Efficient Deep Learning for Skin Lesion Segmentation Using Convolutional Block Attention Mechanism

by

Rejwan Shafi
23241108
Homairah Ferdousia
24141198
Mashrura Ahmed
21301378
Mostakim Mahmud Mugdhha
24141249
Irfanul Islam
22101479

A thesis submitted to the Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
B.Sc. in Computer Science

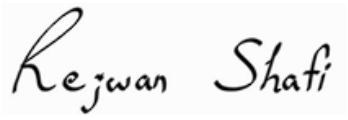
Department of Computer Science and Engineering
BRAC University
June 2025

Declaration

It is hereby declared that

1. The thesis submitted is our own original work while completing the degree at BRAC University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material that has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. We have acknowledged all main sources of help.

Student's Full Name & Signature:

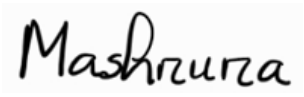


Rejwan Shafi

23241108

Homairah Ferdousia

24141198



Mashrura Ahmed

21301378

Mostakim Mahmud Mugdhha

24141249

Irfanul Islam

22101479

Approval

The thesis titled "Efficient Deep Learning for Skin Lesion Segmentation Using Convolutional Block Attention Mechanism" submitted by

1. Rejwan Shafi (23241108)
2. Homairah Ferdousia (24141198)
3. Mashrura Ahmed (21301378)
4. Mostakim Mahmud Mugdhha (24141249)
5. Irfanul Islam (22101479)

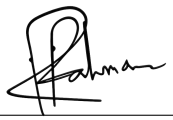
of Summer, 2025 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science in June 20, 2025.

Examining Committee:

Supervisor:

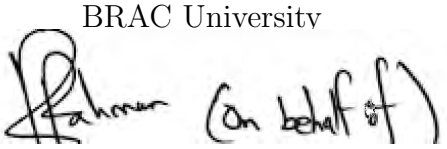
Dr. Muhammad Iqbal Hossain
Associate Professor
Department of Computer Science and Engineering
BRAC University

Secondary Supervisor:



Mr. Rafeed Rahman
Senior Lecturer
Department of Computer Science and Engineering
BRAC University

Program Coordinator:



Dr. Md. Golam Rabiul Alam
Professor
Department of Computer Science and Engineering
BRAC University

Head of Department:

Dr. Sadia Hamid Kazi
Associate Professor
Department of Computer Science and Engineering
BRAC University

Abstract

Skin cancer, particularly melanoma, poses a significant worldwide medical challenge as a result of growing incidence rates and limited awareness until advanced stages. This issue is exacerbated in developing countries, where a shortage of dermatologists hinders timely diagnosis. The subtle color variations in affected skin regions, which often resemble normal skin tones, underscore the critical need for early detection systems. To address this challenge, this paper proposes a hybrid deep learning segmentation model for automated skin lesion segmentation, aimed at improving early diagnosis of skin cancer. The proposed model leverages the HAM10000 dataset and integrates EfficientNetB7 as the encoder, UNet++ as the decoder, and Convolutional Block Attention Modules (CBAM) within skip connections. By incorporating CBAM, the model achieves a Dice coefficient of 0.9468 and an Intersection over Union (IoU) score of 0.9001, reflecting a 1–2% performance improvement compared to the model without CBAM. This noise-robust augmentation strategy enhances model generalization, ensuring robust performance on both noisy and normal dermatoscopic images. To further validate its effectiveness, the model was trained on additional datasets, including ISIC2016, ISIC2017, PH2, and MSK, and evaluated on their respective validation and test splits. The model demonstrated superior generalization and robustness, achieving a Dice coefficient of 0.9301 on the ISIC2016 dataset. The proposed model also achieved superior performance on the other datasets as well. These lightweight models establish a strong foundation for automated skin cancer diagnosis. Future work will focus on refining preprocessing techniques and incorporating multimodal data to further enhance model performance, with the ultimate goal of revolutionizing dermatological care and improving patient outcomes.

Keywords: CBAM, EfficientNet, UNet++

Acknowledgement

We express our deepest gratitude to all those who have supported and guided us throughout the completion of our thesis.

First and foremost, we are grateful to our thesis supervisor Dr. Muhammad Iqbal Hossain sir, and our secondary supervisor Mr. Rafeed Rahman sir. Their continuous support, guidance, and insightful feedback have greatly contributed to the depth and quality of our research.

Additionally, our heartfelt appreciation goes to our friends and family for their constant support and encouragement. Special thanks to our parents for their love, patience, and belief in us.

Lastly, we acknowledge all those who have contributed to this thesis in one way or another.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

Skin cancer arises when the growth of the cells in skin tissue becomes abnormal due to exposure to the Sun's Ultraviolet (UV) radiation. According to the study [1], excessive sun exposure is the main reason for Skin Cancer as it leads to abnormal growth of skin tissue cells. Another study [2] found that almost all skin cancers develop among 95% of individuals due to excessive exposure to the sun's Ultraviolet radiation. Other contributing factors to skin cancer include precancerous skin lesions, moles, fair skin, history of cancer, weakened immune system, and exposure to substances like arsenic. The three main types are Melanoma, basal cell carcinoma (BCC), and squamous cell carcinoma (SCC).

Another study [3] found that Skin Cancer is highly pervasive in the United States with about 9500 new cases diagnosed every day and two deaths occurring hourly. It is estimated [4] that by the age of 70 Melanoma is developed by one in five Americans. However, skin cancer can impact people of any age and race. According to the study [5], Men over age 50 are at a higher risk while women tend to get affected at age under 50. Non-Hispanic white people have a 30% higher likelihood of developing skin cancer than Hispanic and Black individuals. Unfortunately, Black people who are diagnosed at a late stage are more challenging to treat.

Early detection and appropriate treatment help to cure skin cancer and stop cancer from progressing to other areas of the body. Another study [6] found that in most cases, people's unawareness worsens the situation as they don't go to the dermatologists even though they find rashes, bumps, moles itching, burning, and irritation on their skin. Along with dermatologists Deep learning algorithms like CNN, and R-CNN can perform to detect and classify skin cancer.

Diagnosis could be enhanced through segmentation and classification. Segmentation is the process of highlighting the region of interest from dermatoscopic images. It has been helpful for clinicians to help narrow their attention down to specific lesions that will undergo further scrutiny. It is followed by the classification stage, in which these segmented lesions are identified into different skin cancer types, such as melanoma, BCC, and SCC. Normally, both of these processes have been heavily reliant on dermatologists' skills, whose diagnostic accuracy often tends to vary very much because of experiences and training.

With recent interest in computer vision especially in machine learning and deep learning methodologies, dermatology has begun to change. Recent studies have also led to the revelation that such automated systems can do diagnoses comparable to, or even outperforming, those made by trained dermatologists. For instance, a deep learning algorithm recently developed by Esteva et al. [7] was found to classify skin lesions at an accuracy of 91%, which was significantly high compared with the 88% obtained usually from dermatologists. This finding highlights one way in which such technologies have the potential to support healthcare providers in making more informed decisions.

For example, one study, by Tschandl et al. [8], tested CNNs on identifying skin lesions and reported an average accuracy of 94.4%. The paper illustrates how deep learning models can potentially extract complex features and patterns from images that are difficult for human observers to detect. Besides, the very fact that Haenssle et al. [9] managed to couple a deep learning algorithm with the diagnostic performance of expert dermatologists at 95% sensitivity and 87% specificity in distinguishing malignant from benign lesions speaks a lot not only about the efficacy of automated systems but also about how beneficial these technologies can make diagnostic workflows for doctors and specialist.

Such developments, though promising, need to overcome several challenges before machine learning models can be integrated into clinical practice. First, large annotated datasets are required to train the algorithms and make them robust. Publicly available data such as ISIC Archive have helped in trying to address this issue, but representativeness and diversity of data are extremely important concerns. Such biases at the time of training the models may result in substantial differences in diagnostic performance among the subject groups and hence further work in terms of equity and precision is always required. Besides, interpretability is an important challenge in machine learning models. Health experts must understand and trust the system's predictions. This becomes even more true at high-stakes environments, such as in diagnoses related to cancer. Researchers are working on the development of methods that could help create more interpretable models. These models, instead of providing accurate predictions, are able to provide insight into the process leading to a particular decision.

In that respect, automatic segmentation and classification of skin cancer open a new horizon toward diagnostic accuracy and patient outcome. While studies in the area continue to evolve, the clinical translation of machine learning technologies is highly promising for enhancing early detection of skin cancer and, consequently, prognosis.

This thesis introduces a hybrid approach to skin lesion segmentation that overcomes the existing drawbacks of current models by proposing a hybrid segmentation model designed to enhance accuracy, robustness, and computational efficiency.

1.1 Problem Statement

Skin cancer is becoming more widespread every day and is currently regarded as one of the major worldwide health challenges, with millions of new cases diagnosed annually. Due to its high prevalence, it can be potentially fatal if diagnosed at a later stage. In the US, it is estimated that about 9,500 people are diagnosed with skin cancer on a regular basis. Additionally, it is predicted that the UK will have around 20,800 skin cancer diagnoses this year. Early detection and accurate classification are essential for proper and effective patient treatment. Current diagnostic methods rely on dermatologists' visual examination, followed by dermoscopic analysis and biopsy. However, diagnosis is often limited by the availability of healthcare resources and human error, and it can be time-consuming. This is where the implementation of artificial intelligence (AI) models comes in. At present, several AI models are being developed, and many are already in use to automate the diagnosis and classification of skin cancer. However, these models also have limitations. They often require devices with high computational power and sometimes even unique industrial-grade equipment, making them unavailable in most parts of the world. Moreover, most of the existing models struggle to perform effectively on noisy images. To address these challenges, this study proposes a lightweight hybrid segmentation model designed to perform robustly on both noisy and standard dermatoscopic images, offering a viable alternative to industrial grade models.

1.2 Contribution

Our research advances automated skin cancer diagnostics through the following contributions:

- **Hybrid Segmentation Model**: We propose a segmentation framework integrating EfficientNetB7 as the encoder, UNet++ as the decoder, and Convolutional Block Attention Modules (CBAM) across all skip connections. This architecture leverages EfficientNetB7's hierarchical feature extraction, UNet++'s precise localization, and CBAM's attention mechanisms to improve lesion boundary delineation.
- **Noise Robust Augmentation Strategy**: To enhance model generalization and robustness, we implemented a comprehensive data augmentation pipeline. During augmentation, along with adding noise, we used different image enhancement augmentation techniques, including:
  - *Noise Induction*: Applied to 60% of the training data by first enhancing images with unsharp masking to emphasize edges, followed by combining the resulting images with high-boost filtered versions to introduce controlled noise, simulating real-world imaging variations.
  - *Geometric Transformations*: Random rotations (up to 60 degrees), random horizontal and vertical flips, random horizontal and vertical shifts, random shear transformations, and random zooming.
  - *Color Adjustments*: Random modifications to brightness and contrast.

All images were uniformly scaled to a resolution of 128×128 pixels to ensure consistent input dimensions for the models.

These advancements were systematically refined and evaluated using the HAM10000 dataset, demonstrating improved performance over existing state of the art approaches. The proposed model is lightweight and computationally efficient, making them viable for deployment in resource-constrained clinical settings.

1.3 Research Objectives

This research proposes a lightweight and reliable model for skin lesion segmentation task. The proposed approach leverages an hybrid segmentation model that uses EfficientNetB7 as encoder and UNet++ based decoder, and CBAM Module in the skip connection. The main objective of this research is to propose a model that not only outperforms all the current state-of-the-art models but also remains lightweight. The other objectives and goals of this research are as follows:

- Develop a hybrid segmentation model using EfficientNetB7, UNet++, and CBAM for precise lesion delineation.
- Implement a noise-robust augmentation strategy via unsharp masking, high-boost filtering, and diverse geometric and color transformations.
- Evaluate the proposed model against state-of-the-art approaches using the HAM10000 dataset.
- Identify limitations and propose avenues for future improvements.

1.4 Thesis Structure

This thesis is structured to explore the optimization of skin lesion analysis through a noise-augmented EfficientNet approach, presenting hybrid model for segmentation across six chapters.

- Chapter 1 introduces the problem of skin cancer diagnosis, emphasizing its global health impact and the need for early detection. It outlines the research contributions, including a novel hybrid model, and states the objectives of improving accuracy and robustness in skin lesion analysis.
- Chapter 2 reviews the literature on skin lesion segmentation, focusing on deep learning methodologies. It identifies gaps in current approaches, such as computational demands and performance on noisy images, which the proposed models aim to address.
- Chapter 3 describes the datasets used, primarily the HAM10000 dataset for model training and validation. It also introduces additional datasets (ISIC2016, ISIC2017, PH^2, MSK) for evaluating the models' generalization capabilities.

- Chapter 4 details the methodology, presenting the architectures of the hybrid segmentation model. It explains the integration of EfficientNetB7, UNet++, CBAM for the segmentation model along with the noise robust augmentation strategy to enhance model performance.
- Chapter 5 presents the results and analysis, including quantitative comparisons with baseline models and visual evaluations. It discusses the models' performance on the HAM10000 dataset and their generalization to other datasets.
- Chapter 6 concludes the thesis by summarizing the key findings, highlighting the hybrid segmentation model's strengths and limitations, and suggesting future research directions, such as advanced preprocessing techniques and multimodal model integration.

.

Chapter 2

Literature Review

This chapter discusses the literature reviews on skin lesion segmentation and classification, focusing on deep learning methodologies. It identifies gaps in current approaches, such as computational demands and performance on noisy images, which the proposed models aim to address.

In this paper Bindhu et al. [10] presented a skin cancer segmentation method which is Fuzzy U-Net. Skin cancer, especially melanoma, represents a real serious health problem worldwide, with continuously growing incidence over the last five decades. The key to survival is early detection; this can raise cure rates as high as 90%. The traditional method of diagnosis based on visual inspection often has incorrect classification due to similarities among skin lesions. The paper proposes a new MFO-Fuzzy U-Net for the segmentation of melanoma from skin images. The bilateral filter is being used as a preprocessing step to suppress the noise in images before feeding them into Fuzzy U-Net for segmentation. May Fly Optimizer improves the accuracy of the proposed model. The main performance metrics are accuracy, precision, specificity, recall, and F1 score. It outperformed traditional models, namely YOLO, SegNet, and U-Net, with an accuracy of 97.57%. In comparison, this provides an accuracy enhancement of 13.04%, 0.52%, and 22.49% over FRCNN, AlexNet, and CNN, respectively. This work further identifies that automated systems are needed to enhance early detection, thereby increasing the diagnostic reliability of such conditions. Future enhancement upon the methodology will refine the segmentation accuracy for better classification of disease conditions. The model presents a very promising approach to tackling the major issues in melanoma segmentation.

Ashraf et al. [11] excelled in the possibilities of melanoma segmentation using deep learning with test-time augmentations. It shows the usage of conditional random fields (CRF) to model the state of both the input and output images. For this, neighborhood processing was used as it resulted in a better prediction. To achieve the desired result, datasets such as the ISIC-2016 and the ISIC-2017 were used. Also, the combined version of both datasets was used. In such cases, the duplicate data were removed for convenience. For pre-processing they used a morphological filter with the inpainting of an algorithm to effectively remove unwanted hair and preserve the shape. However, rather than depending on the testing or optimizing just the neural networking method, the approach was more prone to successfully

training one to handle dermoscopic images captured under different conditions. The main theme of the methodology is to utilize a combination of U-Net, ResUNet, and ResUNet++ architectures that are trained to predict a fine-tuned segmentation mask using Bayesian learning which is quite efficient in enhancing the overall performance of the lesion segmentation compared to conventional semantic segmentation methods. For post-processing, the authors applied the TTA to improve the robustness of the model during inference. Also, CRF is used to refine the segmentation by considering spatial dependencies. On the ISIC-2016 dataset, the U-Net achieved a Dice coefficient of 93.74%, a Jaccard index of 89.59%, a precision of 95.03%, and a recall of 90.21%. On the ISIC-2017 dataset, the UNet model achieved a Dice coefficient of 83.74%, Jaccard index of 80.65%, precision of 85.00%, and recall of 85.15% proving it to be the best outcome. During the process, over-segmentation and under-segmentation failure cases can be noticed, which is possible to mitigate by incorporating additional post-processing techniques like geodesic active contour level. Thus the adoption of post-processing technique. Again the dispatch is limited by different acquisition states of the dermoscopic images that were captured.

Duque-Vazqueza [12] used an approach to HeLa segmentation using an Image Processing algorithm rather than a Neural Network method because of minimal requirements. The work uses 300 images taken from the SBFSEM. The morphological operation of erosion seemed quite useful in terms of cell segmentation. Cells located at the center were detected using distance calculation where distance and size were considered as parameters. The methodology focuses on the algorithmic approach, outlining steps for approximating cell shapes, binarization, erosion, and verification stages to ensure segmentation accuracy. So, it's more like building an algorithm to achieve the goal rather than the approach being algorithm-centric. As per the cell shape segmentation assessment, the z-test showed better results than Karabağ's. The proposed algorithm scored 87.24% in the Jaccard Index, 85.56% in the NPR Index, 93.18% in Dice, and 94.65% in the F1 score for Median. For Mean the scores are also higher than Karabağ ones in each index and the scores are lower in the case of Standard Deviation. On the other hand, the Nuclei segmentation assessment shows of same types of results which indicates the success of the newly built algorithm for segmentation. Moreover, the visual results are also satisfying where the segmentation of cells presents presents the most accurate results. Then again, the algorithm showed a processing time of 818.85 seconds, smaller than the number of reference algorithms. This proves the speed. However, the algorithm works for images collected from SBFSEM. Here, the structure of the cell and nucleus is well-defined. There can be a possible shortcoming of not being effective under different conditions.

Cai et al. [13] in their research proposed a bi-directional transformer-based UNet segmentation model named BiADATU-Net. This model features a deformable attention Transformer and DAD blocks in the encoder, and a bidirectional attention module in the UNet architecture, consisting of DAD and scSED blocks. DAD blocks are placed before the deformable attention transformer, while scSED blocks are used in skip connections, both utilizing deformable convolution operations. The encoder block consists of Convolutional layers, DAD blocks, embedding layers, and deformable attention transformer layers. The feature maps of the CNN modules

are passed to the DAD block's PAM module and CAM module for position and channel-level feature extraction. Then the feature map from the PAM and CAM modules are merged. DAD-block's deformable convolution layers make the feature extraction capabilities more efficient and flexible. Then the feature maps are passed to the deformable attention transformer. scSED blocks were used in the skip connections; it uses deformable convolution layers to capture the pixel-level spatial information, therefore, resulting in an accurate feature map reconstruction in the decoder. This model uses the default decoder of the UNet.The model underwent training and validation using the ISIC2016, ISIC2017, ISIC2018, and the PH^2 dataset. All the images were preprocessed using the SharpRazor [13] method to remove the hair and ruler marks from the images then different augmentation techniques were used to balance the dataset classes to prevent overfitting problems. SGD optimizer was used for hyperparameter tuning. This model outperformed various state-of-the-art models like U-Net, R2U-Net, Attention U-Net, TransUNet, and FAT-Net. This model achieved an average accuracy of 96.37%, 91.45% sensitivity, 93.07% precision, 91.69% Dice score, and 85.51% Jaccard index. Although this model achieved a great performance, it has some limitations. One of the major limitations is its high computational complexity. Additionally, it relies on the default U-Net decoder, which could be further optimized for better efficiency.

Khan et al. [14] proposed a hierarchical transformer-based segmentation model named DeepSkinFormer. This model is composed of two modules these are scaling module and the hierarchical transformer module. The hierarchical transformer module combines the default architecture of Vision Transformer with the SEEM module. The SEEM block takes the feature maps and the guidance map as input, using semantic difference convolution layers to output a feature map for enhanced skin edge segmentation. This model's scaling unit consists of a multi-head attention block using shifted windows that focus on the major information for global self-attention. This model's feature fusion is done in three steps. Firstly, the feature maps of the semantic difference convolution and encoder features are fused in the SEEM block. Then the outputs of the SEEM blocks are fused with the patch embeddings and then finally, the outputs of the scaling module and the transformer module are fused together in the segmentation head resulting in a more enhanced skin lesion segmentation. This model was developed and evaluated on the three popular datasets, HAM10000, ISIC2017, and PH^2 datasets. It was compared with advanced models like SegNet, Mask R-CNN, UNet, and ResUNet. It outperformed all of these four models achieving an average dice score of 95.92%, an average mIOU of 0.7679, an average 89.03% precision, and an average recall of 89.03%. These results show the potential of the DeepSkinFormer model in effectively segmenting skin lesions.

The SkinSAM model [15] primarily focuses on the segmentation of lesions. Based on the SAM model (pre-trained on the SA-1B dataset), it compares three ViT models (base, large, and huge) to the fine-tuned ViT-b model. The ViT models use 1024x1024 images, so an encoder was used to resize inputs to 3x1024x1024. Utilizing an Adam optimizer with a 1e-5 growth rate, the model's performance was monitored until it reached its peak. Since SAM uses prompts, it was observed that using prompts yielded better results. Now, the following number of parameters are

required to generate outputs: ViT-h : 636 million, ViT-l: 208 million, ViT-b: 91 million. So, the fine-tuned ViT-b model uses less computational power during the feed-forward and self-attention layers. In the end, while the larger models provided better accuracy, the fine-tuned model offered the best balance of efficiency and performance.

Imran et al. [16] paper primarily focuses on classification using the updated MiT-B0. The segmentation is taken from another paper. The second part focuses on creating semantic segmentation using a hierarchical encoder known as the Mix Transformer encoder (MiT). The encoder takes an image input to generate the values of multi-level features to produce features of a higher level. For the decoder, it follows the Vision Transformer (ViT) patch-merging process. They also apply data augmentation and modify the attention head to achieve the following results, an F1 score: 90.8% and an accuracy: 83.1%. Finally, they compare the results with another paper using a modified MiT model to show the class-wise segmentation accuracy of the models.

Toprak and Aruk [17] introduced a hybrid CNN model for skin cancer segmentation and classification that primarily focuses on creating a balance between the specificity and the sensitivity of the model rather than merely focusing on improving accuracy. Their model comprised five modules and these are: segmentation, feature extraction, feature concatenation, feature selection, and classification. ISIC2018 dataset was used for training the model and ISIC2019 and PH^2 datasets were used for evaluating the performance of the model. Data augmentation was done to create a balanced dataset for training. The semantic segmentation module uses the DeepLabv3+ model with ResNet50 as the backbone. After the segmentation, they used three renowned CNN architectures, MobileNetV2, EfficientNetB0, and lastly DenseNet201, for diverse feature extraction from the segmented lesion images. Each of these CNN architectures returned a 1 x 1000 size feature vector. Then the three resulting feature vectors were concatenated into a 1 x 3000 feature vector, followed by applying the ReliefF algorithm to select distinct features. Finally, the K-Nearest Neighbors (KNN) algorithm was applied for classification. It achieved a 94.90% dice score and an 89.55% Jaccard index during segmentation. During classification on ISIC 2019 dataset, this model achieved 94.42% accuracy, 94.13% precision, 92.99% sensitivity, 98.96% specificity, and an F1 Score of 93.49%. On the PH^2 dataset, the performance metrics were somewhat similar and they were as follows: 94.44% accuracy, 94.43% precision, 94.48% sensitivity, 96.68% specificity, and an F1 score of 94.43%. One of the limitations of the model is that it achieved a great performance in classifying melanocytic nevi lesions and vascular lesions, but in some cases, it could not identify melanoma and basal cell carcinoma perfectly, still, it achieved an overall 94.42% accuracy. They further want to make the model more efficient enough so that it can be applied to other medical datasets as well.

Pintelas and Livieirs [18] proposed XSC, a transparent and easily interpretable model for explainable image segmentation and classification. This framework is comprised of three modules and these are clustering, segmentation, and classification. The first two modules collaborate for segmentation purposes. K-means clustering was implemented in the clustering module to perform both texture and

shape-based clustering. Here k was selected based on the silhouette score. Then in the resultant texture and shape-based subregions, they applied mean value and standard deviation value on the five sub-regions with the closest centroids for better segmentation. For transparency and easier interpretability, they also applied feature hierarchy trees in the segmentation procedure. For classification, they used the XGBoost classifier and the statistics value they used for classification are the mean value and the standard deviation value. This model was trained and evaluated on a balanced version of the ISIC2018 dataset. Early stopping techniques were used to prevent overfitting issues during training. This model achieved a good result in malignant cases as there was only little difference between the ground truth mask and the segmented mask. But while identifying benign cases this model's algorithm prioritized segmenting the main lesion area rather than segmenting the border areas. This model's segmentation performance was evaluated against different state-of-the-art models like U-Net, DeepLab, BASNet, and many other models. The segmentation performance metrics of the proposed model achieved a Dice Coefficient of 83% and an IoU score of 0.25 which was close enough to the performance metrics of the U-Net and BASNet [18]. The classification metrics of the proposed model are as follows: 88.1% accuracy, AUC score of 94.7%, 89.0% sensitivity, 87.7% specificity, 87% Positive Predicted Value (PPV) and 89.8% Negative Predicted Value (NPV). There are some limitations of the model. Firstly, the researchers used the K means clustering algorithm and they suggested that if they could have used different clustering algorithms like spectral clustering or hierarchical clustering then it would have resulted in a more precise segmented region which would have increased the overall performance of the model. Moreover, this model was implemented based on handcrafted features like different descriptive statistics which might not have extracted all the necessary features from the images. Furthermore, this model is not robust and generalized as it was tested on a specific dataset only.

Alam et al. [19] in their research proposed a DCNN model named S2C DeLeNet that does segmentation and classification simultaneously. The segmentation module uses a modified U-Net architecture. One of the major drawbacks of U-Net architecture is that it does not have a dilation feature and it cannot extract high-level features. As a result, the vanilla U-Net can not segment regions that have close values with respect to skin tone. Now to address this problem they modified the U-Net by replacing the vanilla encoder of the U-Net with the EfficientNetB4 model as it can extract high-level features. Then the classification module has a feature extraction system that classifies each lesion image based on the transferred parameters from the segmentation subnetwork. Moreover, this module's feature extraction system consists of FCM and 3DLR. FCM was used to mix features from both the encoder and decoder. 3DLR was used to create high-variance feature maps for achieving better classification results. This model was trained and fine-tuned on the HAM10000 dataset, with a batch size of 8 and a learning rate of 0.0001 for the hyperparameters. During the whole experiment, Adam optimizer was used. The model was compared with some state-of-the-art models. During segmentation, this model achieved a mean of 94.89% dice score and a mean accuracy of 97.20% on the validation and test set of HAM10000 dataset. During classification, this model achieved a mean accuracy of 94.03% and a mean precision of 85.875%. Their model

was also evaluated on other skin lesion datasets like ISIC2016, ISIC2017, ISIC2018, and PH^2. This model achieved a mean accuracy of 94.50%, a mean dice score of 91.435%, and a mean Jaccard index of 82.36%. Though their model achieved a good result, it has some limitations due to having limited VRAM they could not have trained their model with a higher batch size which could have resulted in enhanced performance. Also, the dataset was trained only on the HAM10000 dataset which consists of only seven types of skin lesions, but there are numerous types of skin lesions, so if any type of unseen image is passed it may not perform well. So it needs to be trained on such unseen data as well. Lastly, their future goal is to make this model efficient enough so that it can be used for other types of datasets.

He et al. [20] in their research paper proposed a lightweight and faster end-to-end CNN model named MTL-CNN that does edge prediction, segmentation, and classification simultaneously. This model is composed of four modules and these are one shared encoder two parallel decoders and a classification subnet. Because of using a shared encoder, this model resulted in a smaller number of parameters. One of the decoders is used for edge generation and segmentation and the other decoder is used for classification subnet. The encoder uses ResNet50 as its backbone for the feature extraction and then the extracted features are used for edge prediction, segmentation, and classification. Two ASPP modules were used to establish a connection between the segmentation and edge decoders with the encoder. The segmentation and edge decoders have three decoding units connected with the encoder using a skip connection. The decoding unit is used for up-sampling the feature map in decoders with the feature map of the encoder. Among those three decoding units two of them are connected with the encoder using EIE modules. Many EIE modules were used to facilitate the connection from the encoder to the segmentation decoder, which is used for identifying the edge of the lesion area that further enhances the segmentation masks in the segmentation module. Furthermore, they also used many LAE modules that were used between the encoder and the classification module. In this module, they first applied binarization on the segmented region's feature map to remove different artifacts like hair, bubbles, etc. The contour detection algorithm was used to generate the bounding box and then it was dilated by 1.4 times to create a dilated bounding box. Next, ROI projection was used to map the dilated lesion area from the image and then finally, it was passed to the classification subnet to perform classification. In their research, they used three datasets for their model, two of them are publicly available datasets, the ISIC 2016 and the ISIC2017 dataset. The last dataset is a clinical dataset named the Xiangya Clinic Dataset. This dataset consists of 1768 images of different skin lesions and their corresponding segmented mask as ground truth. Different data augmentation techniques like color adjustments, random cropping, random flips, and random rotation were used before training the model. Dice loss and Binary Cross Entropy loss were chosen as the loss functions for segmentation and classification respectively. During segmentation, this proposed model achieved a 93.4% dice score, 87.9% Jaccard index, 97.2% pixel-wise accuracy, 93.8% pixel-wise sensitivity, and 97.5% pixel-wise specificity on the ISIC2016 dataset. Furthermore, it achieved an 88.7% dice score, 81.5% Jaccard index, 95.5% pixel-wise accuracy, 88.9%pixel pixel-wise sensitivity, and 98.3% pixel-wise specificity on the ISIC2017 Dataset. But, this model achieved higher performance on the segmentation of the

Xiangya Clinic dataset, 92.4% dice score, 86.4% Jaccard index, 96.9% pixel-wise accuracy, 93.1% pixel-wise sensitivity, and 97.8% pixel-wise specificity. During classification, it achieved 88.5% accuracy, 67.3% sensitivity, 96.3% specificity, 87.5% AUC on ISIC2016 dataset, and 90.7% accuracy, 76.8% sensitivity, 93% specificity and 93.9% AUC on ISIC2017 dataset. The model achieved 89.4% accuracy, 96% sensitivity, 87.5% specificity, and 97.3% AUC on the Xiangya Clinic dataset, with an inference speed of 0.224 seconds per image [20]. Though this model is faster, it has some limitations. One of the limitations is that this model is not generalized as it is trained on skin lesion dataset. Another limitation is that the dataset they used for their research was limited in size. Their further goal is to make this model more efficient enough so that this model can be used for other medical datasets.

Choudhary and Chauhan [21] proposed a new approach for skin lesion segmentation and classification. They employed UNet for segmentation and a modified VGG network for classification. Before training their segmentation model they first preprocessed the images to remove hair artifacts. For image processing, they first downscaled the image, followed by grayscale, morphological filtering, binary thresholding, morphological cloning, and inpainting to remove the artifacts. Then different data augmentation techniques like random rotation, horizontal and vertical flipping, scaling, and brightness and contrast adjustments were applied to create a balanced dataset. In the UNet segmentation module, they used ReLU activation functions and random dropout layers to prevent the model from overfitting. The sigmoid activation function was used to perfectly identify the boundaries of the lesion region. Modified VGG architecture was used for classification. In that module, they implemented batch normalization and dropout regularization to prevent overfitting, used the ReLU activation function in the dense layers, and applied the SoftMax function to classify the skin lesion based on the segmented region.Categorical cross-entropy loss and Ada-Max optimizer were used to fine-tune the model's hyperparameters. The model was trained and validated using the HAM10000 dataset, while its performance was tested with the ISIC2018 dataset. During segmentation, the model achieved an accuracy of 96.04% and a dice score of 93.1% in the ISIC2018 dataset. In classification, it achieved an accuracy of 97.06%, 99.67% sensitivity, 99.99% specificity, and 96.04% F1 score.

Kavita et al. [22] proposed a method to classify seven types of skin cancer based on CNN and ResNet50. The ISIC database with 2357 (categorized and evenly distributed) images was used in the paper. Firstly, the image quality was enhanced by employing preprocessing. Secondly, segmentation was employed by the authors to prepare the image for analysis. Next, features were extracted from the image. This paper used CNN and ResNet50 for skin lesion classification and detection. In CNN, features were detected using convolution layers, parameters were reduced by pooling layers, and classification tasks were done by fully connected layers (FC). ResNet50 was used to improve the accuracy and efficiency by extracting features from the input image in a combination of 1x1,3x3,5x5 convolutional filters, using batch normalization and ReLu activation. This proposed method showed 76.16% precision, 78.15% recall, 76.92% F1 score, and 91.32% accuracy overall.

Elgohary et al.[23] used HAM10000 and PH^2 datasets and CNN based approach

was employed. In this paper, the variability of the dataset was increased by preprocessing and augmentation. For the classification task, the CNN model was used which starts with the convolutional layers that deploy 16 filters for feature extraction from the images. Then max pooling layers down-sample spatial dimensions. Lastly, a flattening operation was applied to prepare the data for the dense layer after the final convolution layer (FC). The dense layer features 64 and 32 neurons that act as potent feature extractors and the output is generated using the dense layer with seven neurons each of them indicating a class. Softmax function is employed to ensure that the model provides probabilities for each class. For the segmentation, a U-NET-based segmentation model was employed that features an encoder-decoder architecture. The encoder has four convolutional layers and the decoder is just the mirror using transposed convolutions to up-sample feature maps. Models were rained with Jaccard distance loss over 100 epochs using Adam optimizer with a learning rate of 0.003 and the best trained DeepConNet model had impressive 99.5% accuracy, 99.5% recall, and 99.5% precision in classification and 97.204%accuracy,97.2% recall and 97.5% precision in classification.

In 2023, A new method for skin cancer diagnosis by combining optimization techniques and ANN was introduced by Lai et al. [24]. This new method consists of 4 steps. Firstly preprocessing was done to remove unnecessary information from the input images. Secondly, segmentation for each image was done using a Kohonen neural network, and a GSA-based iterative process was performed to improve the segmented region. Thirdly, feature extraction was performed by using CNN from the ROI. Finally, for the classification task, a combination of ANN and IGWO was employed. The proposed method was applied to ISIC-2016 and ISIC-2017 datasets and achieved 97.09% and 95.17% accuracy respectively.

In this paper, Mahin et al. [25] proposed A new segmentation and a new classification model. This model combines SSO and k-means clustering. Minimizing the performance dependency on the image conditions is the main objective of the SSO. Also, CNN is used for feature extraction. In this paper, They also proposed a classification model combining the ECOC model and gamma classifiers. The performance of the model is enhanced by integrating a weighted distance measure Experiments were conducted on both datasets using MATLAB software version 2020. To avoid potential bias in results the authors performed cross-validation on the entire dataset. The proposed method was employed on ISIC-2016 and ISIC-2017 datasets achieving 97.10% and 95.17% accuracy respectively. Some limitations of this model are limited datasets, High time consumption of CNN and Gamma classifiers, and Decisions made by CNN are difficult to understand. In the future, we must work on large data, and explore transfer learning and AI techniques development.

In 2021, Bibi et al. [26] introduced a novel hybrid contrast enhancement approach to improve segmentation tasks. Based on the histogram the best channel was selected and an activation function was proposed by the authors to construct a saliency map and convert the image into binary form which is then mapped onto the original image for final detection. The segmentation task was performed on the ISIC 2017 dataset achieving an accuracy of 95.6%. VGG16 and MobileNetV2

two predefined models were modified and trained by the authors through transfer learning for classification. CCA was used for feature extraction from both models and the MESbS method was used for feature selection. Finally, a C-SVM classifier was used to classify. The Classification Task was performed on the HAM10000 dataset achieving an accuracy of 96.7%.

Akram et al. [27] proposed segmentation and classification using multichannel saliency estimation and M-SVM processes. To extract the features to work on, they followed the ABCD rule and depended on the raw image, too. They used CV aid and exploited ternary color spaces such as CIE-Lab, CIE-YCbCr, and CIE-LUV to figure out the proper features. In the preprocessing part, the Dull Razor software is used to remove hair and any unnecessary ruler lines to avoid noisy segmentation. In the second step of image processing, contrast stretching (also called normalization) is used to improve image quality by changing the intensity range. This technique is especially helpful when working with dermoscopic images, like those from PH^2, ISIC, and ISBI 2016 datasets. These images rely on the assumption that there is a high contrast between the lesion and the background, which helps in identifying salient regions. The method described works well with certain skin tones, but for others, like Asian or African skin tones, results can differ based on the visual differences between the foreground and background. The process starts with applying a Sobel edge filter to the image, followed by splitting the image into blocks of equal size. Weights are then calculated for each block based on gradient intensities. The algorithm uses these weights to adjust gray values in each block. The number of blocks and the block size are important factors in getting optimal results, with 12 blocks being used for the datasets mentioned. Before constructing the feature vector, morphological operations like closing, filling, and reconstruction are applied to enhance the difference between the foreground and background. After that, texture features are analyzed using a method called ETFA (Extended Texture Features Analysis). The algorithm combines PTBD with fuzzy C-means clustering to improve results. This method breaks down the image into binary layers, analyzes texture using fractal analysis, and counts pixels to build the final feature set. In this methodology, both custom and open-source modules are integrated. Most preprocessing tasks, from channel separation to weight calculation, are developed by the authors, except for the ETFA algorithm and Fuzzy C-means, which are open-source and built-in MATLAB functions. This method is quite helpful in calculating feature extraction. For classifying lesions, the built-in category classifier of images leverages the GPU for final categorization. Applying the model to the ISBI 2016 data set, we get an accuracy of 99.2%, a sensitivity of 99.2%, a specificity of 99.4%, a precision of 99.4%, an NPV of 0.6, an FNR of 0.8% and FPR of is0.005, received through M-SVM. For the PH^2 dataset, the method achieved the best testing accuracy at 97.5%, specificity at 99.5%, the precision rate at 97.3%, NPV at 2.7%, FNR at 2.5%, and FPR is 0.010 on M-SVM. ISIC's benchmark data set ISIC UDA base showed an accuracy of 99.5%. sensitivity 99.0%, specificity 100%, FNR 0.5% and FPR 0.01 on M-SVM. On the other hand, ISIC MSK 1 2 says accuracy is 99.0%, sensitivity is 98.5%, specificity is 99.0%, precision rate is 98.5%, and FPR is 0.020 for classification. CNN has shown significant success in various computer vision tasks like skin lesion classification, action recognition, and content-based image retrieval. CNNs learn high-level features in deeper layers, mak-

ing them promising for future implementation and testing on datasets like ISBI 2017.

Abedini [28] proposed a deep learning path to resolve the issue of segmentation and classification as well. The famous and well-known HAM10000 and ISIC 2018 were used for the research. For preprocessing, he used the DHR hair removal model and necessary normalization processes. He went on to develop an accurate deep-learning object detection model and ended up with YOLOv8 being the best as per performance. It identified the location of the skin lesion and provided a bounding box with sheer accuracy. Furthermore, SAM and MedSAM models were used to segment the border within the boundary box. For classification, several CNN models such as InceptionV3, DenseNet, Xception, VGG16, EfficientNet, and Coca-ViT were applied to segmented lesions to remove the background from the optimized images. While detecting the object, YOLOv8 showed the best accuracy which is 85.01%. The precision rate is highest with 73.87% for Faster RCNN. But YOLOv8 dominates in other performance metrics with 81.33% Recall and 75.43% F1-Score. Deep learning Segmentation model SAM MedSAM showed 70.03% and 78.55% of IoU from the ISIC data set. For image classification, the CoCa model uses a cross-modal transfer learning approach. The accuracy, precision, Recall, and F1 scores are chronologically 85.33%, 84.66%, 91.56%, and 87.97%. However, all the models showed better performances. However, further evaluation on larger, more diverse datasets would be needed to assess the efficiency of the mentioned approach fully.

Mahbod [29] developed a dermatoscopic image classification model. To achieve success, they created a baseline classifier which was used as a reference model. Here they restrained themselves from using any segmentation masks. Then, for both the train and test phases they manually and automatically created segmentation masks scenario-wise and recorded the outcome of segmentation. These scenarios were often used to exploit the segmentation masks. The purpose was either to crop lesion images to remove the background. For this, they chose the ISIC-2017 dataset for available labels for train and test sets and used ISIC-2018 for a better evaluation. Using a segmentation mask didn't bring any significant change in any scenario. However, For dilated cropping, the classification had been improved with a significant gesture. On the other hand, removing all background information using segmentation masks makes the overall classification outcome degrade significantly. Thus, inappropriate usage of the masks can significantly degrade classification performances. As a result Manual method had Avg. AUC of 93% and accuracy is 89.9%. Again automated processes such as SkinLinkNet, SkinUNet, and SkinFPN+ had Avg. AUC of 92.6%, 92.7%, and 93%. For accuracy, the numbers are chronologically 89.6%, 89.9%, and 89.9%.

In this paper, Sonia et al. [30] presented how a machine learning framework consisting of FOA combined with SVM can be applied for the segmentation and classification of skin cancer. Melanoma and other types of skin cancer are stated as being major concerns of public health, which should be diagnosed with a high percentage of precision and in time. The manual method of diagnosing the skin lesions is inconclusive, difficult and often involves some level of bias at some instances. The paper also put forward a better way for classification, the

FOA-SVM model. The hyperparameters of SVM are tuned by FOA to improve the pivotal values of penalty ratio and kernel width for the improved classification potential of skin lesions. The steps are image preprocessing, noise and hair removal, segmentation, and finally feature extraction, which improves the efficiency of classification. The system is validated using the data drawn from ISIC and DermNet sources and the results show high accuracy in diagnosing skin cancer compared to other classifiers. Thus, when applying FOA its results are superior to other well-known approaches including Genetic Algorithms or Particle Swarm Optimization. The author deduces that this framework greatly enhances the diagnostic precision, and it may help dermatologists to offer more rapid and precise diagnoses in the future. Future work also plans on improving the system deeper through the addition of other techniques to properly optimize as well as classify. Three skin lesion classes were selected from the Xiangyaderm dataset, the DermnetNZ database, and the HAM10000 dataset. The classification analysis using the proposed SVM method yields comparatively superior performance with classification accuracy of 98%, specificity of 99%, sensitivity of 96%, JSI of 95%, and DSC of 99%.

Karim et al. [31] suggested segmenting and classifying skin cancer using ViT in this paper. This work aims at skin cancer classification for early diagnosis through a deep learning approach employing a Vision Transformer (ViT). The study addresses this problem using the HAM10000 dataset that includes more than 10,000 dermatoscopic images of benign and malignant skin lesions, segments actual cancerous areas using the SAM, and achieves high accuracy with Google's ViT patch-32 model of 96.15%. It employs state-of-the-art preprocessing and optimization algorithms which results in a high classification speed. The analysis shows that there is an opportunity to incorporate the ViT models into smart devices to quickly diagnose skin cancer, mainly affecting fair-skinned people but with future adjustments to cover the deficit of the presented models on a diverse population base.

The paper will be presenting a new model for the segmentation and classification of skin lesions: the collaborative learning model, CL-DCNN, developed by Wang et al. [32]. When the sample data is not enough, this method will give key facts to the dermatologists. This work aims at the learning strategy of the teacher and student respective of segmentation and classification. In this regard, the researchers have employed a class activation map and self-training technique. In the CL-DNN skin lesion segmentation test, it results in a Jaccard index of 79.1% for the segmentation test and an average AUC of 93.7% for the skin disorder class. When this has been compared to other segmentation methods such as FCN, U-net, and DAGAN, among others. For the AUC, Specificity, and Accuracy, this method attained a 0.9% of increase, 0.4% of increase and 0.1% increase respectively, while its overall AUC was 93.7, SP was 94.7% and AC was 90.7%. The type of model has some limitations also. As the model is difficult to implement in clinical practices. Thus, expanding the framework is crucial for the model's generalization.

Chapter 3

Dataset Description

The first stage of our research methodology involves Dataset collection. While reviewing numerous papers, we observed that the HAM10000 [33], ISIC2016 [34], ISIC2017 [35], PH^2 [36] datasets are predominantly used in skin cancer-related research and models. For our segmentation model, we have selected the HAM10000 dataset for training, fine-tuning, and evaluation. After doing some research on the datasets, we selected the HAM10000 dataset as our primary dataset because it has more labels and is better organized. The HAM10000 dataset is a more substantial option in this regard because it represents a comprehensive collection of all key diagnostic categories in pigmented lesions. Additionally, pathology has verified that over 50% of the lesions are correctly diagnosed. The dataset contains 10015 dermatoscopic images, belonging to seven different label columns, with each column representing a specific type of lesion: are the Melanocytic Nevi (NV), Melanoma (MEL), Benign Keratosis (BKL), Basal Cell Carcinoma (BCC), Actinic Keratosis (AKIEC), Dermatofibroma (DF), Vascular Lesions (VASC).

Table 3.1 below provides a detailed description of each column in the HAM10000 dataset, including the variable name, data type, and a brief explanation of the variable. To further illustrate the composition of the HAM10000 dataset, Figure 3.1 displays the frequency distribution of its seven skin lesion types, highlighting class imbalances. Complementing this, Figure 3.4 represents a pie chart that visualizes the percentage distribution of these lesion categories, offering a proportional perspective. For qualitative insights, Figure 3.2 showcases eight randomly sampled images from the dataset alongside their corresponding ground truth masks, demonstrating the diversity of lesion appearances and segmentation complexity. Furthermore, Figure 3.3 provides a curated visual reference of one representative image per lesion type, each labeled with its respective diagnosis, aiding in the interpretation of class-specific morphological features. However, in the ISIC2016 dataset, the labels are not in a one-hot encoded vector format; instead, they are categorized in a binary format, representing the lesion as either benign or malignant. During the data preprocessing stage, the labels in the HAM10000 dataset were transformed into one-hot encoded vectors to ensure consistency across all the datasets used in our research. This standardization ensures smooth dataset integration and preprocessing and supports the effective development of the segmentation task.

Variable	Data Type	Description
Images	String	Specifies the file path to the skin lesion images in the dataset.
Masks	String	Indicates the file path to the corresponding ground truth segmentation masks for the lesions.
7 label skin lesions	One Hot Encoded Vector	There are 7 labels: NV, MEL, BKL, BCC, AKIEC, DF, VASC. This one hot encoded vector indicates the type of skin lesion corresponding to the image.

Table 3.1: Description of variables in the HAM10000 dataset

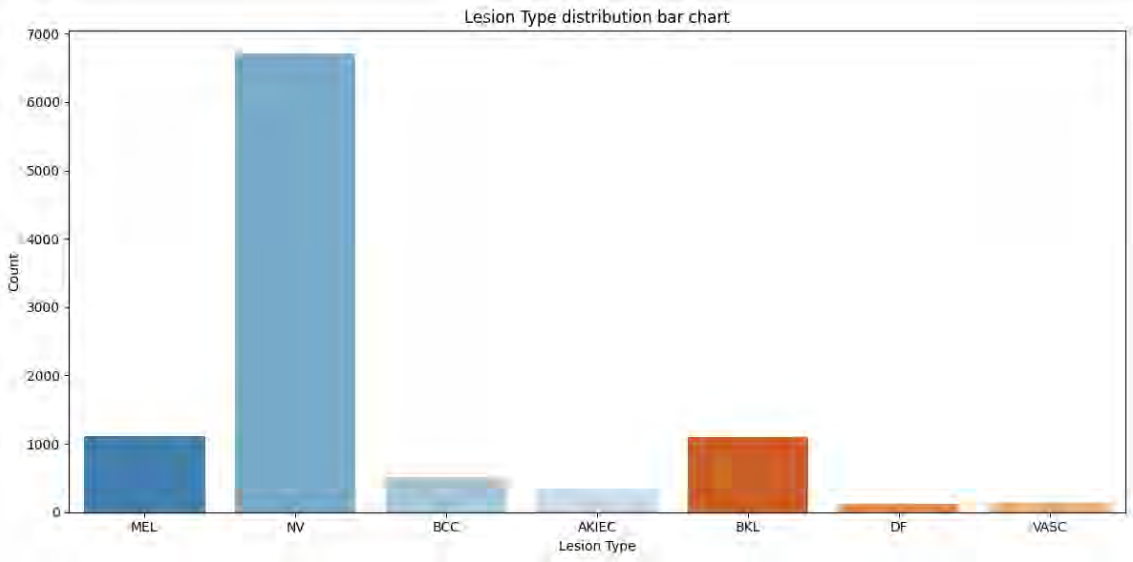


Figure 3.1: Distribution of skin lesion types in the HAM10000 dataset

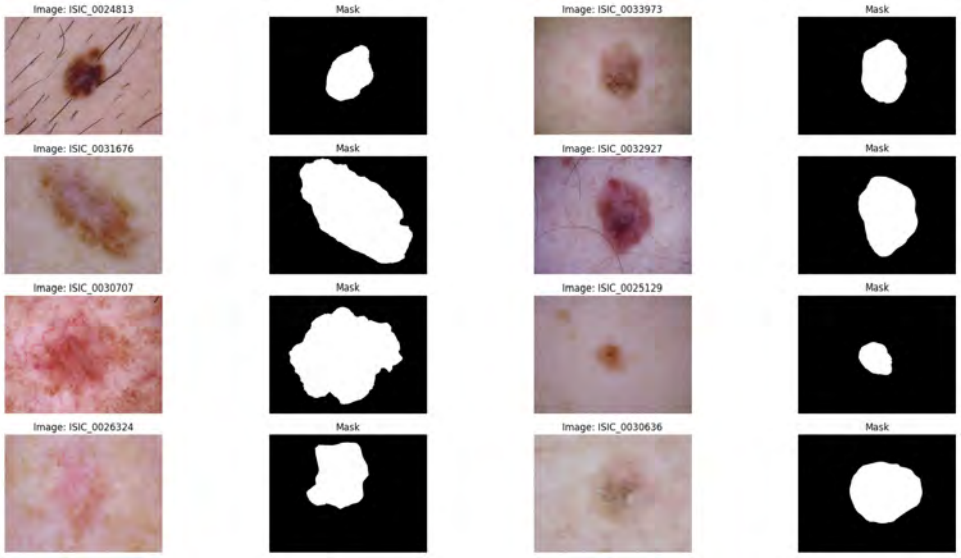
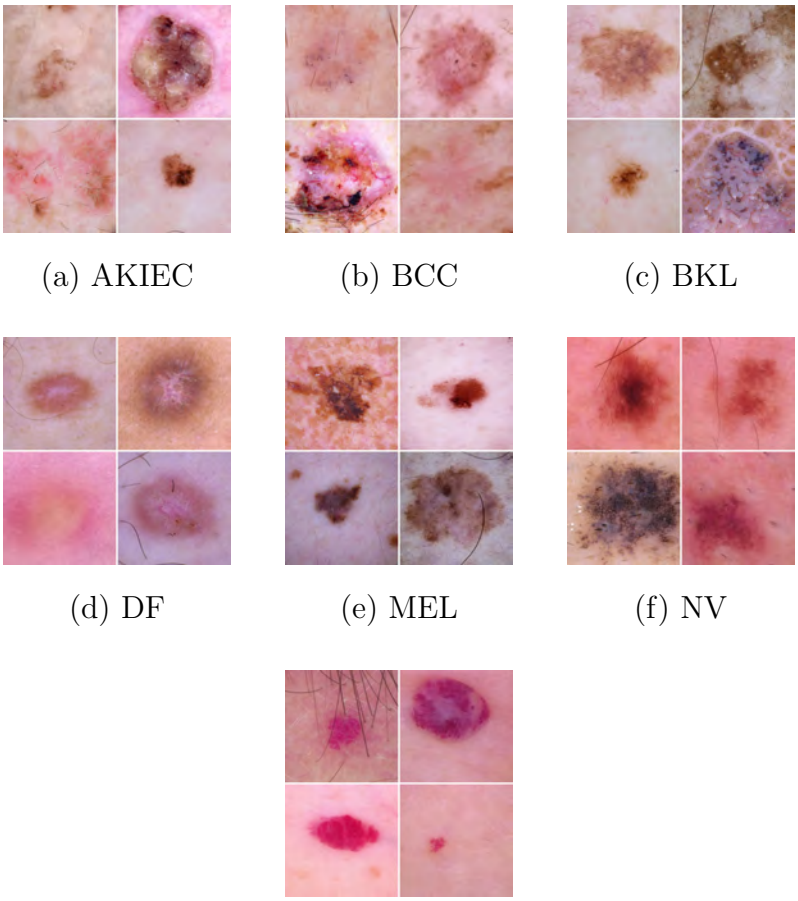


Figure 3.2: Lesion images with their corresponding ground truth segmentation masks



(a) AKIEC (b) BCC (c) BKL

(d) DF (e) MEL (f) NV

(g) VASC

Figure 3.3: Sample images of each skin lesion types from the HAM10000 dataset

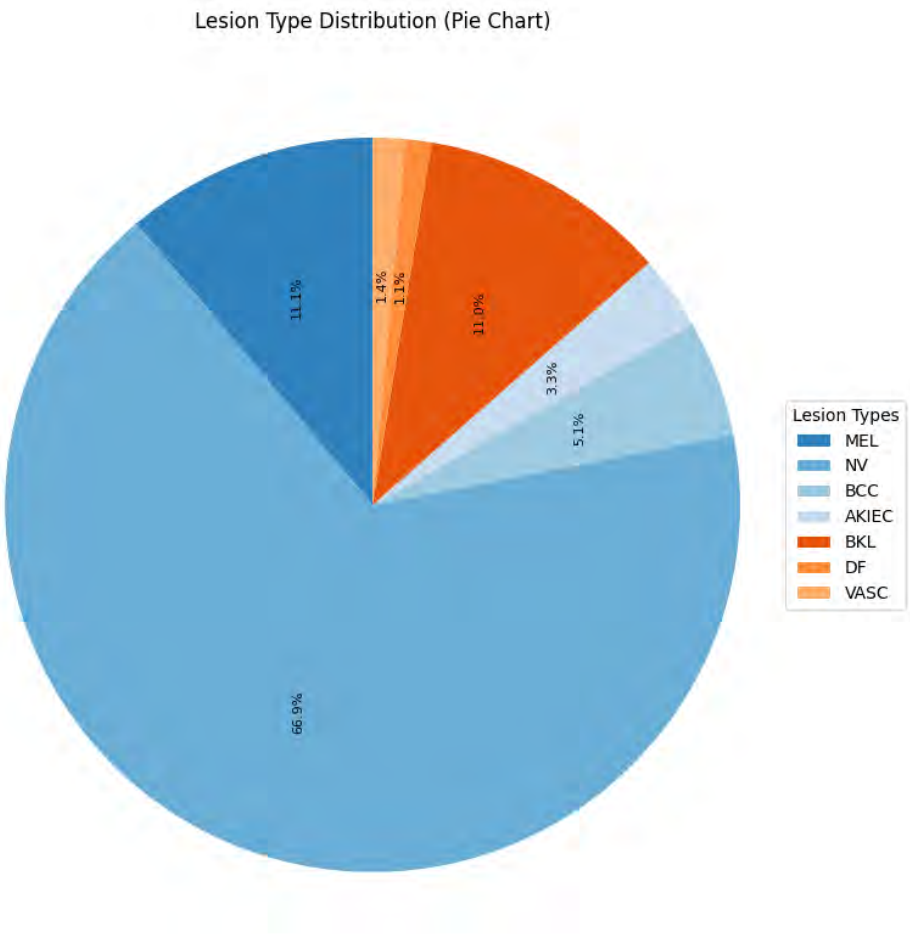


Figure 3.4: Percentage distribution of skin lesion types in the HAM10000 dataset

Labels	**Amount**	**Percentage**
MEL	1113	11.1%
NV	6705	66.9%
BCC	514	5.1%
AKIEC	327	3.3%
BKL	1099	11.0%
DF	115	1.1%
VASC	142	1.4%

Table 3.2: Frequency and percentage of each skin lesion type in the HAM10000 dataset

As seen from Table 3.2, the dataset is highly imbalanced, where the majority of skin lesion cases are NV (Nevus 66.9%). This imbalance might create potential bias in the segmentation model, creating over-fitting issues. To address these issues, weighted data augmentation approaches were applied during the data pre-processing stage to increase the diversity of the dataset, so that the hybrid segmentation model performance is generalized and suffers from no potential bias.

Chapter 4

Methodology

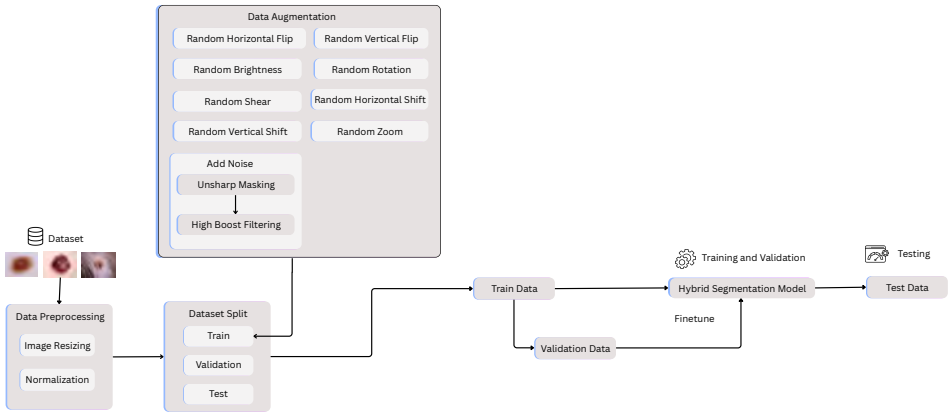


Figure 4.1: Proposed Methodology

This research methodology centers on developing an EfficientNetB7-based hybrid segmentation model to precisely delineate skin lesion areas. Initially, the datasets are collected, and all the images and their corresponding labels are preprocessed using advanced preprocessing methods. After completing the required preprocessing, the datasets are split into train, validation, and test sets following the 80:10:10 ratio, respectively. Afterwards, the hybrid segmentation model is implemented. During training, the segmentation model takes images and their corresponding ground truth masks as input, and then, based on it, generates segmentation masks. After training,the segmentation model's performance is evaluated. If the results are unsatisfactory, then the model was fine tuned by adjusting the learning rate and optimizing the number of filters to improve it's overall performance. Figure 4.1 illustrates the various stages of our research methodology.

In this stage the datasets are preprocessed and augmented to increase their diversity, ensuring that the input data is in an optimal form for training our hybrid segmentation model. During augmentation, input data is normalized to ensure that the

model can learn effectively, resulting in an enhanced segmentation of skin lesions. This data preprocessing stage is necessary for achieving high performance and reliable results in our segmentation model. Additionally, in Figure 4.3, we can see the class count of each lesion type before and after augmentation.

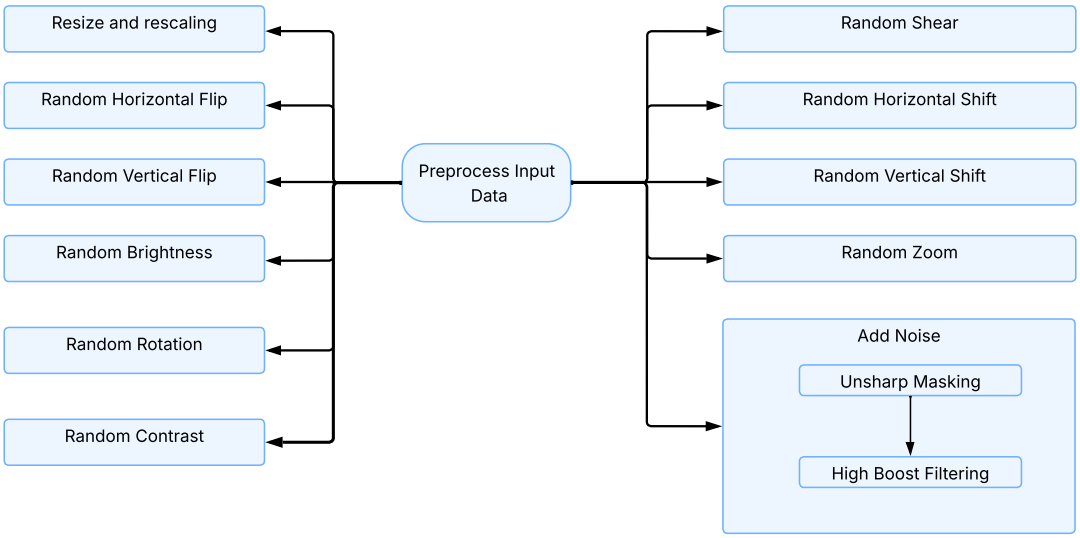


Figure 4.2: Data Pre-processing

Resize Image: Note that the default input size required by EfficientNetB7 and Swin Transformer is 224x224 pixels. All the input images in the dataset are resized to 128x128 pixels. This is done to make the convolution layer operations faster and more efficient. After resizing, all the pixel values of the images were normalized. This normalization is necessary to keep our models numerically stable during the training phase by reducing variance in pixel values.

Random Horizontal and Vertical Flip: Images are randomly flipped horizontally or vertically. This augmentation method was used to simulate real world variations. As some people, might take images of skin lesion affected regions by holding their camera in a horizontal orientation whereas some might hold their camera in a vertical orientation. So this augmentation improves the generalization capabilities of our model and makes it invariant to horizontal and vertical orientation.

Random Rotation: Images are randomly rotated within the range of - 60° degrees to + 60° degrees to introduce rotational variability in the dataset. This augmentation helps simulate real world scenarios where the orientation of the subject may vary depending on how the camera is held or the angle at which the photograph is taken.

Random Brightness: Random brightness was applied on the images within the range of -27% to +27%. This augmentation method helps simulate real-world scenarios where images may be taken under different lighting conditions. It helps the model learn to identify skin lesion features independently of lighting conditions,

enhancing its robustness

Random Contrast: Random contrast was applied to the images within the range of -27% to +27%. This augmentation technique simulates variations in image contrast that may occur due to differences in camera settings or environmental factors. By exposing the model to images with varying contrast levels, it learns to recognize skin lesion features regardless of contrast changes, thereby improving its generalization and robustness.

Width, Height, and Shear Shift: Images are randomly shifted horizontally or vertically by up to 10% of their actual dimensions. This augmentation simulates slight misalignment or off-center captures of the skin lesion region, as often seen in real-world scenarios. Additionally, some images are randomly sheared by up to 10% to simulate slanted or skewed views of the skin lesions. This augmentation helps the model become invariant to subtle distortions caused by the camera angle or movement during image capture, thereby enhancing the models' generalization capabilities.

Filling up the empty part of the canvas by the nearest pixels: After applying horizontal, vertical, or shear shift there might be some empty black regions in the image. These empty black regions were filled up by taking the nearest pixel values from the images and then sampling them. This step prevents the artifacts that could confuse the model during the training phase.

Random Zoom: Images are randomly zoomed in or zoomed out within the range of - 5% to + 5%. This augmentation improves the model's competence to recognize lesions at varying scales and ensures it focuses on key features rather than the absolute size of the lesion in the image.

Noise Injection via Combined Enhancement Techniques: To further enhance generalization, noise was deliberately introduced into the training data using a two step process: (1) Unsharp masking, and (2) High boost filtering. The results of Unsharp masking were combined with those of high boost filtering to simulate noisy, high frequency image content. For each class in the training set, 60% of the samples were augmented with these noise inducing techniques, while the remaining 40% underwent only standard augmentations. This strategy ensures that the hybrid segmentation model is robust and generalize well, performing effectively on both noisy and clean images.

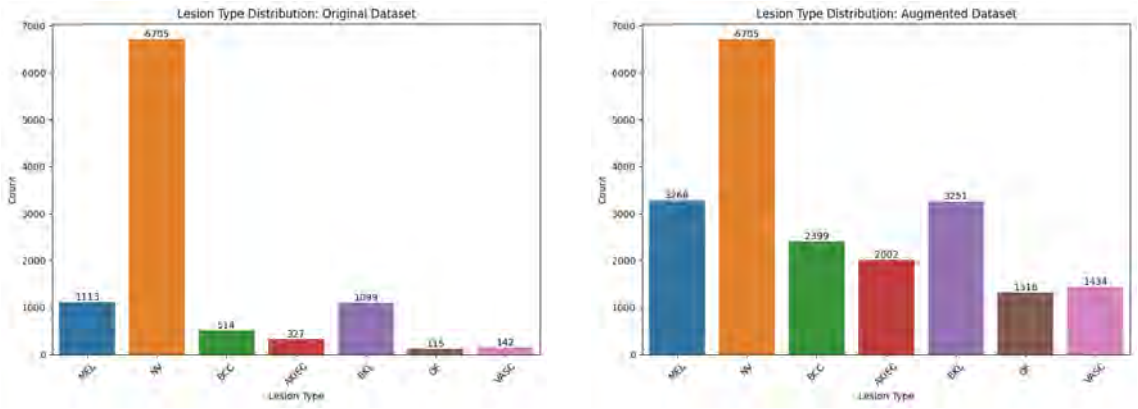


Figure 4.3: Lesion type distribution before and after data augmentation.

4.1 Segmentation Model Architecture

The segmentation model is based on a hybrid architecture that integrates the EfficientNetB7 encoder with a UNet++ decoder, incorporating Convolutional Block Attention Module (CBAM) modules into every skip connection. This architecture was chosen to leverage the strengths of both EfficientNetB7's rich hierarchical feature extraction capabilities and UNet++'s precise localization for pixel-wise segmentation tasks. The proposed hybrid segmentation model was inspired by the S2C DeLeNet model [37], which integrates segmentation and classification tasks using a parameter transfer approach.

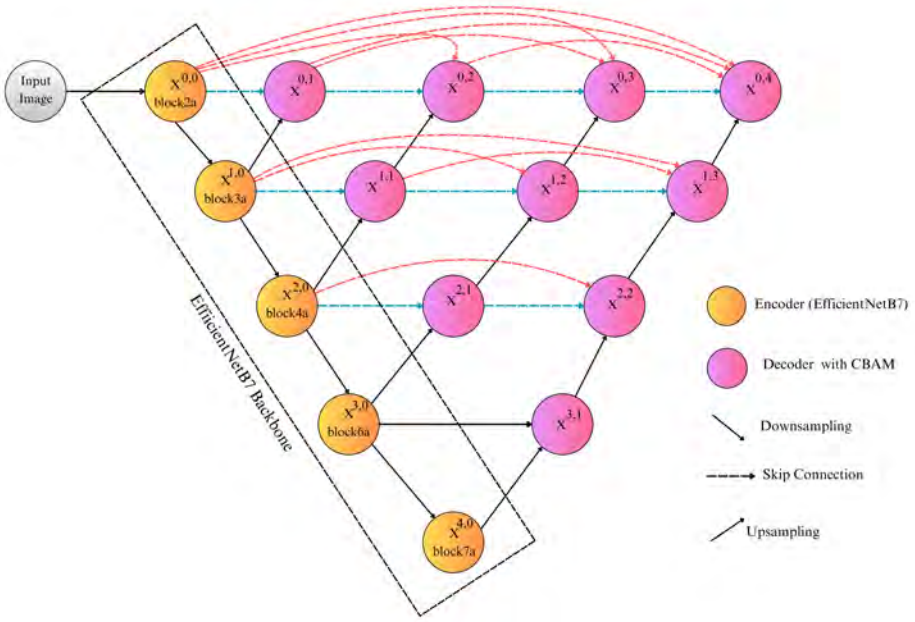


Figure 4.4: Architecture of our proposed Hybrid Segmentation Model

4.1.1 Encoder: EfficientNetB7

EfficientNet models are recognized for their ability to balance accuracy and computational efficiency through a compound scaling method. This method uniformly scales the network's depth, width, and resolution. The EfficientNet family comprises variants from B0 to B7, with EfficientNetB7 being the largest, offering an optimal trade-off between complexity and performance. A pre-trained EfficientNetB7 was employed as the encoder to leverage its robust feature extraction capabilities, derived from training on large image datasets. It was chosen for our hybrid segmentation model due to its efficiency in achieving high accuracy with lesser parameters and lower computational cost contrasted to traditional convolutional neural network (CNN) models such as ResNet50, VGG19, and DenseNet This makes it well-suited for resource-constrained environments. Additionally, EfficientNetB7 excels at handling noisy images more effectively, capturing both low-level and high-level features through its hierarchical feature extraction.

4.1.2 Decoder: UNet++

UNet++ was selected as the decoder for its ability to capture fine-grained spatial details, facilitated by its dense skip connections. These connections enhance feature fusion across different resolution levels, resulting in a more precise segmentation mask compared to its predecessor, UNet.

4.1.3 Skip Connections with CBAM:

CBAM modules are applied to every skip connection, enabling the model to selectively enhance features transferred from the encoder to the decoder. This process allows the decoder to utilize the most informative features, potentially improving the model's capacity to capture critical lesion characteristics and delineate boundaries accurately. CBAM contributes by capturing rich global context information alongside image-specific positional and channel features.

4.1.4 Model Architecture Details:

In this hybrid segmentation model, the encoder extracts feature maps from specific layers of EfficientNetB7 namely, block2a_expand_activation, block3a_expand_activation, block4a_expand_activation, block6a_expand_activation, and block7a_expand_activation which are then utilized in the skip connections of the UNet++ decoder. The decoder comprises a series of transposed convolutional layers for upsampling and convolutional blocks for feature refinement. Skip connections concatenate feature maps from the encoder with the corresponding upsampled features in the decoder. Following concatenation, CBAM modules apply both channel and spatial attention. Each convolutional block in the decoder includes two convolutional layers with ReLU activation, followed by batch normalization. Dropout layers are incorporated after each concatenation step to mitigate overfitting.

4.1.5 Training Configuration:

The model was compiled using the Adam optimizer with a learning rate of 10^{-4}, selected for its adaptive learning capabilities. The Dice coefficient loss function was chosen to quantify the overlap between predicted and actual masks, suiting it well for segmentation tasks. Training incorporated early stopping to pause the process if validation loss showed no further decrease, thus avoiding overfitting. When validation loss remained unchanged over five consecutive epochs, the learning rate was halved by a factor of 0.5. Performance metrics, including accuracy, precision, recall, Dice coefficient, Intersection over Union (IoU), Matthews Correlation Coefficient (MCC), and XOR Error, were monitored throughout training. The model was trained for 100 epochs, with the validation set used to oversee the training process and fine-tune the model. The predicted segmentation mask is generated from the sigmoid function's output.

4.1.6 Parameter Optimization:

The combination of EfficientNetB7, UNet++, and CBAM results in a substantial number of parameters and feature maps, which could potentially overwhelm the model and impede performance optimization. To address this, a 50% dropout was applied after each CBAM module in the skip connections.

This hybrid architecture leverages the strengths of its components. UNet++ provides a robust framework for segmentation, preserving spatial details through dense skip connections. EfficientNetB7 serves as a powerful and efficient feature extractor for the encoder. CBAM modules enhance the skip connections by enabling the network to focus on the most salient features at each scale, refining the information passed to the decoder. This synergy improves segmentation accuracy, particularly for precise boundary delineation, making the model highly effective for complex segmentation tasks.

Chapter 5

Result and Analysis

In this section, a comprehensive analysis of the experimental results obtained from the proposed hybrid segmentation model for real time skin lesion detection is presented. The evaluation is structured to provide a detailed understanding of model performance, benchmarking, and generalization capability across various scenarios.

The discussion begins with hyper-parameter tuning, outlining the rationale behind the selection of key hyper-parameters such as batch size, optimizer configuration, learning rate scheduling, regularization techniques, and loss function. All relevant evaluation metrics employed for segmentation task are subsequently defined.

The results are presented in three main parts. First, the performance of the segmentation models is analyzed, comparing the proposed hybrid model against established architectures such as UNet, UNet++. Comparative results are provided in both tabular and visual formats, including validation curves and qualitative predictions on validation and test sets.

In this section, the results of the proposed hybrid segmentation model are compared with other competing segmentation models. It is noteworthy that the segmentation model was initially trained and fine-tuned on the HAM10000 dataset. Finally, to assess generalization and robustness, the model was also trained and evaluated on external datasets: ISIC 2016, ISIC 2017, MSK, and PH^2.

Through this systematic result analysis, the effectiveness of the proposed model, as well as their capacity to generalize to new, previously unseen data distributions, is demonstrated.

5.1 Evaluation Metrics

Suitable performance evaluation metrics were thoughtfully chosen for the proposed approach. Metrics utilized for the segmentation model have been summarized in Figure 5.1. For the segmentation model, the evaluation metrics reported include accuracy, precision, recall, Dice score, Intersection over Union (IOU), XOR error, and Matthews Correlation Coefficient (MCC). The Dice score is set as the key metric for the segmentation model, with greater values denoting superior model

performance. Notably, XOR error acts as a metric to minimize, where reduced values indicate enhanced results.

Segmentation Evaluation Metric	
Accuracy	$\frac{TP + TN}{TP + TN + FP + FN}$
Precision	$\frac{TP}{TP + FP}$
Recall	$\frac{TP}{TP + FN}$
Dice Score	$\frac{2 \times TP}{(TP + FP) + (TP + FN)}$
IoU	$\frac{TP}{(TP + FP) + (TP + FN)}$
XOR Error	$\frac{XOR(\text{Proposed Mask, Ground Truth Mask})}{\text{Area of Ground Truth Mask}}$
XOR Error	$\frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}$

Figure 5.1: Evaluation Metrics

Here, n = class index, N = total number of classes, TP = true positive, TN = true negative, FP = false positive, FN = false negative.

5.2 Hyper-parameter tuning

Hyper-parameter tuning is a critical process in the development of deep learning models, as it involves selecting appropriate values for the model's hyper parameters to maximize performance, minimize overfitting issues, and reduce inference time. Proper hyper-parameter optimization can substantially affect the model's ability to generalize to unseen data and its computational efficiency. In this thesis, we systematically explored and set the following hyper-parameters for our experiments:

Batch Size: Batch sizes ranging from 1 to 16 was explored. Although larger batch sizes generally contribute to more stable training, hardware limitations specifically GPU VRAM constraints necessitated the selection of a batch size of 8. Consequently, eight images were processed per training step in each epoch.

Optimizer: The Adam optimizer was used for the segmentation model. Adam is widely known for its adaptive learning rate and momentum capabilities, which facilitates faster convergence and improved generalization compared to other traditional optimizers. The weight update rule for Adam is expressed as:

$$w_n = w_{n-1} - \delta_{n-1}\nabla\phi(w_{n-1}) + \beta_{n-1}(w_{n-1} - w_{n-2}) \tag{1}$$

$$\beta_k = \beta_1 \frac{1-\beta_1^{k-1}}{1-\beta_1^k} \tag{2}$$

$$\delta_k = \delta \frac{1-\beta_1}{1-\beta_1^k} \tag{3}$$

w_n = Weights of the n-th epoch from the network
δ = Initially provided learning rate
δ_k = Decayed learning rate at k-th epoch
β_1 = Momentum estimation parameter (set to 0.99)
$\nabla\phi(w_{n-1})$ = Computed gradient from the loss function ϕ at w_{n-1} weights

Loss Function: For the segmentation model, Dice loss was selected due to its effectiveness in medical image segmentation, The formula for dice loss and categorical cross entropy loss is as follows:

$$L_{DL} = 1 - D(y, \hat{y}) \tag{4}$$

$$D(y, \hat{y}) = 2 \times \frac{\sum y \cdot \hat{y}}{\sum y + \sum \hat{y}} \tag{5}$$

L_{DL} = Dice Loss
$D(y, \hat{y})$ = Dice Score
y = Ground truth probability
$\hat{y}$ = Predicted probability

Learning Rate: Various initial learning rates, including 10^{-2}, 10^{-4}, 10^{-5}, and 10^{-6} were evaluated. Optimal results were obtained with a learning rate of 10^{-4}. Higher rates caused instability, whereas lower rates slowed the convergence process. Given Adam's ability to adjust the learning rate dynamically during training, a learning rate scheduler was also implemented to decrease the rate when training loss stabilized, assessed using Dice score for segmentation.

Regularization: To mitigate overfitting issues , L2 regularization was applied with a coefficient of 10^{-3}. This approach penalizes large weights and encourages model generalization.

5.3 Quantitative Segmentation Results

To evaluate the proposed segmentation model, it is compared with state-of-the-art models, including baseline UNet and UNet++, and a variant of the proposed model without the Convolutional Block Attention Module (CBAM), using the HAM10000 training and validation datasets. The proposed model, leveraging EfficientNetB7, UNet++, and CBAM, was also trained and evaluated on the ISIC2016, ISIC2017,

MSK, and PH^2 datasets to assess its generalization across diverse datasets. Performance metrics, including Dice coefficient, Intersection over Union (IoU), accuracy, loss, precision, recall, Matthews Correlation Coefficient (MCC), XOR error, Area Under the Curve (AUC), and macro F1 score, were computed on the corresponding validation and test splits. The comparative results without noise induction are presented in Tables 5.1, 5.3, 5.5, 5.7, and 5.9, with the CBAM variant achieving a Dice coefficient of 0.9484 and IoU of 0.9029 on the HAM10000 test set, a Dice coefficient of 0.9192 and IoU of 0.8532 on the ISIC2016 test set, and a Dice coefficient of 0.9203 and IoU of 0.8525 on the PH^2 test set. To assess robustness to noise, Tables 5.2, 5.4, 5.6, 5.8, and 5.10 present the performance on the HAM10000, ISIC2016, ISIC2017, MSK, and PH^2 datasets, respectively, after noise induction, with the CBAM-removed variant achieving a Dice coefficient of 0.9471 and IoU of 0.9004 on the HAM10000 test set, the CBAM variant achieving a Dice coefficient of 0.9218 and IoU of 0.8566 on the ISIC2016 test set, a Dice coefficient of 0.8284 and IoU of 0.7167 on the ISIC2017 test set, a Dice coefficient of 0.8869 and IoU of 0.8018 on the MSK test set, and a Dice coefficient of 0.8924 and IoU of 0.8058 on the PH^2 test set.

Foundation	Models	Acc	Loss	Prec	Rec	Dice	IoU	MCC	XORerr	AUC	Macro F1
HAM10000 Validation Set											
State of the Art	UNet	0.9628	0.0789	0.9353	0.9378	0.9387	0.8677	0.9155	0.1128	0.9644	0.9612
	UNet++	0.9605	0.0864	0.9456	0.9433	0.9322	0.8828	0.9100	0.1135	0.9666	0.9566
Introduced Networks	Proposed Model	**0.9703**	**0.0594**	**0.9476**	**0.9433**	**0.9442**	**0.8957**	**0.9243**	**0.1117**	**0.9683**	**0.9616**
	Proposed Model excluding CBAM	0.9702	0.0602	0.9453	**0.9451**	**0.9442**	0.8956	0.9241	0.1118	0.9692	**0.9616**
HAM10000 Test Set											
State of the Art	UNet	0.9580	0.0898	0.9400	0.9356	0.9382	0.8801	0.9023	0.1098	0.9699	0.9630
	UNet++	0.9678	0.0659	0.9373	0.9420	0.9368	0.7977	0.9015	0.1068	0.9702	0.9639
Introduced Networks	Proposed Model	**0.9730**	**0.0558**	**0.9560**	0.9431	**0.9484**	**0.9029**	**0.9484**	**0.1029**	**0.9704**	**0.9649**
	Proposed Model excluding CBAM	0.9723	0.0576	0.9516	**0.9446**	0.9471	0.9004	0.9288	0.1054	**0.9706**	0.9640

Table 5.1: Skin Lesion Segmentation Performance on HAM10000K (No Noise)

Foundation	Models	Acc	Loss	Prec	Rec	Dice	IoU	MCC	XORerr	AUC	Macro F1
HAM10000 Validation Set (Noise-Induced)											
State of the Art	UNet	0.9623	0.0997	0.9278	0.9350	0.9300	0.8713	0.9046	0.1412	0.9544	0.9516
	UNet++	0.9605	0.0737	0.9281	0.9283	0.9259	0.8648	0.9000	0.1474	0.9538	0.9490
Introduced Networks	Proposed Model	0.9702	**0.0607**	**0.9518**	0.9391	0.9442	0.8955	0.9245	**0.1109**	0.9680	0.9618
	Proposed Model excluding CBAM	**0.9708**	0.0610	0.9502	**0.9411**	**0.9445**	**0.8961**	**0.9250**	0.1106	**0.9691**	**0.9620**
HAM10000 Test Set (Noise-Induced)											
State of the Art	UNet	0.9679	0.0619	0.9424	0.9363	0.9384	0.8853	0.9171	0.1227	0.9592	0.9581
	UNet++	0.9671	0.0630	0.9366	0.9403	0.9373	0.8834	0.9156	0.1255	0.9622	0.9573
Introduced Networks	Proposed Model	0.9724	**0.0583**	**0.9581**	0.9380	0.9468	0.9001	0.9289	**0.1005**	**0.9697**	0.9639
	Proposed Model excluding CBAM	**0.9725**	0.0588	**0.9581**	**0.9383**	**0.9471**	**0.9004**	**0.9291**	0.1046	0.9695	**0.9641**

Table 5.2: Skin Lesion Segmentation Performance on HAM10000 (With Noise)

Foundation	Models	Acc	Loss	Prec	Rec	Dice	IoU	MCC	XORerr	AUC	Macro F1
ISIC2016 Test Set											
State of the Art	UNet	0.9479	0.1356	0.9103	0.9127	0.9112	0.8301	0.8798	0.1835	0.9610	0.9355
	UNet++	0.9552	0.1184	0.9192	0.9205	0.9198	0.8443	0.8881	0.1720	0.9655	0.9415
Introduced Networks	Proposed Model	**0.9619**	**0.1069**	**0.9280**	0.9260	**0.9192**	**0.8532**	**0.8983**	**0.1671**	**0.9743**	0.9482
	Proposed Model excluding CBAM	0.9601	0.1070	0.9141	**0.9397**	0.9074	0.8330	0.8968	0.1947	0.9473	**0.9776**

Table 5.3: Skin Lesion Segmentation Performance on ISIC2016 (No Noise)

Foundation	Models	Acc	Loss	Prec	Rec	Dice	IoU	MCC	XORerr	AUC	Macro F1
ISIC2016 Test Set (Noise-Induced)											
State of the Art	UNet	0.9445	0.1429	0.9012	0.9098	0.9026	0.8257	0.8722	0.1951	0.9581	0.9291
	UNet++	0.9523	0.1215	0.9136	0.9184	0.9159	0.8395	0.8839	0.1822	0.9634	0.9378
Introduced Networks	Proposed Model	**0.9614**	**0.1037**	0.9218	0.9346	**0.9218**	**0.8566**	0.8991	0.1602	**0.9738**	0.9487
	Proposed Model excluding CBAM	0.9612	0.1062	**0.9362**	**0.9383**	0.9208	0.8548	**0.9004**	**0.1599**	0.9726	**0.9493**

Table 5.4: Skin Lesion Segmentation Performance on ISIC2016 (With Noise)

Foundation	Models	Acc	Loss	Prec	Rec	Dice	IoU	MCC	XORerr	AUC	Macro F1
ISIC2017 Validation Set											
State of the Art	UNet	0.9410	0.2225	0.8880	0.7912	0.8210	0.7010	0.8042	0.3620	0.9125	0.9012
	UNet++	0.9465	0.2072	0.8955	0.8073	0.8290	0.7009	0.8115	0.3408	0.9189	0.9067
Introduced Networks	Proposed Model	0.9504	0.2005	0.8984	0.8141	0.8314	0.7175	0.8203	0.3266	**0.9217**	0.9065
	Proposed Model excluding CBAM	**0.9552**	**0.1580**	**0.9104**	**0.8209**	**0.8532**	**0.7487**	**0.8326**	**0.2768**	0.9144	**0.9134**
ISIC2017 Test Set											
State of the Art	UNet	0.9291	0.2380	0.9223	**0.7989**	**0.8801**	0.6901	0.7854	0.3410	0.9114	0.8802
	UNet++	0.9342	0.2245	0.9288	0.7425	0.8081	0.6996	0.7910	0.3331	0.9162	0.8889
Introduced Networks	Proposed Model	0.9341	0.2161	**0.9379**	0.7582	0.8153	0.6981	0.7989	0.3299	**0.9200**	0.8923
	Proposed Model excluding CBAM	**0.9373**	**0.1752**	0.9329	0.7780	0.8354	**0.7266**	**0.8089**	**0.2943**	0.8984	**0.9054**

Table 5.5: Skin Lesion Segmentation Performance on ISIC2017 (No Noise)

Foundation	Models	Acc	Loss	Prec	Rec	Dice	IoU	MCC	XORerr	AUC	Macro F1
ISIC2017 Validation Set (Noise-Induced)											
State of the Art	UNet	0.9372	0.1850	0.8855	0.7871	0.8144	0.6958	0.8123	0.3172	0.8932	0.9024
	UNet++	0.9448	0.1392	0.9010	0.7998	0.8322	0.7163	0.8198	0.2985	0.9044	0.9083
Introduced Networks	Proposed Model	**0.9508**	**0.1055**	0.9117	0.8105	0.8484	0.7435	0.8255	0.2823	**0.9038**	0.9082
	Proposed Model excluding CBAM	0.9499	0.1648	**0.9127**	**0.8071**	**0.8456**	**0.7391**	**0.8238**	**0.2850**	0.8986	**0.9072**
ISIC2017 Test Set (Noise-Induced)											
State of the Art	UNet	0.9284	0.2221	0.9191	0.7455	0.7971	0.6958	0.7924	0.3281	0.8894	0.8888
	UNet++	0.9335	0.1989	0.9302	0.7497	0.8127	0.7054	0.7992	0.3145	0.8935	0.8909
Introduced Networks	Proposed Model	0.9348	0.1946	**0.9396**	0.7582	**0.8284**	**0.7167**	0.8006	**0.3024**	0.8975	0.8931
	Proposed Model excluding CBAM	**0.9353**	**0.1815**	0.9373	**0.7612**	0.8283	**0.7167**	**0.8015**	**0.3024**	**0.8976**	**0.8938**

Table 5.6: Skin Lesion Segmentation Performance on ISIC2017 (With Noise)

Foundation	Models	Acc	Loss	Prec	Rec	Dice	IoU	MCC	XORerr	AUC	Macro F1
MSK Validation Set											
State of the Art	UNet	0.9448	0.1206	0.8751	0.9130	0.8714	0.8213	0.8651	0.1894	0.9615	0.9309
	UNet++	0.9532	0.1122	0.8829	0.9341	0.8926	0.8321	0.8715	0.1833	0.9609	0.9405
Introduced Networks	Proposed Model	**0.9570**	**0.1000**	**0.8942**	**0.9516**	**0.9199**	**0.8528**	**0.8917**	**0.1654**	**0.9658**	**0.9450**
	Proposed Model excluding CBAM	0.9472	0.1172	0.8709	0.9488	0.8956	0.8139	0.8719	0.2182	0.9617	0.9340
MSK Test Set											
State of the Art	UNet	0.9288	0.1566	0.8698	0.8826	0.8629	0.7659	0.8351	0.2691	0.9406	0.9168
	UNet++	0.9343	0.1486	0.8761	0.8973	0.8764	0.7792	0.8425	0.2435	0.9405	0.9100
Introduced Networks	Proposed Model	**0.9385**	**0.1347**	**0.8791**	0.9065	**0.8861**	**0.8033**	**0.8467**	**0.2245**	**0.9407**	**0.9204**
	Proposed Model excluding CBAM	0.9342	0.1404	0.8571	**0.9205**	0.8729	0.7798	0.8388	0.2628	0.9471	0.9162

Table 5.7: Skin Lesion Segmentation Performance on MSK (No Noise)

Foundation	Models	Acc	Loss	Prec	Rec	Dice	IoU	MCC	XORerr	AUC	Macro F1
MSK Validation Set (Noise-Induced)											
State of the Art	UNet	0.9400	0.1102	0.8648	0.9126	0.8886	0.8026	0.8699	0.1989	0.9600	0.9342
	UNet++	0.9521	0.1195	0.8753	0.9345	0.9041	0.8245	0.8734	0.1867	0.9615	0.9347
Introduced Networks	Proposed Model	**0.9523**	0.1083	**0.8821**	0.9488	**0.9108**	**0.8382**	**0.8807**	**0.1846**	**0.9623**	**0.9386**
	Proposed Model excluding CBAM	0.9487	**0.1061**	0.8648	**0.9567**	0.9046	0.8269	0.8738	0.2019	**0.9623**	0.9352
MSK Test Set (Noise-Induced)											
State of the Art	UNet	0.9288	0.1522	0.8659	0.8826	0.8677	0.7659	0.8376	0.2684	0.9400	0.9169
	UNet++	0.9333	0.1429	0.8642	0.9056	0.8802	0.7888	0.8426	0.2486	0.9389	0.9189
Introduced Networks	Proposed Model	**0.9374**	**0.1343**	**0.8650**	**0.9207**	**0.8869**	**0.8018**	**0.8456**	**0.2308**	**0.9454**	**0.9201**
	Proposed Model excluding CBAM	0.9328	0.1353	0.8579	0.9113	0.8756	0.7853	0.8342	0.2502	0.9370	0.9136

Table 5.8: Skin Lesion Segmentation Performance on MSK (With Noise)

Foundation	Models	Acc	Loss	Prec	Rec	Dice	IoU	MCC	XORerr	AUC	Macro F1
PH^2 Validation Set											
State of the Art	UNet	0.9407	0.1956	0.9200	0.9357	0.8592	0.7982	0.8700	0.3325	0.9800	0.9432
	UNet++	0.9458	0.1900	0.9298	0.9348	0.8793	0.8279	0.8879	0.2619	0.9828	0.9502
Introduced Networks	Proposed Model	0.9551	0.1886	0.9322	**0.9401**	**0.8954**	**0.8109**	0.8923	**0.2194**	0.9837	0.9446
	Proposed Model excluding CBAM	**0.9609**	**0.1709**	**0.9505**	0.9353	0.8489	0.7404	**0.9038**	0.3338	**0.9838**	**0.9508**
PH^2 Test Set											
State of the Art	UNet	0.9549	0.1629	0.9389	0.9562	0.9115	0.8425	0.9056	0.1700	0.9875	0.9575
	UNet++	0.9552	0.1631	0.9375	0.9502	0.9102	0.8426	0.9058	0.1680	0.9888	0.9570
Introduced Networks	Proposed Model	**0.9590**	0.1639	0.9416	**0.9626**	**0.9203**	**0.8525**	**0.9187**	**0.1656**	**0.9908**	**0.9590**
	Proposed Model excluding CBAM	0.9581	**0.1477**	**0.9435**	0.9580	0.8771	0.7814	0.9169	0.2667	0.9885	0.9582

Table 5.9: Skin Lesion Segmentation Performance on PH^2 (No Noise)

Foundation	Models	Acc	Loss	Prec	Rec	Dice	IoU	MCC	XORerr	AUC	Macro F1
PH^2 **Validation Set (Noise-Induced)**											
State of the Art	UNet	0.9408	0.3216	0.9256	0.9146	0.8426	0.7489	0.8596	0.4555	0.9802	0.9420
	UNet++	0.9551	0.3156	0.9365	0.9149	0.8576	0.7486	0.8822	0.3156	0.9826	0.9426
Introduced Networks	Proposed Model	0.9555	0.2506	**0.9473**	0.9230	**0.8672**	**0.7658**	0.8897	**0.2774**	0.9863	0.9430
	Proposed Model excluding CBAM	**0.9537**	**0.2539**	0.9257	**0.9421**	0.7860	0.6512	**0.8922**	0.5054	**0.9892**	**0.9452**
PH^2 **Test Set (Noise-Induced)**											
State of the Art	UNet	0.9578	0.2356	0.9389	0.9384	0.8525	0.7880	0.8978	0.2568	0.9802	0.9405
	UNet++	0.9572	0.2362	0.9452	0.9387	0.8658	0.7798	0.8988	0.2456	0.9888	0.9432
Introduced Networks	Proposed Model	**0.9582**	0.2233	**0.9493**	0.9471	**0.8924**	**0.8058**	**0.9122**	**0.2230**	0.9905	**0.9557**
	Proposed Model excluding CBAM	0.9529	**0.2040**	0.9292	**0.9559**	0.8225	0.6987	0.9016	0.3993	**0.9895**	0.9503

Table 5.10: Skin Lesion Segmentation Performance on PH^2 (With Noise)

5.3.1 Standard Augmentation Approach

From the tabular data, it can be analyzed that the proposed model with CBAM generally outperforms the proposed model without CBAM and state-of-the-art baseline models (U-Net and U-Net++) in Dice Score and IoU across the HAM10000, ISIC-2016, ISIC-2017, MSK, and PH^2 datasets. In the HAM10000 (Test) dataset, the CBAM model achieved a Dice Score of 0.9484 and IoU of 0.9029, slightly better than the non-CBAM model's 0.9471 and 0.9004, and significantly surpassing U-Net's 0.9382 and 0.8801, and U-Net++'s 0.9368 and 0.7977. For ISIC-2016 (Test), the CBAM model recorded a Dice Score of 0.9192 and IoU of 0.8532, outperforming the non-CBAM model's 0.9074 and 0.8330, U-Net's 0.9112 and 0.8301, and U-Net++'s 0.9198 and 0.8443. However, in ISIC-2017 (Test), the CBAM model achieved a Dice Score of 0.8153 and IoU of 0.6981, compared to 0.8354 and 0.7266 for the non-CBAM model, with the latter leading in IoU. In MSK (Test), the CBAM model excelled with a Dice Score of 0.8861 and IoU of 0.8033, surpassing the non-CBAM model's 0.8729 and 0.7798. For PH^2 (Test), the CBAM model led with a Dice Score of 0.9203 and IoU of 0.8525, compared to 0.9115 and 0.8425 for U-Net, 0.8426 and 0.9058 for U-Net++, and 0.8771 and 0.7814 for the non-CBAM model.

For MCC, Macro F1, AUC, and Recall, the CBAM model displayed superior performance compared to the other models. In HAM10000 (Test), the CBAM model's MCC was 0.9484 and Macro F1 was 0.9649, compared to 0.9288 and 0.9640 for the non-CBAM model, with AUC values of 0.9704 and 0.9706, respectively. In ISIC-2016 (Test), the CBAM model's MCC was 0.8983 and Macro F1 was 0.9482, outperforming the non-CBAM model's 0.8968 and 0.9776. In ISIC-2017 (Validation), the CBAM model's Recall was 0.8141, slightly lower than the non-CBAM model's 0.8209, while its MCC of 0.8203 was slightly below the non-CBAM's 0.8326. For Precision, Accuracy, Loss, and XOR Error, the CBAM model performed comparably. In PH^2 (Test), the CBAM model's Precision was 0.9416 versus 0.9435 for the non-CBAM model, with Accuracy at 0.9590 versus 0.9581. Loss and XOR Error were competitive, with the CBAM model achieving 0.1639 and 0.1656, compared to 0.1477 and 0.2667 for the non-CBAM model. Overall, the CBAM model's consistent improvements in Dice Score and IoU, along with robust performance in MCC, Macro F1, AUC, and Recall, demonstrate its effectiveness for skin lesion image segmentation, although the non-CBAM model occasionally outperforms it in specific metrics, such as IoU in ISIC-2017.

5.3.2 Noise Induced Augmentation Approach

To evaluate the models' effectiveness in real-life scenarios where noise is prevalent, they were trained and tested on the same datasets with added noise. In these noise-induced datasets, the proposed model with CBAM generally outperformed the proposed model without CBAM and state-of-the-art models in Dice Score and IoU, with additional advantages in secondary segmentation metrics. In HAM10000 (noise-induced test), the non-CBAM model slightly outperformed with a Dice Score of 0.9471 and IoU of 0.9004, compared to 0.9468 and 0.9001 for the CBAM model. Both models surpassed U-Net++'s 0.9373 and 0.8834. For ISIC-2016 (noise-induced test), the CBAM model recorded a Dice Score of 0.9218 and IoU of 0.8566, outperforming the non-CBAM model's 0.9208 and 0.8548, and U-Net++'s 0.9159 and 0.8395. In ISIC-2017 (noise-induced test), the CBAM model recorded a Dice Score of 0.8284 and IoU of 0.7167, closely matching the non-CBAM model's 0.8283 and 0.7167, and exceeding U-Net's 0.7971 and 0.6958, and U-Net++'s 0.8127. For MSK (noise-induced test dataset), the CBAM model led with a Dice Score of 0.8869 and IoU of 0.8018, compared to 0.8756 and 0.7853 for the non-CBAM model. In PH^2 (noise-induced test dataset), the CBAM model excelled with a Dice Score of 0.8924 and an IoU of 0.8058, better than the non-CBAM model's 0.8225 and 0.6987.

For MCC, Macro F1, AUC, and Recall, the CBAM model showed robust performance. In HAM10000 (noise-induced test dataset), the CBAM model's MCC was 0.9289 and Macro F1 was 0.9639, closely matching the non-CBAM model's 0.9291 and 0.9641, with AUC values of 0.9697 and 0.9695, respectively. In ISIC-2016 (noise-induced test dataset), the CBAM model's MCC was 0.8991 and Macro F1 was 0.9487, compared to 0.9004 and 0.9493 for the non-CBAM model. In ISIC-2017 (Validation with Noise), the CBAM model's Recall was 0.8105, slightly higher than the non-CBAM model's 0.8071, with MCC at 0.8255 compared to 0.8238. For Precision, Accuracy, Loss, and XOR Error, the CBAM model remained competitive. In PH^2 (noise-induced test dataset), the CBAM model's Precision was 0.9493 compared to 0.9292 for the non-CBAM model, with Accuracy at 0.9582 versus 0.9529. Loss and XOR Error were favourable, with the CBAM model achieving 0.2233 and 0.2230, compared to 0.2040 and 0.3993 for the non-CBAM model.

It can be inferred that while the CBAM model generally outperforms in Dice Score, IoU, MCC, Macro F1, AUC, and Recall under noisy conditions, the non-CBAM model still shows occasional advantages in specific cases, such as the noise-induced HAM10000 dataset. Nevertheless, the inclusion of CBAM consistently enhances segmentation performance particularly in challenging, noise heavy environments making it a robust and reliable choice for medical imaging applications.

5.4 Visual Segmentation Results

Figure 5.2 presents a comparative analysis of the proposed skin lesion segmentation model against three state-of-the-art models, including UNet and UNet++, evaluated using the Dice coefficient. As observed in Figure 5.2, the proposed model outper-

forms baseline models such as UNet and UNet++. Additionally, the proposed model excluding the Convolutional Block Attention Module (CBAM) performs reasonably well, though its results are approximately 1-2% lower than the full proposed model. The integration of noise-injection based augmentation and CBAM enables the proposed model to detect lesion edges with greater precision and refinement compared to other models. Notably, some ground truth segmentation masks contain slight inaccuracies, such as over annotation of the affected region beyond the actual lesion area. However, the proposed model's results demonstrate that it focuses almost perfectly on the lesion area alone, mitigating these annotation errors. To illustrate the training progression of the proposed model, Figures 5.3, 5.4 , 5.5,5.6, 5.7 depict the evolution of key evaluation metrics, including accuracy, loss, Dice coefficient, Intersection over Union (IoU), precision, recall, Matthews Correlation Coefficient (MCC), macro F1 score, and XOR error, over training epochs.

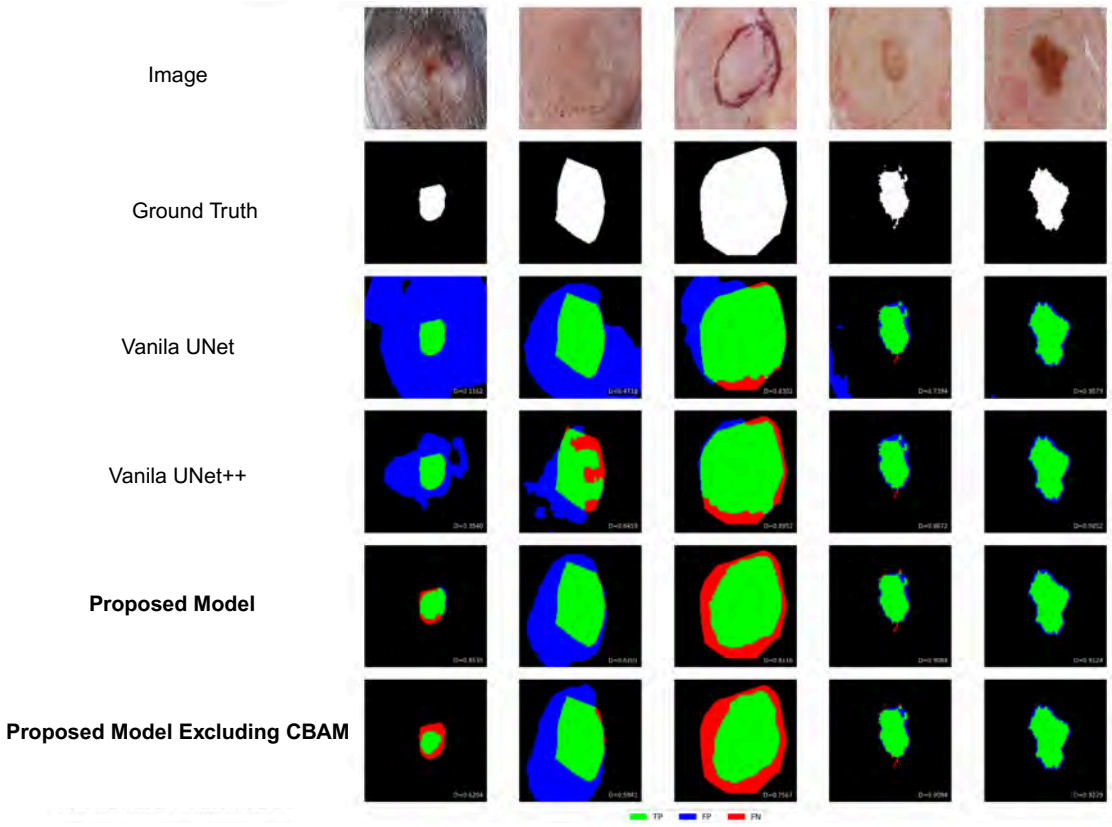


Figure 5.2: Comparative Analysis of Skin Lesion Segmentation Models

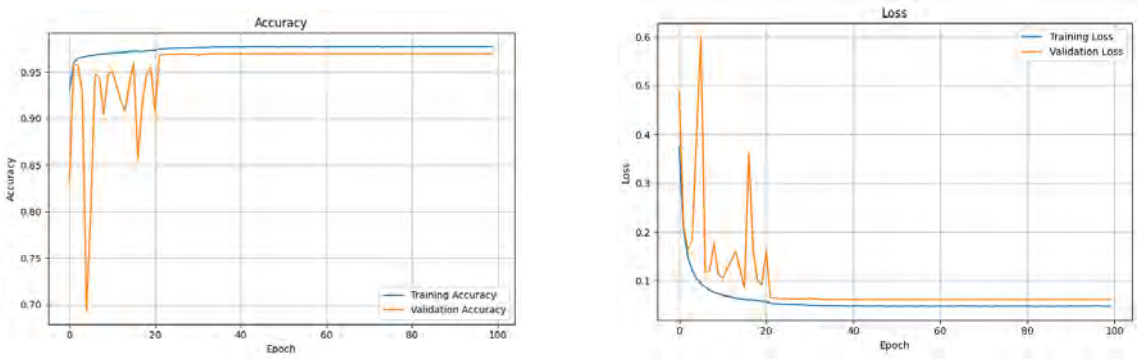


Figure 5.3: Segmentation Model Curve for Accuracy and Loss

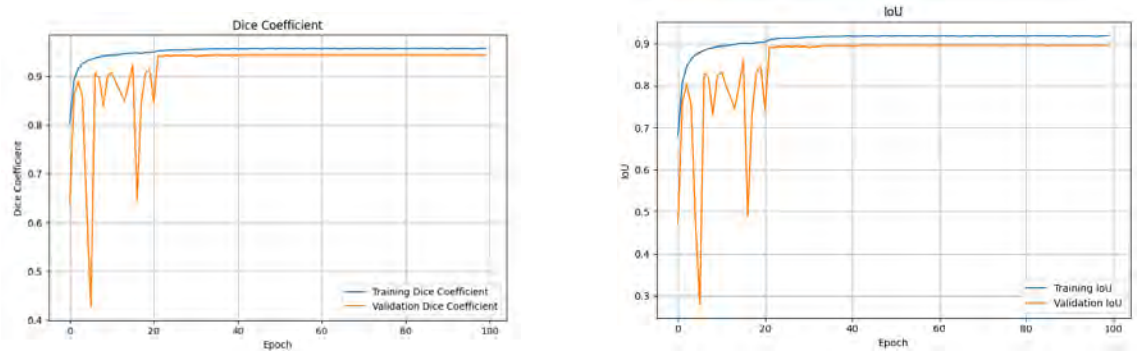


Figure 5.4: Segmentation Model Curve for Dice Coefficient and IOU

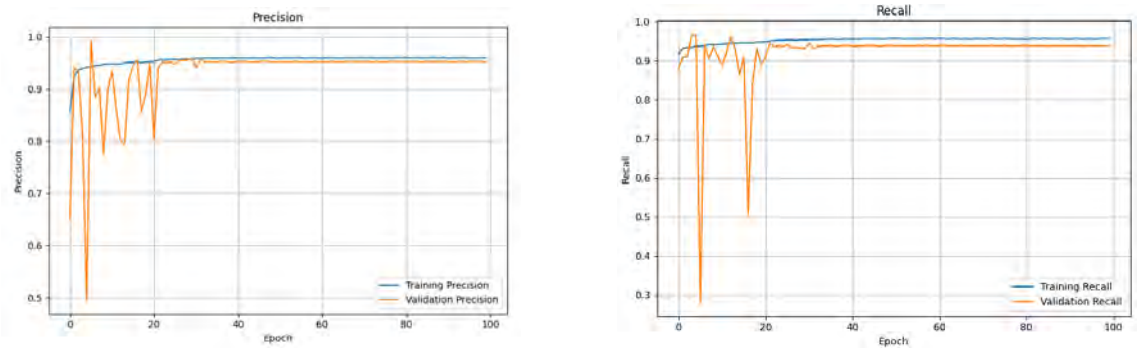


Figure 5.5: Segmentation Model Curve for Precision and Recall

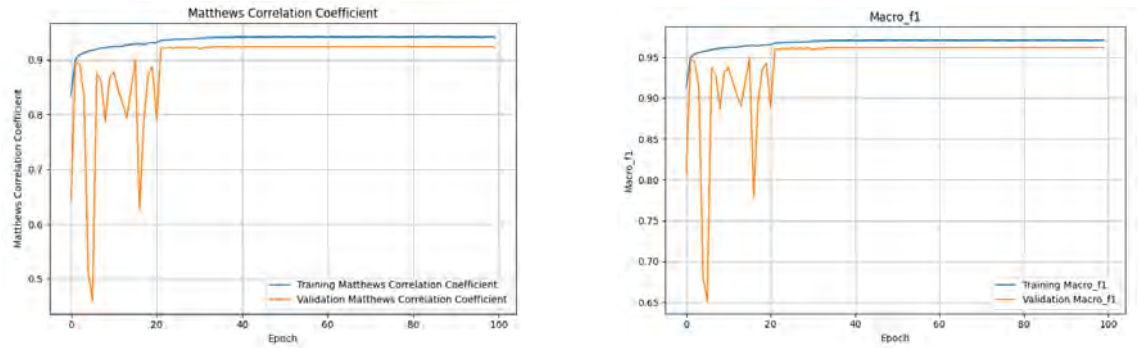


Figure 5.6: Segmentation Model Curve for MCC and Macro F1

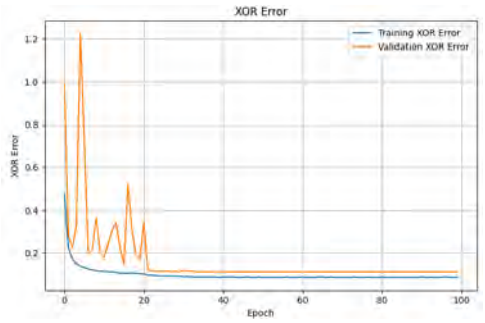


Figure 5.7: Segmentation Model Curve for XOR Error

5.5 Comparison with Existing Methods

In this part, we compare our proposed methodology to various recent deep learning algorithms for skin cancer analysis. The proposed model, integrates a hybrid segmentation architecture comprising an EfficientNetB7 encoder, a UNet++ decoder with Convolutional Block Attention Module (CBAM) in skip connections. This study evaluates its performance against state-of-the-art methods published in 2024 and earlier, focusing on dataset utilization, architectural innovations, and key contributions. The comparison highlights the proposed model's noise robustness, lightweight design, and computational efficiency, validated across diverse datasets including HAM10000, ISIC2016-2018, and PH^2. This analysis provides a comprehensive benchmark to underscore the advancements and potential of our approach in the domain of automated skin lesion analysis.

Reference (Author, Year)	Approach/Model	Task(s)	Dataset(s)	Key Metrics Mentioned	Noise Added or Not	Contribution/Highlights
Himel et al. (2024)	Vision Transformer	Segmentation, Classification	HAM10000	Segmentation: IoU(0.9601), Dice Score (0.98.14) Classification: Accuracy (0.9615)	Preprocessing includes normalization	Utilizes Vision Transformer for skin cancer classification
Duque-Vazquez et al. (2024)	Digital Image Processing	Segmentation	EMPIAR-10478, GitHub dataset	IoU (0.9702), Dice Score (0.9801), Macro F1 (98.49)	Digital hair removal (DHR) algorithm, Random Rotation, Random Zoom, Random Brightness	Image processing algorithms for cell segmentation, alternative to deep learning
Abedini (2024)	Segment Anything Model (SAM)	Segmentation, Classification	HAM10000	IoU (0.7855)	Not specified	Leverages SAM for skin lesion segmentation, emphasizes early detection
Elgohary et al. (2024)	Deep Learning	Classification, Segmentation	HAM10000, PH2	Accuracy (0.95), Precision (0.95), Recall (0.95)	Not specified	General approach for skin cancer classification and segmentation
Khan et al. (2024)	DeepSkinFormer (DSF) with SEEM	Segmentation	HAM10000, ISIC2017, PH2	Dice (0.8452), mIoU (0.7604), Precision (0.8675), Recall (0.8679)	Mentions handling "interference from background noise" using Skin Edge Enhancement Module (SEEM)	End-to-end framework for skin lesion segmentation with edge enhancement
Imran et al. (2024)	Transformer-based framework	Multi-class Segmentation	Custom Dataset	Macro F1 (0.908), Accuracy (0.831), mIoU (0.653)	Not specified	Transformer-based framework for multi-class segmentation of histopathology images
Toprak & Aruk (2024)	Hybrid CNN (DeepLabV3+, MobileNetV2, EfficientNetB0, DenseNet201, KNN)	Classification, Segmentation	ISIC2018, ISIC2019, PH2	F1 score (93.49%), accuracy (94.42%)	Not specified	Novel hybrid deep-learning model with segmentation and feature selection
Mahin et al. (2024)	SSO, ECOC, Weighted Hamming Distance (ML, Optimization)	Segmentation, Classification	ISIC2016, ISIC2017	Accuarcy (0.9517)	Data pre-processing involves removal of unnecessary noise artifact	Integrates optimization methods with supervised learning for diagnosis
Sonia et al. (2024)	Fruit Fly Optimization Algorithm, ML Framework	Segmentation, Classification	Xiangyaderm, Dermnet NZ,HAM10000	Segmentation: Accuracy (0.98), IoU (95.5) Classification: Accuarcy (0.985)	Data pre-processing involves removal of unnecessary noise and hair artifact	Uses a fruit fly optimization algorithm with an ML framework
Cai et al. (2024)	Deformable Attention Transformer U-Net	Segmentation	ISIC2016, ISIC2017, ISIC2018, PH2	Accuracy (0.9626), IoU (0.8343), Dice Score (0.9008)	Not specified	Intelligent skin lesion segmentation using a deformable attention Transformer U-Net
Alam et al. (202X)	S2C-DeLeNet (CNN with EfficientNet-B4)	Segmentation, Classification	HAM10000, ISIC2016, ISIC2017, ISIC2018 (MSK), PH2	Segmentation: Dice Score (0.9467), IoU(0.9007), Precision (0.9593), Recall (0.9373) Classification: Accuracy (0.9058), F1 Score (0.8390), Precision (0.8444), Recall (0.8336)	Geometric Transformations (rotation, flip, shift)	Proposes a parameter transfer based segmentation-classification integration for detecting skin cancer lesions
Proposed Model	Segmentation: Hybrid Model (EfficientNetB7, UNet++, CBAM)	Segmentation	HAM10000, ISIC2016, ISIC2017, MSK, PH2	Dice Score (0.9468), IoU(0.9001), Precision (0.9581), Recall (0.9380), F1 Score (0.9639)	Yes: Unsharp Masking, High Boost Filtering for noise injection (60% of training data), Geometric Transformations (rotation, flip, shift, zoom), Color Adjustments (brightness, contrast)	Proposes a lightweight and computationally efficient Hybrid Segmentation Model (EfficientNetB7 encoder, UNet++ decoder, CBAM in skip connections) using Noise Robust Augmentation Strategy (noise induction and image enhancement techniques);

Figure 5.8: Summary of Recent Deep Learning Models for Skin Cancer Analysis and Proposed Model

The proposed methodology outperforms existing approaches through its advanced hybrid architecture, combining EfficientNetB7, UNet++, and CBAM for enhanced segmentation. Its superiority is evidenced by a 10% improvement in performance metrics over state-of-the-art methods, achieving Dice (0.9468), IoU (0.9001), Recall (0.9380) , and Precision (0.9581), attributable to its optimized design and effective noise handling via advanced augmentation techniques.

Chapter 6

Conclusion

This study presents a significant advancement in automated skin cancer diagnosis through the development of two novel hybrid segmentation models that integrate EfficientNetB7, UNet++ and Convolutional Block Attention Modules (CBAM). The comprehensive evaluation across five diverse datasets demonstrates the superior performance and robustness of the proposed architecture.

The experimental results validate the effectiveness of our approach across multiple benchmark datasets. On the ISIC2016 test dataset, our proposed model with CBAM achieved an accuracy of 96.14%, Dice coefficient of 0.9218 and IoU of 0.8566, representing substantial improvements over established architectures, including UNet (94.45% accuracy, 0.9026 Dice) and UNet++ (95.23% accuracy, 0.9159 Dice). The model demonstrated consistent superior performance across all evaluated datasets, with particularly notable achievements on HAM10000 (97.02% accuracy, 0.9442 Dice, 0.8955 IoU) and strong generalization capabilities on the challenging PH^2 dataset (95.55% accuracy, 0.8672 Dice).

One of the study's key findings is the influence of the CBAM attention mechanism. The consistency of results across various imaging conditions and datasets indicates that the attention mechanism enhances precision and feature discrimination, even though both variants of the models, with and without CBAM, achieve competitive performance. A significant gap in clinical viability is filled when the noise-robust augmentation technique and CBAM integration guarantee consistent performance in real-world imaging scenarios.

Regardless of these successes, some restrictions still exist. Images with dense hair follicles that blur lesion boundaries and scenarios with noticeable imaging artifacts that produce inaccurate edge responses present challenges for the model. Because of these difficulties, non-biological characteristics are occasionally identified as wrong lesion boundaries, leading to false positives and pointing out areas that need more work.

Beyond performance metrics, this work has clinical significance. From resource-constrained primary care facilities to specialized dermatology centers, the lightweight architecture design guarantees deployment viability across a range of healthcare environments. The model exhibits the dependability required for

clinical decision support systems, with Dice coefficients consistently above 0.84 and performance metrics consistently surpassing 95% accuracy across all tested datasets.

Advanced preprocessing techniques, such as complex hair removal algorithms and improved edge detection systems that can differentiate between biological and artificial features, should be given top priority in future research directions. Adding supplementary imaging modalities like polarized dermoscopy and extending training datasets to include multi-lesion scenarios could improve diagnostic accuracy and increase clinical applicability. Moreover, multimodal models can come in handy in this regard as well.

Finally, this study establishes a strong framework for AI driven skin cancer detection that effectively addresses the complexities of dermatoscopic image analysis. The architecture outperforms traditional models while maintaining computational efficiency appropriate for medical deployment. The thorough evaluation across multiple datasets, which resulted in consistently superior performance metrics including 1-2% accuracy improvements and Dice coefficient enhancements of 0.02-0.04 over existing methods, provides strong evidence for the approach's practicality. While there is still room for improvement, especially when dealing with complicated imaging conditions, this study contributes significantly to improving the consistency, efficiency and accessibility of skin cancer diagnosis in clinical practice.

Bibliography

[1] T. S. C. Foundation, *We don't want to scare you, but ...* en-US, May 2024. [Online]. Available: https://www.skincancer.org/blog/we-dont-want-to-scare-you-but/.

[2] C. C. NSW, *About skin cancer — cancer council nsw*, en-US, Jun. 2024. [Online]. Available: https://www.cancercouncil.com.au/skin-cancer/about-skin-cancer/.

[3] *Skin cancer*, (n.d.) [Online]. Available: https://www.aad.org/media/stats-skin-cancer.

[4] T. S. C. Foundation, *Skin cancer facts statistics - the skin cancer foundation*, en-US, Feb. 2024. [Online]. Available: https://www.skincancer.org/skin-cancer-information/skin-cancer-facts/.

[5] *Skin cancer*, Sep. 2024. [Online]. Available: https://my.clevelandclinic.org/health/diseases/15818-skin-cancer.

[6] *Skin cancer - symptoms and causes*, (n.d.). Mayo Clinic. [Online]. Available: https://www.mayoclinic.org/diseases-conditions/skin-cancer/symptoms-causes/syc-20377605.

[7] A. Esteva, B. Kuprel, R. A. Novoa, *et al.*, "Dermatologist-level classification of skin cancer with deep neural networks," *nature*, vol. 542, no. 7639, pp. 115–118, 2017. DOI: 10.1038/nature21056.

[8] P. Tschandl, C. Rinner, Z. Apalla, *et al.*, "Human–computer collaboration for skin cancer recognition," *Nature medicine*, vol. 26, no. 8, pp. 1229–1234, 2020. DOI: 10.1038/s41591-020-0942-0.

[9] H. A. Haenssle, C. Fink, R. Schneiderbauer, *et al.*, "Man against machine: Diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists," *Annals of oncology*, vol. 29, no. 8, pp. 1836–1842, 2018. DOI: 10.1093/annonc/mdy166.

[10] A. Bindhu and K. Thanammal, "Segmentation of skin cancer using fuzzy u-network via deep learning," *Measurement: Sensors*, vol. 26, p. 100 677, 2023. DOI: 10.1016/j.measen.2023.100677.

[11] H. Ashraf, A. Waris, M. F. Ghafoor, S. O. Gilani, and I. K. Niazi, "Melanoma segmentation using deep learning with test-time augmentations and conditional random fields," *Scientific Reports*, vol. 12, no. 1, p. 3948, 2022. DOI: 10.1038/s41598-022-07885-y.

[12] E. F. Duque-Vazquez, R. E. Sanchez-Yanez, N. Saldaña-Robles, M. F. León-Galván, and J. Cepeda-Negrete, "Hela cell segmentation using digital image processing," *Heliyon*, vol. 10, no. 5, 2024. DOI: 10.1016/j.heliyon.2024.e26520.

[13] L. Cai, K. Hou, and S. Zhou, "Intelligent skin lesion segmentation using deformable attention transformer u-net with bidirectional attention mechanism in skin cancer images," *Skin Research and Technology*, vol. 30, no. 8, e13783, 2024. DOI: 10.1111/srt.13783.

[14] U. Khan, U. Nawaz, M. Khan, W. Gueaieb, and A. El Saddik, "Deepskinformer: Skin lesion segmentation using hierarchical transformers and edge enhancement," in *2024 IEEE International Conference on Image Processing (ICIP)*, IEEE, 2024, pp. 3868–3874. DOI: 10.1109/ICIP51287.2024.10647899.

[15] M. Hu, Y. Li, and X. Yang, "Skinsam: Empowering skin cancer segmentation with segment anything model," *arXiv preprint arXiv:2304.13973*, 2023. DOI: 10.48550/arXiv.2304.13973.

[16] M. Imran, M. Islam Tiwana, M. M. Mohsan, N. S. Alghamdi, and M. U. Akram, "Transformer-based framework for multi-class segmentation of skin cancer from histopathology images," *Frontiers in Medicine*, vol. 11, p. 1 380 405, 2024. DOI: 10.3389/fmed.2024.1380405.

[17] A. N. Toprak and I. Aruk, "A hybrid convolutional neural network model for the classification of multi-class skin cancer," *International Journal of Imaging Systems and Technology*, vol. 34, no. 5, e23180, 2024. DOI: 10.1002/ima.23180.

[18] E. Pintelas and I. E. Livieris, "Xsc—an explainable image segmentation and classification framework: A case study on skin cancer," *Electronics*, vol. 12, no. 17, p. 3551, 2023. DOI: 10.3390/electronics12173551.

[19] M. J. Alam, M. S. Mohammad, M. A. F. Hossain, *et al.*, "S2c-delenet: A parameter transfer based segmentation-classification integration for detecting skin cancer lesions from dermoscopic images," *Computers in Biology and Medicine*, vol. 150, p. 106 148, 2022. DOI: 10.1016/j.compbiomed.2022.106148.

[20] X. He, Y. Wang, S. Zhao, and X. Chen, "Joint segmentation and classification of skin lesions via a multi-task learning convolutional neural network," *Expert Systems with Applications*, vol. 230, p. 120 174, 2023. DOI: 10.1016/j.eswa.2023.120174.

[21] M. Choudhary and N. Chauhan, "Enhancing skin lesion analysis: Leveraging unet and vgg architectures in deep learning models," in *2024 International Conference on Integrated Circuits, Communication, and Computing Systems (ICIC3S)*, IEEE, vol. 1, 2024, pp. 1–8. DOI: 10.1109/ICIC3S61846.2024.10603184.

[22] C. Kavitha, S. Priyanka, M. P. Kumar, and V. Kusuma, "Skin cancer detection and classification using deep learning techniques," *Procedia Computer Science*, vol. 235, pp. 2793–2802, 2024, International Conference on Machine Learning and Data Engineering (ICMLDE 2023), ISSN: 1877-0509. DOI: https://doi.org/10.1016/j.procs.2024.04.264.

[23] M. Badawi, R. Elgohary, M. Tarek, *et al.*, "Skin cancer classification and segmentation using deep learning," *International Journal of Telecommunications*, vol. 4, no. 01, pp. 1–23, 2024. DOI: 10.21608/ijt.2024.280957.1045.

[24] W. Lai, M. Kuang, X. Wang, P. Ghafariasl, M. H. Sabzalian, and S. Lee, "Skin cancer diagnosis (scd) using artificial neural network (ann) and improved gray wolf optimization (igwo)," *Scientific Reports*, vol. 13, no. 1, p. 19 377, 2023. DOI: 10.1038/s41598-023-45039-w.

[25] J. Mahin, X. Xu, L. Li, and C. Zhang, "Diagnosing skin cancer using social spider optimization (sso) and error correcting output codes (ecoc) with weighted hamming distance," *Scientific Reports*, vol. 14, no. 1, p. 22 997, 2024. DOI: 10.1038/s41598-024-73219-9.

[26] A. Bibi, M. A. Khan, M. Y. Javed, *et al.*, "Skin lesion segmentation and classification using conventional and deep learning based framework," *Comput. Mater. Contin*, vol. 71, no. 2, pp. 2477–2495, 2022. DOI: 10.32604/cmc.2022.018917.

[27] T. Akram, M. A. Khan, M. Sharif, and M. Yasmin, "Skin lesion segmentation and recognition using multichannel saliency estimation and m-svm on selected serially fused features," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–20, 2024. DOI: doi.org/10.1007/s12652-018-1051-5.

[28] M. Abedini, "Skin cancer classification by leveraging segment anything model for semantic segmentation of skin lesion," *IJARCCE - International Journal of Advanced Research in Computer and Communication Engineering*, DOI: 10.17148/IJARCCE.2024.13835.

[29] A. Mahbod, P. Tschandl, G. Langs, R. Ecker, and I. Ellinger, "The effects of skin lesion segmentation on the performance of dermatoscopic image classification," *Computer Methods and Programs in Biomedicine*, vol. 197, p. 105 725, 2020. DOI: 10.1016/j.cmpb.2020.105725.

[30] R. Sonia, J. Joseph, D. Kalaiyarasi, *et al.*, "Segmenting and classifying skin lesions using a fruit fly optimization algorithm with a machine learning framework," *Automatika*, vol. 65, no. 1, pp. 217–231, 2024. DOI: 10.1080/00051144.2023.2293515.

[31] G. M. Shahriar Himel, M. Islam, K. Abdullah Al-Aff, S. Ibne Karim, M. K. U. Sikder, *et al.*, "Skin cancer segmentation and classification using vision transformer for automatic analysis in dermatoscopy-based non-invasive digital system," *arXiv e-prints*, arXiv–2401, 2024. DOI: 10.48550/arXiv.2401.04746.

[32] Y. Wang, J. Su, Q. Xu, and Y. Zhong, "A collaborative learning model for skin lesion segmentation and classification," *Diagnostics*, vol. 13, no. 5, p. 912, 2023. DOI: 10.3390/diagnostics13050912.

[33] P. Tschandl, C. Rosendahl, and H. Kittler, "The HAM10000 dataset: A large collection of multi-source dermatoscopic images of common pigmented skin lesions," *CoRR*, vol. abs/1803.10417, 2018. arXiv: 1803.10417. [Online]. Available: http://arxiv.org/abs/1803.10417.

[34] D. Gutman, N. C. F. Codella, E. Celebi, *et al.*, *Skin lesion analysis toward melanoma detection: A challenge at the international symposium on biomedical imaging (isbi) 2016, hosted by the international skin imaging collaboration (isic)*, 2016. arXiv: 1605.01397 [cs.CV]. [Online]. Available: https://arxiv.org/abs/1605.01397.

[35] N. C. F. Codella, D. Gutman, M. E. Celebi, *et al.*, *Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (isbi), hosted by the international skin imaging collaboration (isic)*, 2018. arXiv: 1710.05006 [cs.CV]. [Online]. Available: https://arxiv.org/abs/1710.05006.

[36] T. Mendonça, P. M. Ferreira, J. S. Marques, A. R. S. Marcal, and J. Rozeira, "Ph2 - a dermoscopic image database for research and benchmarking," in *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2013, pp. 5437–5440. DOI: 10.1109/EMBC.2013.6610779.

[37] ClockWorkKid, *Github - clockworkkid/skin-cancer-segmentation-classification: Skin cancer detection, classification and segmentation based on the ham10000 and isic dataset archives.* en. [Online]. Available: https://github.com/ClockWorkKid/Skin-Cancer-Segmentation-Classification.