally, the U-Net model was designed to process the entire dataset in a single pass; however, due to computational constraints, we adapted it to train on smaller subsets of 500 images. The model was trained for 20 epochs on each subset, with the best-performing model saved after each iteration. This approach achieved high accuracy, successfully segmenting almost 66,000 images. After processing the images with the U-Net, binary masks are generated, which are then applied to the original images to isolate the cancerous areas. These masked images are subsequently fed into a modified ResNet-50 architecture. ResNet-50 is a deep CNN that utilizes residual blocks to learn complex patterns by employing skip connections, which address the vanishing gradient problem. In our approach, we modify ResNet-50 by removing the final fully connected layer, transforming it from a classifier into a feature extractor that captures high-level representations of the masked images. These features are subsequently processed through our optimized MLP classifier. Following the feature extraction process, the low-dimensional representations of the images are flattened and passed through the classifier component. To optimize the classifier, we employed a random search for hyperparameter tuning, creating and validating over 100 MLP architectures featuring a range of values for the number of layers, neurons, and learning rate. Our final model features a classifier component with 6 layers. Figure 5 provides a detailed illustration of this final classifier.

6 BASELINE MODEL

For our baseline model, we utilized an SVM classifier with a radial basis function (RBF) kernel, leveraging its capability to handle non-linear relationships inherent in image data. The datasets mentioned in the Data Processing section were used for training and testing the SVM classifier. After fitting the model to the training set, predictions were then made on the test set. The model's performance was evaluated by comparing these predictions with the ground truth labels using sklearn's built-in evaluation functions: accuracy score, precision score, recall score, and F1 score. The resulting scores were 0.57 for accuracy, 0.60 for precision, 0.58 for recall, and 0.58 for F1 score. These metrics and observations provide a basis for comparison with the performance of other models.A notable qualitative observation from the baseline model's performance is its tendency to misclassify samples between Melanoma (mel) and Benign Keratosis-Like Lesions (bkl).

This baseline SVM model was implemented using the following process: we firstly flattened the pixel values into a single vector per image using PyTorch's stack() and view() methodsm we then used the scikit-learn library to import the SVM classifier. Next, we applied the radial basis function (RBF) kernel, effective for capturing non-linear relationships in image data. After that, we trained the SVM Model: Fit the SVM on the training set. Finally, we evaluated the model for accuracy, precision, recall, and F1-score.

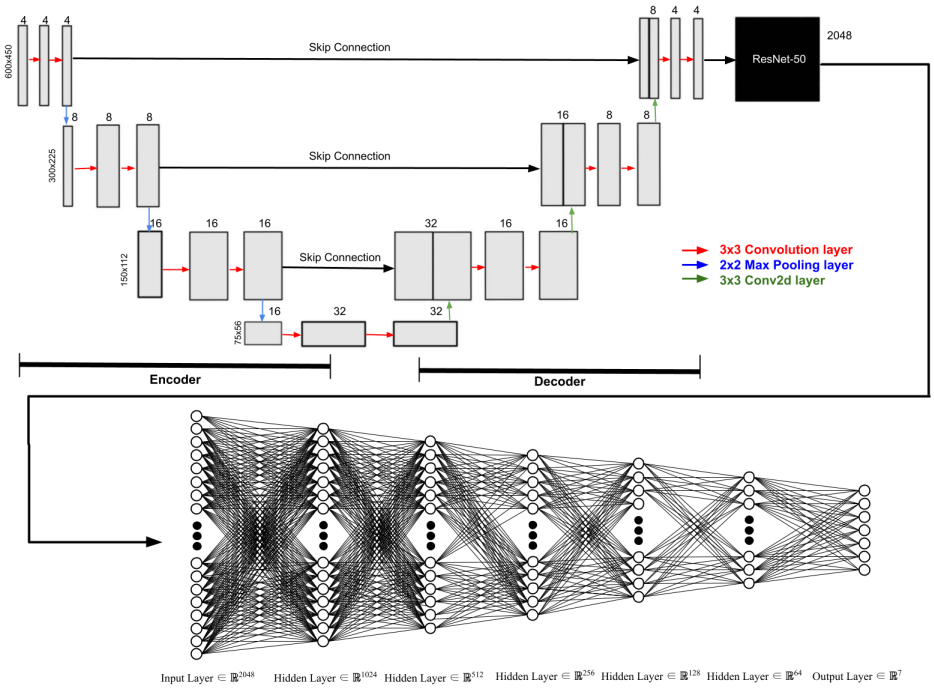


Figure 5: Overview diagram of model with detailed information

7 Quantitative Results

7.1 Overall Results

The model's overall performance on the test dataset achieved an average precision of 78.4%, recall of 87.1%, F1-Score of 0.825, and accuracy of 85.4%. Since the dataset was balanced, accuracy is a key measure here. In biomedical DL applications, high recall (indicating fewer false negatives) is particularly important because it minimizes the risk of missing severe cases. Although precision is relatively lower, which means there are more false positives, this trade-off is acceptable in this context. It is preferable for the model to classify less severe cases as more severe, as this approach helps ensure that potentially serious conditions are not overlooked and prompts timely medical examination and treatment.

7.2 U-Net

When tuning the U-Net, the team opted for manual, iterative parameter adjustments due to time and resource constraints. This approach led to gradual improvements in the training process. By fine-tuning the hyperparameters, we observed substantial improvements in the model's performance across different data subsets of 500 images. Initially, as shown in Figure 6, the model was unstable, frequently overfitting to the specific subset it was trained on; resulting in poor generalization as moving to new subsets caused large spikes in loss/accuracy. However, after several iterations of parameter tuning, the final U-Net model demonstrated improved generalization when transitioning between data subsets, as illustrated in Figure 7. This iterative approach also helped mitigate overfitting on any particular data subset, as new data was regularly incorporated for training and validation.

7.3 ResNet-50 and MLP

To integrate the MLP classifier with ResNet-50, the team began by applying the ResNet-50 feature extraction layers to the masked augmented images generated by our U-Net model. Subsequently, these features and their corresponding labels were split into training (80%), validation (10%), and

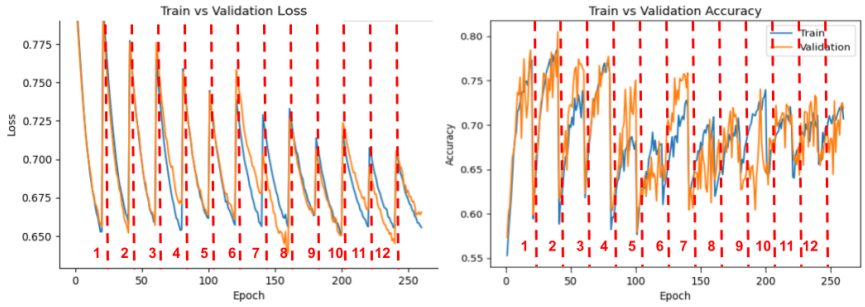


Figure 6: Initial model behaviour across data subsets (red lines indicate dataset used up to that point)

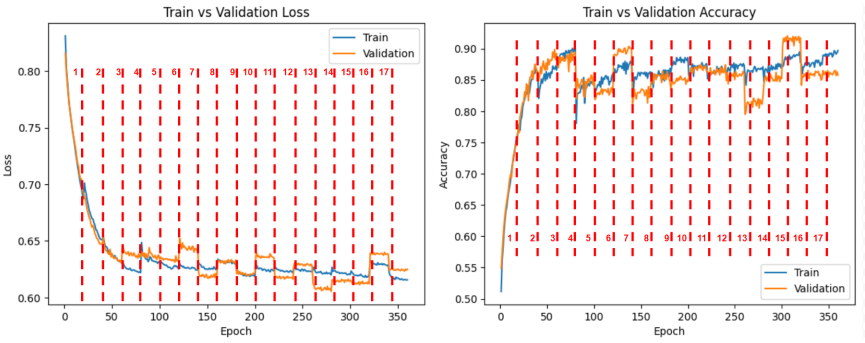


Figure 7: Loss and accuracy behavior of the model after iterative tuning

testing (10%) subsets. Afterwards, the team utilized a random search for MLP classifier hyperparameter tuning. After two iterations of random search, the team narrowed the hyperparameter space and arrived at a 6-layer NN (as seen at the bottom of Figure 5 which was further trained for 1000 epochs. The training and validation curves can be observed in Figure 8.

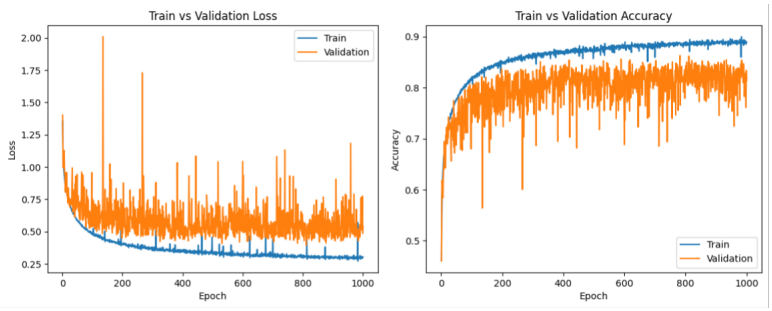


Figure 8: Training and validation progress of MLP classifier over epochs

8 Qualitative Results

Analyzing the graphical results of the U-Net assists in contextualizing the statistical performance of the model. The U-Net was trained on a set of ground truth masks for each HAM10000 image and was tasked with segmenting the images accordingly.

As illustrated in part a) and b) of Figure 9, the model effectively learned to black out the irrelevant areas of the skin while highlighting the cancerous areas white. The segmented images are then used as a mask for the original image producing a masked image that you see in part c) of Figure 9

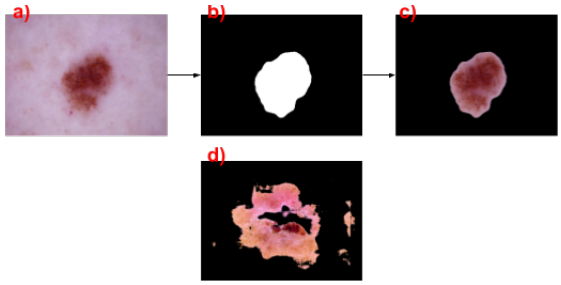


Figure 9: Example of a lesion going through a segmentation and masking stage

This U-Net and masking process was crucial for accurately isolating and highlighting the cancerous regions within the images, allowing our subsequent CNN model to focus on the most relevant features for classifying the eight different skin cancer types.

Figure 9 d) demonstrates a less optimal example of this output. While the U-Net successfully segmented major components of the skin lesion, there are noticeable noise and inconsistencies in the segmentation. For instance, the area in the center of the lesion appears to be cut out, and there is arbitrary segmentation along the edges of the image. These graphical observations provide insight into the quantitative results discussed in section 7. Specifically, the segmentation of edges and the omission of central lesion areas could negatively impact the performance of subsequent components, particularly the ResNet-50 feature extraction.

9 Evaluating the model on new data

To thoroughly evaluate our principal model after training the individual components, the team utilized both the ISIC: The International Skin Imaging Collaboration dataset (Katanskiy, 2019) and the HAM10000 test dataset the team created. The images from the ISIC dataset were all categorized into the 7 categories our model was designed to predict and resized to meet the model's input requirements. Subsequently, these images were put into a data loader and used as an evaluation metric not used for hyperparameter tuning. This analysis provided insights to the model's biasing, with the most successful diagnosis being of type nevus and least successful being of type being dermatofibromas (often confused for nevus) which aligns with the original data distribution of the HAM10000. Similarly, we evaluated the model on the HAM10000 test dataset, which contained data the model had never been exposed to. This evaluation achieved an accuracy of 85.4% and a loss of 0.439, closely matching our validation results. Our model outperformed several state-of-the-art approaches. For example, it exceeded the accuracy of the best model in "Skin Cancer Detection using CNN with Swish Activation Function" (Fathima et al., 2020) by 9.4%. Additionally, it attained average precision, recall, and F1-Score of 0.784, 0.871, and 0.825, respectively. Compared to our baseline model, our DL approach improved average precision, recall, F1-Score, and accuracy by 18.4%, 29.1%, 0.245, and 28.4% respectively. This further reinforces the superior performance of our model across all metrics on the same test dataset.

10 Discussion

Reflecting on our model's implementation, despite Google Colab's constraints, our model exceeded expectations by achieving an accuracy upwards of 80% across the HAM10000 dataset, demonstrating performance on par with previous research. The limitations the team encountered during the creation of the segmentation portion, feature extraction, and classification phases necessitated innovation and adaptation of course concepts. For example, faced with RAM limitations (e.g. 12.5 GB) while training the U-Net, we addressed this challenge by training the model on data subsets. This approach not only circumvented memory issues but also helped combat overfitting through frequent exposure to new data. To effectively extract meaningful features from the masked images, the team employed transfer learning using the feature extraction layers of the ResNet-50 model. Additionally, the team applied random search for hyperparameter tuning of the MLP. Overall, our model demonstrated consistent performance throughout the training process and showed remarkable

learning capabilities, especially given the size of the dataset and multiple compounded augmentations. This interesting observation suggests that our model's learning could potentially be extended to other types of skin cancer in the future, since the results across various evaluations exhibited minimal biases. In summary, the implementation of this model has provided the team with valuable insights and skills in developing innovative and efficient DL models, highlighting our growth and the team's ability to tackle complex ML problems.

11 Ethical Considerations

When classifying skin cancer varieties using the HAM10000: Skin Cancer MNIST dataset, we have identified several potential ethical limitations associated with the model trained on this data.

11.1 Model Misdiagnosis Risk

It is crucial to consider the ethical implications of utilizing our skin cancer classifier model in practice. The dataset is specifically composed of skin cancer samples and is relatively devoid of other skin lesions or clean skin samples (with the exception of benign melanocytic nevi). This means that the model might mistakenly classify non-cancerous skin conditions, such as morphea, as innately cancerous. A misdiagnosis presents risks to patients, including unnecessary physical, emotional, and financial burdens. As a result, it is ethically imperative to ensure that the model is only used for detecting a type of skin cancer, meaning a healthcare professional must first confidently identify the skin lesion as cancerous. This demonstrates the importance of clear communication in the intended application of our model.

11.2 Ethnic Diversity Concerns in the Data

The dataset lacks information regarding the ethnic distribution of the subjects. This absence presents a significant risk of inaccurate classification for underrepresented ethnic groups, such as African-Americans, who often suffer from higher morbidity and mortality as a result of their less common skin cancer prevalence (Gloster Jr. & Neal, 2006). Without proper representation of various ethnicities, the model might fail to generalize effectively, leading to potential misdiagnoses and disparities in healthcare outcomes.

11.3 Gender Imbalances in the Data

There are known structural differences between men's skin and women's skin. Androgen (testosterone) stimulation causes a roughly 25% increase in skin thickness in men (der). Our previous data analysis showed that HAM10000 has a 45% to 55% distribution of females to males. This imbalance may lead to a model that more accurately predicts skin cancer in males. Additionally, the dataset does not include concrete information about non-binary individuals, whose data is not adequately represented. Consequently, the model's predictive capability for non-binary individuals is unknown and potentially invalid. This underrepresentation may negatively impact non-binary individuals, highlighting the importance of a more inclusive dataset.

12 Project Quality

The complexity of skin cancer classification necessitates the careful application of machine learning (ML) techniques to achieve reliable results. Previous studies have demonstrated the effectiveness of transfer learning with pre-trained deep network architectures, achieving accuracies averaging around 91%, as seen in the most successful models outlined in Section 3. These studies reveal that simpler CNNs and MLPs often fall short in performance. In contrast, our approach integrates a more comprehensive and challenging model. It combines an untrained U-Net for segmentation with a customized ResNet-50 for feature extraction and an MLP for classification. Additionally, we implemented innovative data augmentation techniques to enhance the model's robustness. This approach required a thorough understanding of model architecture and training processes, particularly when customizing the ResNet-50. Despite the complexity of generalizing across multiple skin cancer types with subtle visual differences, our model achieved impressive results.

REFERENCES

Is a man's skin really different from a woman's? URL https://www.dermalogica.com/blogs/living-skin/is-a-mans-skin-really-different-from-a-womans.

Canadian Skin Cancer Foundation. Skin Cancer, n.d. URL https://www.canadianskincancerfoundation.com/skin-cancer/#:~:text=Other%20Skin%20Cancer%20Facts&text=One%20in%20every%20three%20cancers,deadliest%20form%20of%20skin%20cancer.

S. S. Chaturvedi, J. V. Tembhurne, and T. Diwan. A multi-class skin cancer classification using deep convolutional neural networks. *Multimed Tools Appl*, 79:28477–28498, 2020. doi: 10.1007/s11042-020-09388-2. URL https://doi.org/10.1007/s11042-020-09388-2.

Cleveland Clinic. Skin Lesions, 2020. URL https://my.clevelandclinic.org/health/diseases/24296-skin-lesions#:~:text=What%20are%20skin%20lesions%3F,like%20infections%20or%20autoimmune%20diseases.

Mariam Fathima, Misba Farheen, M. Manjushree, and Manish Kumar Pandit. Skin cancer detection using cnn with swish activation function. *International Journal of Engineering Research & Technology (IJERT)*, 8, 2020. ISSN 2278-0181. URL https://www.ijert.org. NCETESFT - 2020 Conference Proceedings, Special Issue - 2020.

Hugh M. Gloster Jr. and Kenneth Neal. Skin cancer in skin of color. *Journal of the American Academy of Dermatology*, 55(5):741–760, November 2006. Available online 17 October 2006.

E. Gomathi, M. Jayasheela, M. Thamarai, and M. Geetha. Skin cancer detection using dual optimization based deep learning network. *Biomedical Signal Processing and Control*, 84: 104968, 2023. ISSN 1746-8094. doi: https://doi.org/10.1016/j.bspc.2023.104968. URL https://www.sciencedirect.com/science/article/pii/S1746809423004019.

P. Jiang. Cnn-based diagnosis system on skin cancer using ensemble method weighted by cubic precision. In *2021 2nd International Seminar on Artificial Intelligence, Networking and Information Technology (AINIT)*, pp. 145–152, 2021. doi: 10.1109/AINIT54228.2021.00038.

A. Katanskiy. Skin cancer isic, August 2019. URL https://www.kaggle.com/datasets/nodoubttome/skin-cancer9-classesisic/data. Accessed: 2024-07-04.

Utkarsh Kumar Lilhore, Suraj Simaiya, Yogendra Kumar Sharma, and et al. A precise model for skin cancer diagnosis using hybrid u-net and improved mobilenet-v3 with hyperparameters optimization. *Sci Rep*, 14:4299, 2024.

Mayo Clinic. Skin Cancer: Diagnosis & Treatment. *Mayo Clinic*, 2022. URL https://www.mayoclinic.org/diseases-conditions/skin-cancer/diagnosis-treatment/drc-20377608.

Medical News Today. Is Skin Cancer Deadly? *Medical News Today*, 2023. URL https://www.medicalnewstoday.com/articles/is-skin-cancer-deadly#:~:text=Skin%20cancer%20can%20be%20deadly,dangerous%20type%20of%20skin%20cancer.

R. Mohakud and R. Dash. A hybrid model for classification of skin cancer images after segmentation. *International Journal of Image and Graphics*, 31(2550022), 2023.

Zairza-CETB. HAM10000 - Human Against Machine with 10000 training images. GitHub repository, n.d. URL https://github.com/zairza-cetb/HAM10000.