INTERNATIONAL JOURNAL OF Telecommunications
AIR DEFENSE COLLEGE

Skin Cancer Classification and Segmentation using Deep Learning

Citation: Elgohary, Rania; Tarek, Mostafa.; EzzAlRegal, Mohamed; Ahmed, Abdulrahman; Samir, Amr; Ehab, Nour; M.B. Badawi
Inter. Jour. of Telecommunications, ***IJT'2024****, Vol. 04, Issue 01, pp. 01-23, 2024.*

Editor-in-Chief: Youssef Fayed.

Received: 01/04/2024.

Accepted: 10/05/2024.
Published: 10/05/2024.

Publisher's Note: The International Journal of Telecommunications, IJT, stays neutral regard-ing jurisdictional claims in published maps and institutional affiliations.

Rania Elgohary [1], Mostafa Tarek[1], Mohamed EzzAlRegal[1], Abdulrahman Ahmed[1], Amr Samir[1], Nour Ehab[1], and M.B. Badawy [1,2, *]

[1] Department of Artificial Intelligence, College of Information Technology, Misr University for Science and Technology (MUST), 6th of October City 12566, Egypt.

[2] Mechanical Engineering Department, Faculty of Engineering, Alexandria University, El-Chatby, Alexandria 21544, Egypt.

* Correspondence author(s). E-mail(s): mohammed.badawi@must.edu.eg,

Contributing authors: Rania.elgohary@must.edu.eg, mostafat4438@gmail.com, mezz7409@gmail.com, abdulrahmn477@gmail.com, amrsalatif@gmail.com ,nourehaab21@gmail.com

Abstract: This paper integrates medical science and artificial intelligence, focusing on using convolutional neural networks (CNNs) to improve skin cancer diagnosis accuracy. Given the rising global incidence of skin cancers such as melanoma and basal cell carcinoma, this research is becoming increasingly important. This study uses the HAM10000 and PH2 datasets, which are known for their diverse skin cancer images, and employs a CNN-based approach informed by previous research findings.

The proposed methodology includes extensive preprocessing and augmentation to increase the dataset's variability, allowing for thorough training and evaluation. The CNN model, which was developed using advanced training methods and includes convolutional and pooling layers, is the result of previous research demonstrating the efficacy of CNNs in skin lesion detection. Furthermore, the U-NET-based segmentation model contributes to the comprehensive analysis by precisely delineating lesion boundaries, which improves the understanding of skin cancer. The CNN model's performance is evaluated using a variety of metrics, including accuracy, classification reports, confusion matrices, and segmentation-specific metrics like the Dice coefficient and IOU. These metrics provide valuable insights into the changing landscape of skin cancer diagnosis, allowing for the development of effective, precise, and accessible healthcare solutions in the dynamic field of dermatology. The experimental results for skin cancer classification are promising, indicating that the proposed approach outperforms other models. The best-trained classification model had an impressive 99.5% accuracy, 99.5% precision, and 99.5% recall. The test data was 97.204% accurate, 97.5% precise, and 97.2% recall. In addition, the U-NET model performed admirably in skin cancer lesion segmentation, with segmentation metrics such as an accuracy of 96.68%, precision of 95.39%, recall of 94.24%, Dice coefficient of 93.58%, and IOU of 97.09% for training data and an accuracy of 96.14%, precision of 93.44%, recall of 94.09%, Dice coefficient of 92.55%, and IOU of 96.43% for testing data.

Keywords: Deep learning; computer vision; skin cancer; multi-class classification; segmentation; PH2 dataset

Abbreviation	Definition
CNN	Convolutional Neural Network
VGG	Visual Geometry Group
Resnet	Residual Network
IOU	Intersection over Union
Lr	Learning Rate
SSD-KD	Single Shot MultiBox Detector with Knowledge Distillation
Xception	Extreme Inception
SVM	Support Vector Machine

1. Introduction

To improve diagnostic approaches for skin cancer, it is critical to understand the significant global concern caused by the disease's rising prevalence. Skin cancer, including types such as melanoma and basal cell carcinoma, represents a significant public health challenge, with its prevalence on the rise worldwide. Skin cancer, including types such as melanoma and basal cell carcinoma, represents a significant public health challenge, with its prevalence on the rise worldwide. Early detection and classification of skin lesions are critical for effective intervention and treatment. With sunlight exposure, genetic factors, and lifestyle choices all contributing to rising rates of skin cancer, there is an urgent need for novel solutions that can keep up with the growing demand for precise diagnostics.

This research is especially important considering this context, as it harnesses the capabilities of Convolutional Neural Networks (CNNs) and utilizes the HAM10000 and PH2 datasets to address the challenges of skin cancer diagnosis. By integrating medical expertise with innovative technology, the proposed study has the potential to revolutionize the field. Additionally, with the incorporation of the segmentation model, which precisely delineates lesion boundaries, this research offers a comprehensive approach to improving skin cancer diagnosis accuracy. In the broader context of dermatological healthcare, this study aims to make a significant contribution to the ongoing fight against skin cancer. By incorporating advanced deep learning techniques into the diagnostic process, we hope to provide a solution that not only addresses the current challenges posed by skin cancer but also anticipates and adapts to the changing landscape of this complex health issue.

Panda et al. [1] compared various deep learning models using a transfer learning approach, emphasizing the method's effectiveness in skin lesion classification. Similarly, Wang et al. [2] investigated deep learning-based melanoma segmentation and classification, with the latter developing the SSD-KD method, a self-supervised approach for lightweight classification. Sirotkin et al. [3] proposed an improved recognition system using a self-supervised curricular deep learning approach, while Aldhyani et al. [4] created a multi-class skin lesion classification system using a dynamic kernel deep-learning-based CNN. Maqsood and Damaševičius [5] proposed a framework for localizing and classifying multiple skin lesions, focusing on feature fusion and selection for smart healthcare applications. Similarly, Baig et al. [6] presented novel CNN-based diagnostic tools for multi-class skin lesions, emphasizing lightweight and machine-learning-based approaches.

Shetty et al. [7] and Ali et al. [8] emphasized the use of convolutional neural networks (CNNs) in skin lesion classification, leveraging machine learning techniques to improve accuracy. Zhuang et al. [9] and Hosna et al. [10] provided thorough overviews and introductions to transfer learning, a critical technique in this field. Additional studies by Mohammed and Kora [11], Nie et al. [12], and Popescu et al. [13] investigated the opportunities and challenges of ensemble deep learning, advances in dermoscopic image diagnosis, and neural network collective intelligence, respectively. Khan et al. [14] and Anand et al. [15] investigated the extraction and optimal selection of features for skin lesion classification via multi-model deep neural networks and enhanced transfer learning-based classification systems. Alam et al. [16], Aladhadh et al. [17], and Jain et al. [18] made additional contributions to the field by addressing issues with imbalanced datasets, the use of medical vision transformers, and transfer learning in skin cancer classification. Finally, Al-masni et al. [19] and Panthakkan et al. [20] proposed integrated deep convolutional networks and a novel hybrid approach that combines Xception and ResNet50 for accurate skin cancer prediction, respectively, building on the foundation laid by RD Seeja and A Suresh [21], who used deep learning for skin lesion segmentation and melanoma classification using SVM, and the diagnostic tool developed by A Tajerian et al. [22], who used machine learning for dermatoscopic skin cancer image differentiation, recent advances have significantly improved the field. Researchers have made significant progress by addressing challenges such as imbalanced datasets, integrating medical vision transformers, and leveraging transfer learning techniques.

This paper addresses the need to improve skin cancer diagnostic methods by combining medical science and artificial intelligence, with a focus on Convolutional Neural Networks (CNNs) for improved accuracy. Using the large HAM10000 dataset, which is known for its diverse skin cancer images, a CNN-based approach informed by previous research is employed. The dataset's variability is increased for training and evaluation through extensive preprocessing and augmentation.

2. Methodology

2.1. Data Collection

The datasets employed in this research are the "HAM10000" dataset and the "PH2" dataset. The "HAM10000" dataset comprises skin cancer images depicting various skin lesions. This dataset comprises a total of 10,015 images, each with dimensions (450, 600, 3). Each image is linked to a specific diagnosis, categorized into seven classes: Melanocytic nevi (nv), Melanoma (mel), Benign keratosis-like lesions (bkl), Basal cell carcinoma (bcc), Actinic keratoses (akiec), Vascular lesions (vasc), and Dermatofibroma (df). The dataset also provides additional information for each image, including diagnosis and age. The "PH2" dataset is another dataset used in this research, which consists of skin lesion images as well. It includes a total of 200 images, with dimensions of 768 × 560 pixels. These images were acquired in RGB color as BMP files.

2.2. Data Preprocessing

HAM10000: Following the retrieval of image files, all images were resized from 450x600 pixels to 28x28 pixels. Subsequently, the dataset was partitioned into training and testing sets, with 80% of the data allocated to the training set and the remaining 20% to the test set. Both the training and test sets were normalized to ensure consistency in the data distribution. Additionally, a label mapping was created, consisting of a dictionary that associates the names of the seven classes with key values ranging from 0 to 6, facilitating classification tasks.

PH2: All images were resized from 768x560 pixels to 224x224 pixels. Subsequently, the dataset was partitioned into training and testing sets, with 80% of the data allocated to the training set and the remaining 20% to the test set, ensuring a balanced distribution for training and evaluation. This resizing process enables compatibility with models that expect input images of uniform dimensions.

2.3. Data Augmentation

HAM10000: To address the class imbalance and augment the training dataset, various methods were employed. Skin images were augmented using transformations such as rotation, width shift, height shift, shear, horizontal flip, and vertical flip. This augmentation strategy increased the number of images from 10,015 to 45,756, while preserving identical dimensions of twenty-eight pixels in width, twenty-eight pixels in height, and three-color channels. By introducing variability into the training set, this augmentation approach enhances the model's generalization and robustness to different skin lesion variations.

PH2: Random rotation and horizontal flipping were applied to augment the PH2 dataset. These transformations introduce variations in the dataset, which helps in improving the model's ability to generalize to unseen data and enhances its robustness. This augmentation strategy diversifies the dataset while maintaining consistency in dimensions, facilitating more effective training of the model.

2.4. Model Architecture:

Classification using the HAM10000 dataset: The model consists of twelve layers. The model initiates with convolutional layers, proficient at capturing intricate patterns within images. The initial layer deploys sixteen filters, followed by a max-pooling layer strategically down-sampling spatial dimensions. This pattern iterates, progressively escalating complexity with 32, 64, and 128 filters in subsequent convolutional layers. The corresponding max-pooling layers strike a balance between preserving crucial features and reducing spatial dimensions, culminating in a final convolutional layer followed by a flattening operation. This transition readies the data for the fully connected layers, establishing a connection between spatial hierarchies and the dense layers. The subsequent dense layers, featuring 64 and 32 neurons, act as a potent feature extractor, refining the learned representations. The output is realized through a dense layer with seven neurons, each representing a distinct class in the present classification task. The SoftMax activation function ensures the model provides well-calibrated probabilities for each class, facilitating confident predictions. This comprehensive architecture carefully considers both spatial intricacies and hierarchical feature extraction, contributing to the model's robust performance in computer vision tasks. Figure 1 shows the used classification model architecture. Table 1 shows the hyperparameters for Classification Methodology.

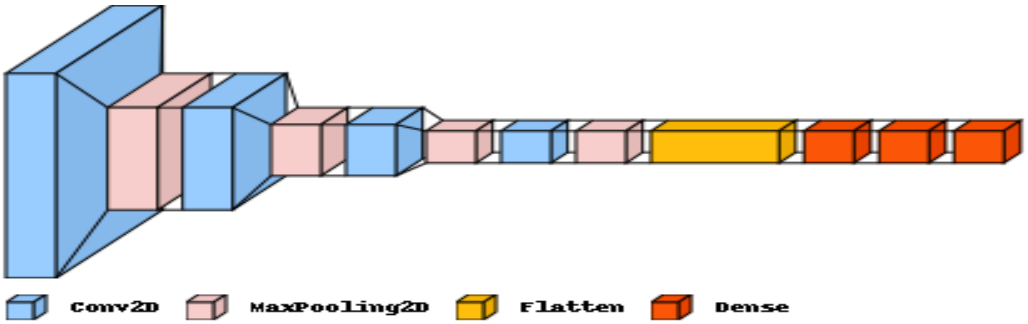

Figure 1: Proposed Classification Model Architectural Framework.

Table 1: Hyperparameters for classification process.

Hyperparameter	Value	Description
Learning Rate	0.001	The rate at which the model adjusts its weights during training.
Rotation Range	10	Range (in degrees) for random rotations applied to the images.
Width Shift Range	0.2	Range for random horizontal shifts applied to the images.
Height Shift Range	0.2	Range for random vertical shifts applied to the images.
Shear Range	0.2	Shear intensity (in radians) for geometric transformations.
Horizontal Flip	TRUE	Randomly flip images horizontally during training.
Vertical Flip	TRUE	Randomly flip images vertically during training.
Batch Size	64	Number of samples processed per gradient update during training.
Epochs	20	Number of complete passes through the entire training dataset.

Segmentation using the PH2 dataset: The model architecture includes one encoder and one decoder pathway. The encoder pathway initiates with four convolutional layers, followed by max-pooling layers for down-sampling, progressively increasing the complexity with deeper layers. Each convolutional block is composed of two convolutional layers with batch normalization and ReLU activation, ensuring effective feature extraction while mitigating the risk of overfitting. Additionally, spatial dropout is incorporated to enhance the model's robustness by introducing randomness during training.

The decoder pathway mirrors the encoder in terms of the number of layers, with four transposed convolutional layers for up-sampling. These layers are used to sample the feature maps to the original image resolution. This symmetric architecture facilitates the precise localization of skin lesion boundaries. Furthermore, the final layer employs a 1x1 convolution followed by a sigmoid activation function. Figure 2 shows the used Segmentation model architecture.

The Jaccard distance loss function is used to train the model by calculating the difference between predicted and ground truth segmentation masks. The segmentation model is trained for 100 epochs with the Adam optimizer and a learning rate of 0.003. Throughout the training process, different evaluation measures such as Intersection over Union (IoU), Dice coefficient, precision, recall, and accuracy are tracked to assess the model's performance on both the training and validation sets. Table 2 shows the hyperparameters for Segmentation Methodology.

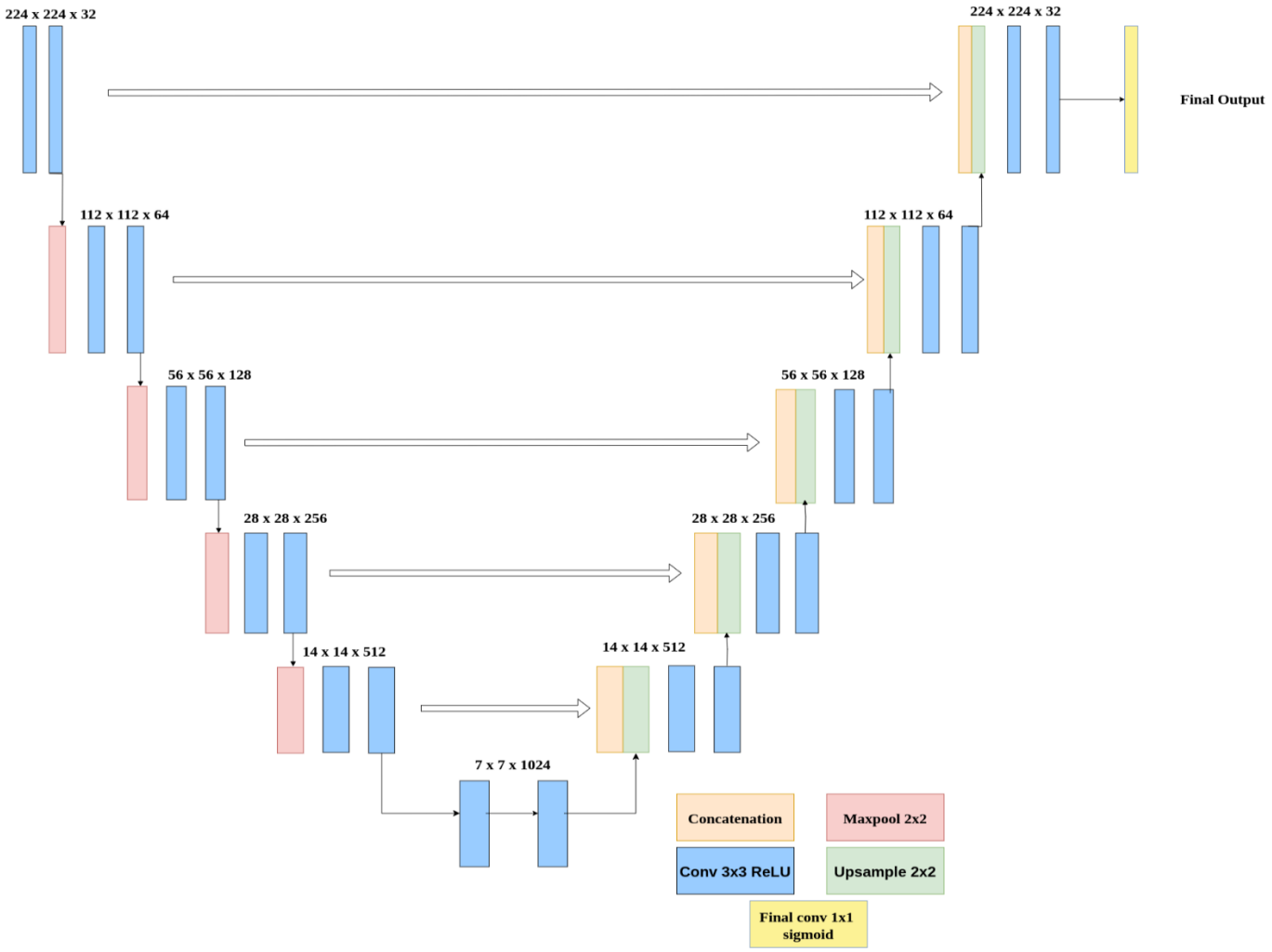

Figure 2: Segmentation Model Architectural Framework.

Table 2: Hyperparameters for Segmentation process.

Hyperparameter	Value	Description
Rotation Range	-40 to 40	Range (in degrees) for random rotations applied to the images.
Horizontal Flip	TRUE	Randomly flip images horizontally during training.
Dropout	0.4	Dropout rate for spatial dropout applied to convolutional layers.
Learning Rate	0.003	The rate at which the model adjusts its weights during training.
Optimizer	Adam	Optimizer algorithm used for training the model.
Batch Size	16	Number of samples processed per gradient update during training.
Epochs	Variable	Number of complete passes through the entire training dataset.

3. Dataset Insights

The HAM10000 dataset [23], also known as the Human Against Machine with ten thousand training images dataset, features high-quality images of skin lesions. It encompasses several types of skin lesions, ranging from benign to malignant. The lesions are categorized into seven distinct classes. However, there is an imbalance among the classes in the dataset, as illustrated in Table 3 and Figure 3.

Table 3: Class Distribution Analysis of the dataset.

Class	Counts
Melanocytic nevi (nv)	6705
Melanoma (mel)	1113
Benign keratosis-like lesions (bkl)	1099
Basal cell carcinoma (bcc)	514
Actinic keratoses and intraepithelial carcinoma (akiec)	327
Vascular lesions (vasc)	142
Dermatofibroma (df)	115

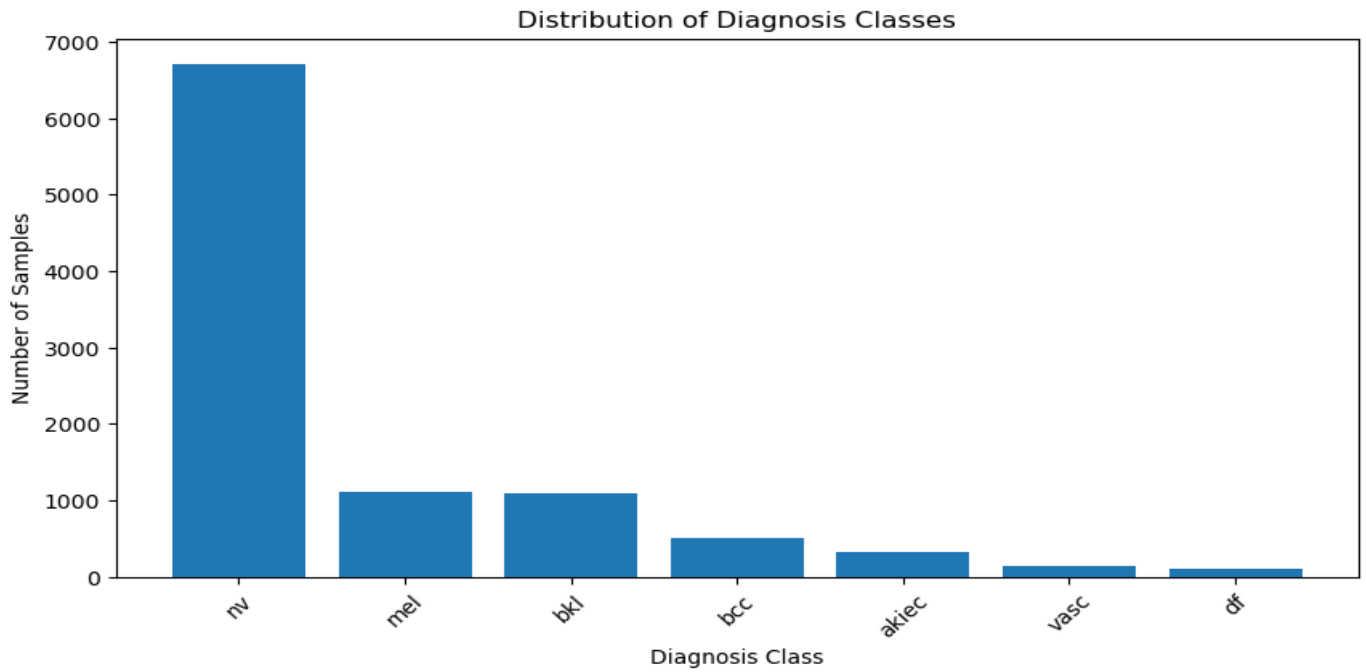

Figure 3: Class Distribution Analysis of the dataset.

To address the challenge of an unbalanced dataset within the HAM1000 dataset, a strategy of duplicating images was employed for augmentation purposes. This duplication process did not result in any new augmentation transformations, and to preserve the NV class's integrity in the HAM1000 dataset, no augmentation tech-

niques were used. During data preparation for training and testing, augmentation transformations such as rotation, shifting, and flipping were applied to the training set only, leaving the original dataset unchanged.

Table 4 and Figure 4 summarize the data augmentation strategies used to address the unbalanced dataset in HAM1000. The approach involved duplicating images for augmentation without introducing new transformations, while maintaining the integrity of the NV class. Augmentation techniques such as rotation, shifting, and flipping were exclusively applied to the training set.

Table 4: Class Distribution after using the factor.

Class	Counts	Factor used	Counts + (Counts*factor) + Counts
Melanoma (mel)	1113	4	6678
Benign keratosis-like lesions (bkl)	1099	4	6594
Basal cell carcinoma (bcc)	514	11	6682
Actinic keratoses and intraepithelial carcinoma (akiec)	327	17	6213
Vascular lesions (vasc)	142	45	6674
Dermatofibroma (df)	115	52	6210

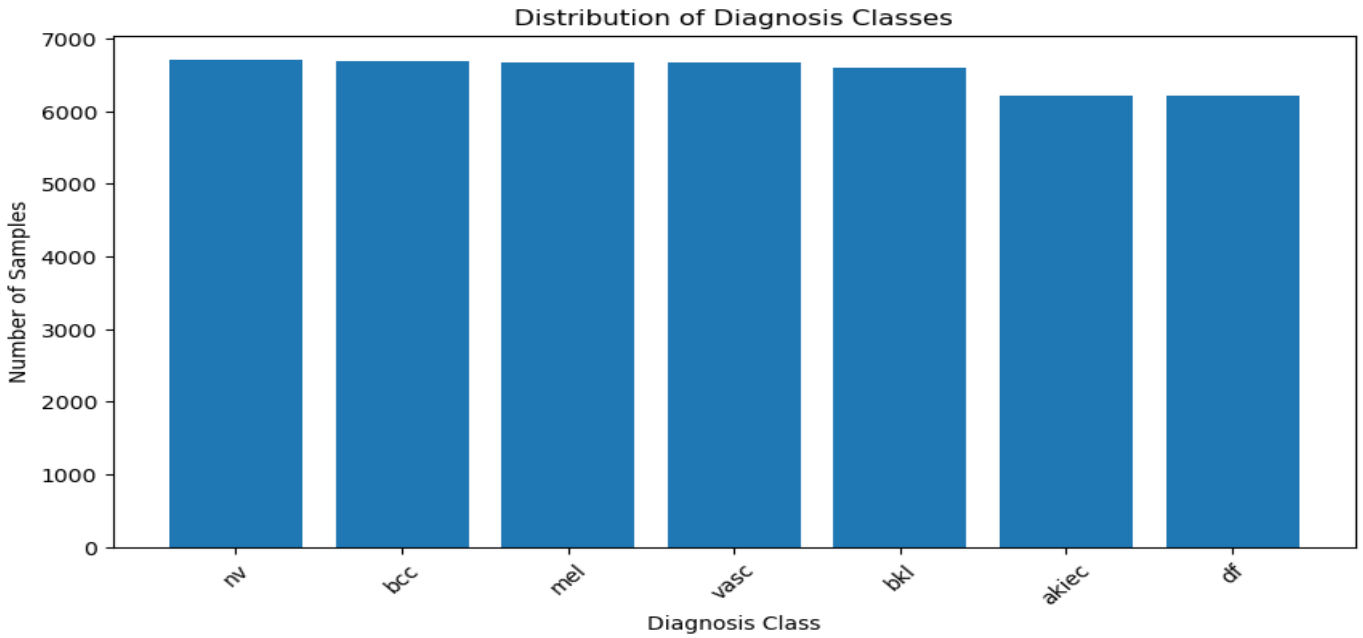

Figure 4: All Class Distribution after using the factor.

The PH2 dataset [20] is a well-known dataset in the field of dermatology and medical image analysis. It consists of 200 high-resolution images acquired in RGB color format as BMP files, with dimensions of 768 × 560 pixels.

4. Evaluation Metrics for both classification and segmentation

Before we delve into the results obtained from training various deep learning algorithms on the HAM10000 and PH2 datasets for predicting skin cancer, it is essential to understand the significance of each metric used in the evaluation process. Accuracy, precision, and recall for classification and Jaccard Distance, Intersection over Union (IoU), Dice Coefficient, precision, recall, and Accuracy for segmentation.

4.1 Accuracy [24]

Definition: Accuracy is a measure of the overall correctness of the model. It calculates the ratio of correctly predicted instances to the total instances.

$$Accuracy = \frac{True\ Positives + True\ Negatives}{Total\ Predictions} \quad (1)$$

Usefulness: While accuracy provides a general sense of how well the model is performing, it might not be the best metric for imbalanced datasets. In the case of skin cancer classification, where the occurrence of malignant cases might be significantly lower than benign cases, accuracy alone may not provide a complete picture.

4.2 Precision [24]

Definition: Precision measures the accuracy of positive predictions. It calculates the ratio of true positives to the total predicted positives.

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives} \quad (2)$$

Usefulness: Precision is crucial in scenarios where false positives are costly. In skin cancer classification, high precision means that when the model predicts a sample as malignant, it is likely to be correct. It is particularly important in medical contexts where misdiagnosing benign cases as malignant could lead to unnecessary treatments.

4.3 Recall [24]

Definition: Recall measures the ability of the model to capture all the relevant instances. It calculates the ratio of true positives to the total actual positives.

Equation:

$$Recall = \frac{TruePositives}{TruePositives + FalseNegatives} \quad (3)$$

Usefulness: Recall is vital when the cost of false negatives is high. In the context of skin cancer classification, high recall indicates that the model is effective in identifying malignant cases, minimizing the chances of missing potentially dangerous lesions.

4.4 Jaccard Distance [25]

Definition: The Jaccard distance, also known as the Intersection over Union (IoU), quantifies the dissimilarity between the predicted and ground truth segmentation masks. It measures the ratio of the intersection to the union of the two masks. A lower Jaccard distance indicates better segmentation accuracy.

$$JaccardIndex = \frac{|A \cap B|}{|A \cup B|} \quad (4)$$

$$JaccardDistance = 1 - JaccardIndex \quad (5)$$

Where:

- A and B are the ground truth and predicted segmentation masks, respectively.
- $|A \cap B|$ denotes the number of pixels common to both masks.
- $|A \cup B|$ represents the total number of pixels in both masks.

Usefulness: Jaccard distance is useful for evaluating the similarity between two segmentation masks. It provides a measure of how well the predicted segmentation aligns with the ground truth. A lower Jaccard distance indicates better segmentation accuracy.

4.5 Intersection over Union (IoU) [25]

Definition: IoU is a measure of the overlap between the predicted and ground truth segmentation masks. It calculates the ratio of the intersection to the union of the two masks, providing insights into the model's ability to accurately delineate skin lesion boundaries. Higher IoU values signify better segmentation performance.

$$IoU(A, B) = \frac{|A \cap B|}{|A \cup B|} \quad (6)$$

Where:

- A and B are the ground truth and predicted segmentation masks, respectively.
- $|A \cap B|$ denotes the number of pixels common to both masks.
- $|A \cup B|$ represents the total number of pixels in both masks.

Usefulness: IoU is commonly used in image segmentation tasks to assess the quality of segmentation results. Higher IoU values indicate better agreement between the predicted and ground truth segmentation masks, reflecting improved segmentation accuracy.

4.6 Dice Coefficient [25]

Definition: The Dice coefficient assesses the similarity between the predicted and ground truth segmentation masks. It computes the ratio of twice the intersection to the sum of the volumes of the two masks. A higher Dice coefficient indicates greater overlap and similarity between the predicted and ground truth masks.

Equation:

$$Dice(A, B) = \frac{2|A \cap B|}{|A| + |B|} \quad (7)$$

Where:

- A and B are the ground truth and predicted segmentation masks, respectively.
- $|A \cap B|$ denotes the number of pixels common to both masks.
- $|A|$and $|B|$denotes the total number of pixels in each mask.

Usefulness: The Dice coefficient is particularly useful in evaluating the performance of segmentation models. It provides a robust measure of segmentation accuracy, especially in scenarios with class imbalance, where accurately capturing small structures is essential.

5. Results and Discussion

5.1. Results for classification

As shown in Table 5 and Figure 5, DeepConvNet achieved the highest accuracy, precision, and recall scores, indicating its effectiveness in accurately classifying skin cancer lesions. Auto Encoder, while having a relatively high precision score, exhibited lower accuracy and recall scores compared to DeepConvNet, suggesting that it may have struggled with correctly identifying some instances of skin cancer. CNN decay lr, VGG16, ResNet50, InceptionV3, and Xception all demonstrated varying degrees of performance, with accuracy, precision, and recall scores falling below those of DeepConvNet but still showcasing some level of effectiveness in skin cancer classification.

Table 5: Performance Analysis: Metric Comparison across Training Algorithms.

Training Metrics			
Model Name	Accuracy	Precision	Recall
DeepConvNet	99.5	99.5	99.5
Auto Encoder	70.17	82.26	58.09
CNN with decay lr	81.23	89.72	73.28
VGG16	67.13	84.03	54.65
ResNet50	66.99	66.99	66.99
InceptionV3	66.97	85.86	54.2
Xception	66.8	85.92	54.31

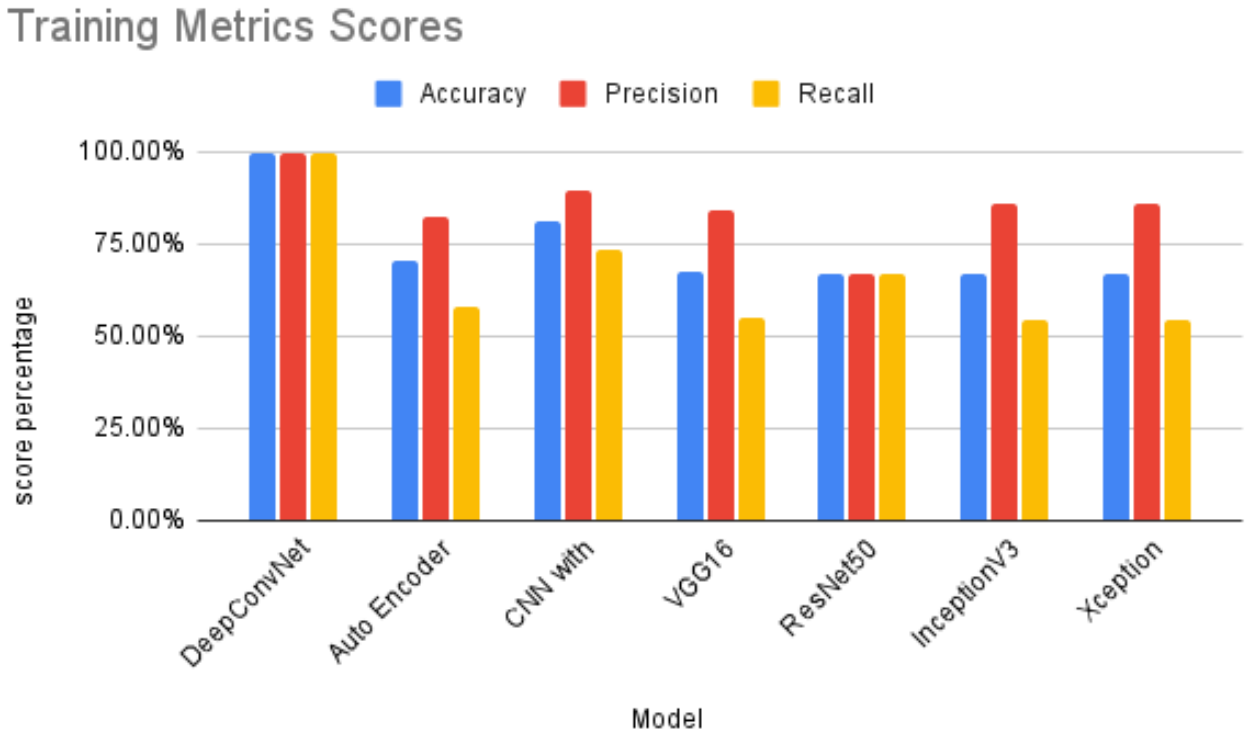

Figure 5: Performance Analysis: Metric Comparison across Training Algorithms.

Table 6 and Figure 6 show that the DeepConvNet outperformed all other testing algorithms in terms of accuracy, precision, and recall, demonstrating its ability to accurately categorize skin cancer lesions.

Table 6: Performance Analysis: Metric Comparison across Testing Algorithms.

Testing Metrics			
Model Name	Accuracy	Precision	Recall
DeepConvNet	97.204	97.5	97.2
Auto Encoder	70.17	82.49	58.17
CNN with decay lr	73.64	81.1	67.9
VGG16	66.99	83.52	60.03
ResNet50	66.83	66.83	66.83
InceptionV3	66.89	85.19	60.58
Xception	66.73	83.29	61.04

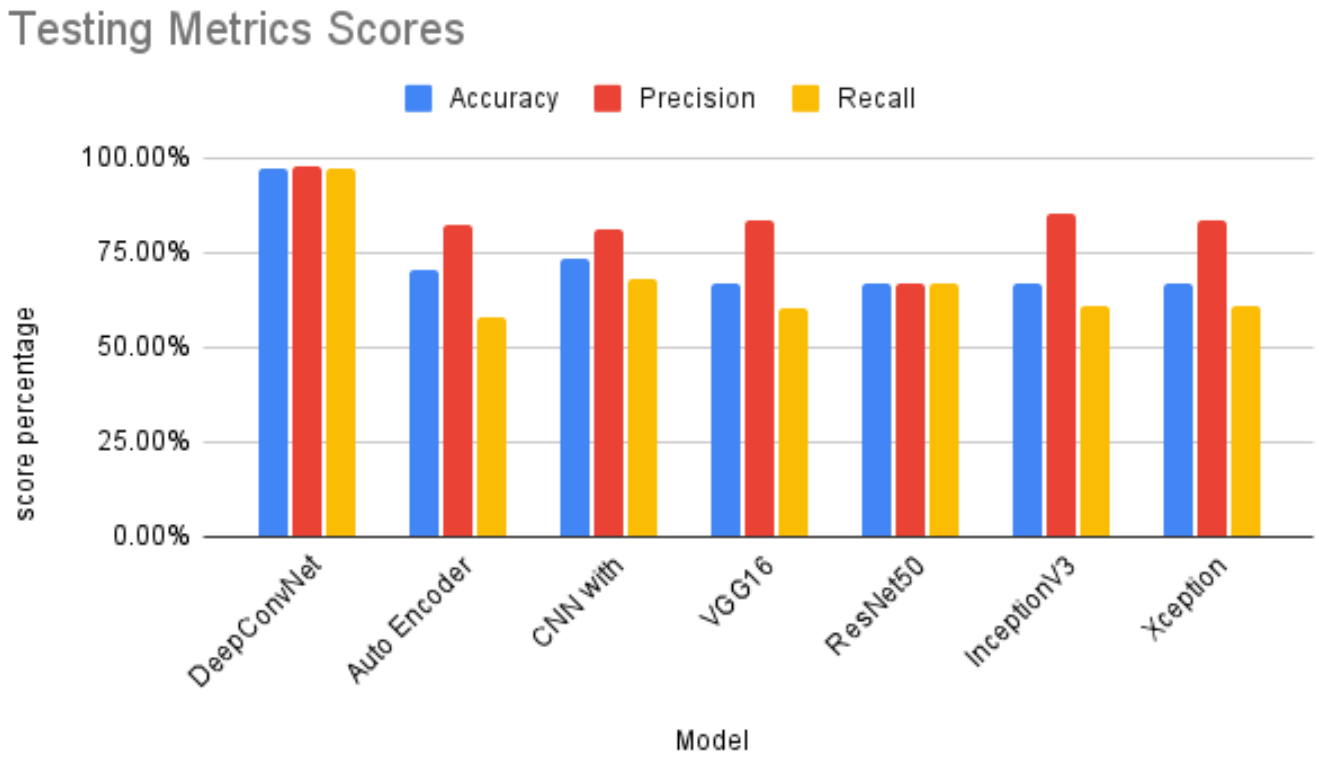

Figure 6: Performance Analysis: Metric Comparison across Testing Algorithms.

Figure 7 illustrates the progressive enhancement in accuracy over time, showcasing the positive trend in performance as training progresses.

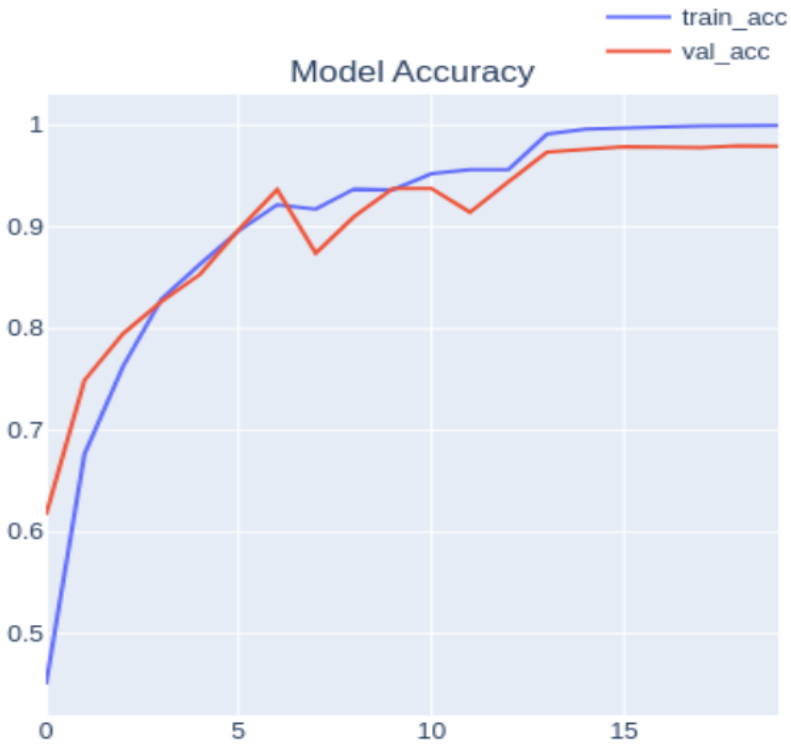

Figure 7: Tracking Progress: Evaluation of Accuracy for Proposed Model Architecture.

Figure 8 demonstrates the concurrent decline in loss over time, highlighting the iterative refinement and optimization of the model.

Figure 8: Tracking Progress: Evaluation of Loss for Proposed Model Architecture.

Figure 9 provides a visual representation of the model's performance through a confusion matrix graph for each class of data on the test set, including its performance on the unbalanced classes, as presented in Table and Figure 3. Each row in the confusion matrix corresponds to the actual class labels, while each column represents the predicted class labels. The values in the cells of the matrix indicate the number of instances that were classified into each class.

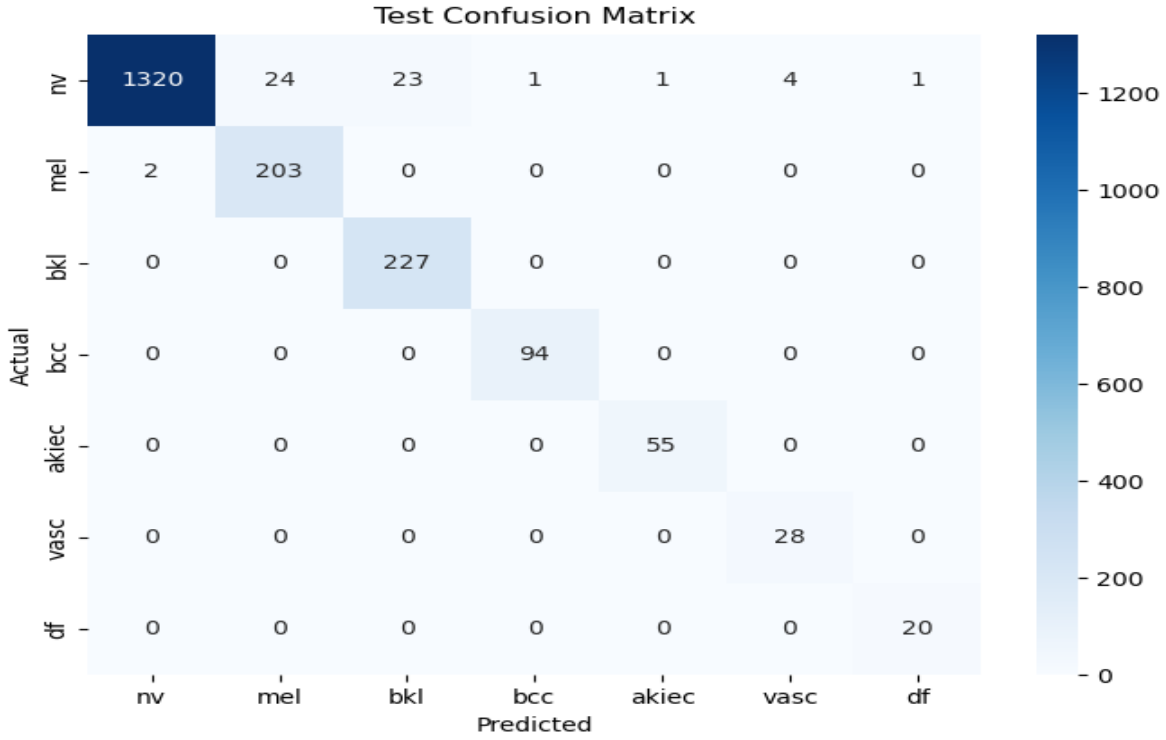

Figure 9: Visualizing Model Performance: Confusion Matrix Graph for Test Data.

This research paper provides a comprehensive overview of recent advances in skin lesion classification using deep learning models. The complexities of improving diagnostic accuracy and efficiency in skin disease detection are examined in detail using a variety of methodologies demonstrated by leading researchers, including transfer learning, knowledge distillation, and innovative network architectures.

1. S Panda et al. [1]: The research paper on skin lesion classification utilizing Deep Learning models employed various methods to achieve its objectives. The study utilized transfer learning with pre-trained models such as VGG16, ResNet50, InceptionV3, and Xception to classify skin lesions into different categories. The models were trained on a dataset consisting of images of various skin diseases, including melanoma, nevus, and seborrheic keratosis. The training process involved 30 epochs with a batch size of 16 for training and 10 for validation.

2. Y Wang et al. [2]: In a comparative study of skin lesion classification methods, deep learning models, traditional machine learning algorithms, and knowledge distillation techniques have been evaluated for their effectiveness in improving diagnostic accuracy. Deep learning models have shown promise in achieving higher accuracy rates due to their ability to learn complex patterns. Knowledge distillation techniques aim to enhance the performance of lightweight models by transferring knowledge from larger models. These methods have demonstrated improvements in accuracy, sensitivity, and specificity in skin lesion classification tasks. Lever-

aging advanced techniques like knowledge distillation can enhance the diagnostic accuracy of skin disease classification, contributing to more efficient diagnostic tools for skin diseases.

3. TH Aldhyani et al. [4]: The study focuses on the development and implementation of a lightweight dynamic kernel deep-learning-based convolutional neural network for multi-class skin lesion classification. The methodology employed in the research includes the use of variable size kernels and activation functions in the network, with a strategic allocation of fewer kernels in the initial layers for efficient utilization. Additionally, class-wise data balancing was performed to ensure unbiased training.

4. S Maqsood & R Damaševičius [5]: In this study on multiclass skin lesion localization and classification using deep learning, a novel approach was developed to enhance the accuracy and efficiency of skin cancer detection. The methodology involved the utilization of a customized Convolutional Neural Network (CNN) for automatic feature extraction, incorporating well-known networks such as Xception, ResNet-50, ResNet-101, and VGG16 to reduce computation time. The feature selection process was optimized using a unique Univariate Measurement of Pairwise Dependence (UMPD) approach, which effectively selected the best features for recognition.

5. B Shetty et al. [7]: In this research study on skin lesion classification, a variety of methods were employed to enhance the accuracy of the classification models. Machine learning models including Decision Tree, Random Forest, Support Vector Machine, K-Nearest Neighbor, Logistic Regression, Gaussian Naïve Bayes, and Linear Discriminant Analysis were evaluated, with Random Forest exhibiting the highest accuracy among them.

6. MS Ali et al. [8]: The research focuses on utilizing a deep convolutional neural network (DCNN) model combined with transfer learning techniques to enhance the classification of skin cancer based on dermoscopy images. The proposed DCNN model was developed to accurately classify skin lesions, particularly in the early stages of cancer. By training the model on a large dataset and fine-tuning it over multiple epochs, the researchers achieved significant improvements in classification accuracy compared to existing deep learning models. The results demonstrated that the DCNN model outperformed traditional transfer learning models such as AlexNet, ResNet, VGG-16, DenseNet, and MobileNet in terms of accuracy and execution time. Through a comprehensive evaluation on the HAM10000 dataset, the DCNN model showed superior performance in distinguishing between benign and malignant skin lesions, with promising implications for early detection and treatment of skin cancer.

7. V Anand et al. [15]: The research focuses on enhancing the classification of skin cancer through a transfer learning approach using the VGG16 architecture. The proposed model incorporates additional layers, including a flatten layer and dense layers with LeakyReLU and sigmoid activation functions, to improve accuracy. Data augmentation techniques are employed during pre-processing to increase dataset randomness and stability.

8. TM Alam et al. [16]: In their study of an Efficient Deep Learning-Based Skin Cancer Classifier for an Imbalanced Dataset, Alam et al. used a comprehensive methodology to address the challenges posed by imbalanced data.

9. S Aladhadh et al. [17]: In their research employed a two-tier framework to address the challenges associated with accurate skin cancer classification. The first stage involved data augmentation techniques to enhance the training dataset, mitigating issues related to insufficient labeled data. Subsequently, they developed a Medical Vision Transformer (MVT)-based classification model for skin cancer. This innovative approach involved splitting input images into patches and feeding them to the transformer in a sequence structure, akin to word embedding. The final classification was performed using a Multi-Layer Perceptron (MLP). The experimental results, conducted on the Human Against Machine (HAM10000) dataset, demonstrated the superiority of the proposed MVT-based model over existing state-of-the-art techniques.

10. RD Seeja & A Suresh [21]: The study focuses on utilizing deep learning technology for skin lesion segmentation and classification of melanoma. The methodology employed in this research involves the initial segmentation of dermoscopy images using a Convolutional Neural Network (CNN) based U-net algorithm. Subsequently, color, texture, and shape features are extracted from the segmented images using techniques such as Local Binary Pattern (LBP), Edge Histogram (EH), Histogram of Oriented Gradients (HOG), and Gabor method. These extracted features are then fed into various classifiers including Support Vector Machine (SVM), Random Forest (RF), K-Nearest Neighbor (KNN), and Naïve Bayes (NB) for the diagnosis of melanoma or benign lesions.

11. A Tajerian et al. [22]: In this study, we employed a methodological approach that leveraged dermoscopy images from the HAM10000 dataset to develop a machine-learning-based diagnostic tool for the classification of dermatoscopic skin cancer images. The process involved image pre-processing techniques such as labeling, resizing, and data augmentation to enhance the dataset. Transfer learning was utilized to create a model architecture based on EfficientNET-B1, incorporating a global average pooling 2D layer and a softmax layer with 7 nodes for classification.

Table 7 and Figure 10 present a comparison between the proposed method, implemented through the DeepConvNet architecture, and several previous research papers. Various training metrics, including accuracy, precision, and recall, are used for each model.

Table 7: Comparison between the proposed method and the other papers.

Training Metrics			
Paper	Accuracy	Precision	Recall
S Panda et al. [1]	-	97	95.2
Y Wang et al. [2]	84.6	-	-
TH Aldhyani et al. [4]	97.8	98.1	98
S Maqsood & R Damaševičius [5]	98.57	-	-
B Shetty et al. [7]	91.77	-	-
MS Ali et al. [8]	93.16	96.57	93.66
V Anand et al. [15]	89.09	-	-

TM Alam et al. [16]	91	-	-
S Aladhadh et al. [17]	96.14	96	96.50
RD Seeja & A Suresh [21]	85.19	42.59	50
A Tajerian et al. [22]	94	88	85
Current Proposed Method DeepConvNet	99.5	99.5	99.5

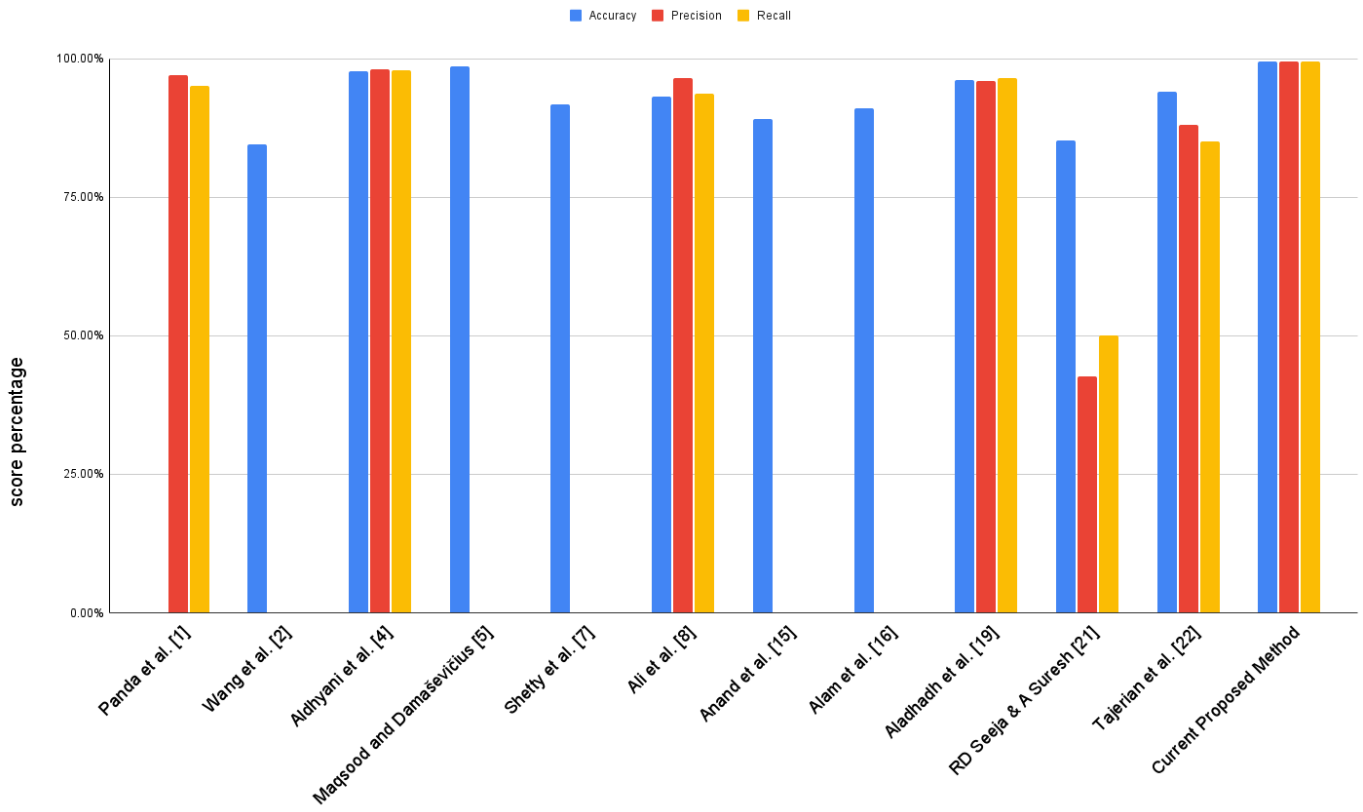

Figure 10: Benchmarking Proposed Method Against Existing Papers: A Comparative Study.

5.2. Results for Segmentation

In Table 8 and Fig 11, the U-Net model's performance metrics are examined across both training and testing datasets. The U-Net model achieves high accuracy, precision, recall, Dice Coefficient, and IoU scores, underscoring its efficacy in accurately segmenting skin cancer lesions.

Table 8: Performance Analysis: Metric Comparison across Training and Testing Sets.

Training Metrics					
Model Name	Accuracy	Precision	Recall	Dice Coefficient	IoU
U-NET	96.68	95.39	94.24	93.58	97.09
Testing Metrics					
Model Name	Accuracy	Precision	Recall	Dice Coefficient	IoU
U-NET	96.14	93.44	94.09	92.55	96.43

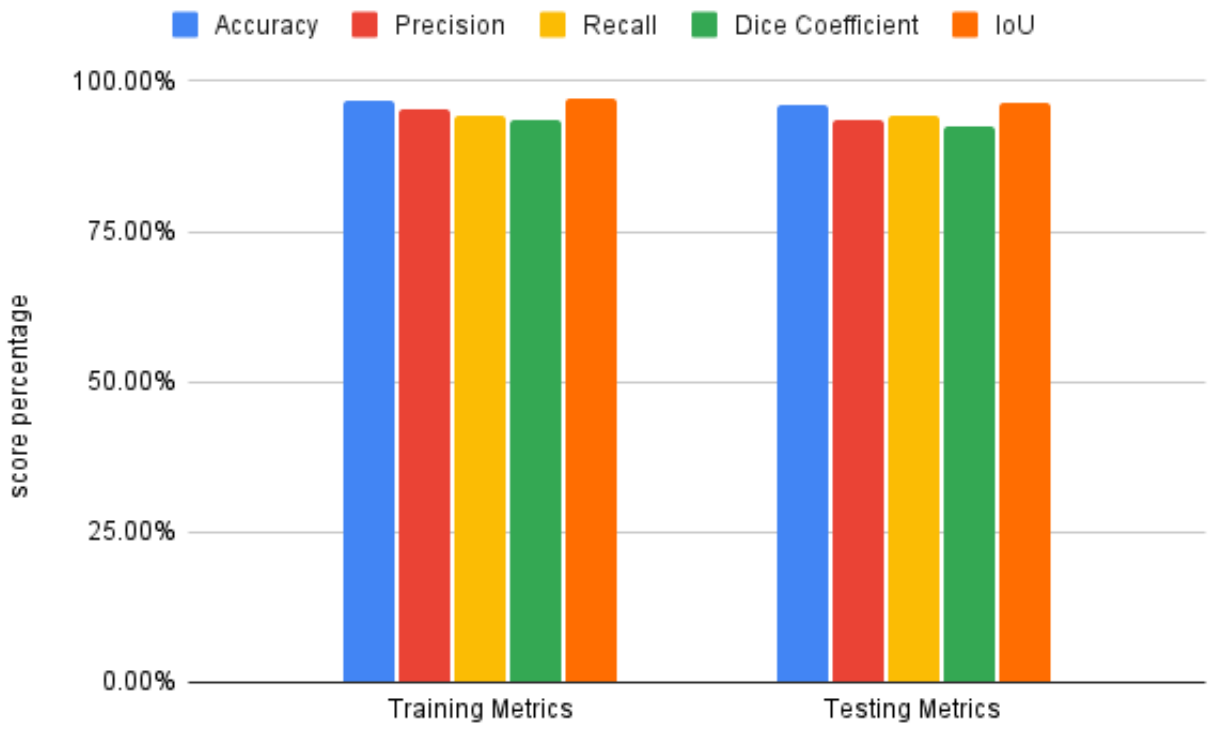

Figure 11: Performance Analysis: Metric Comparison across Training and Testing Sets.

Figures 12, 13, and 14 illustrate the progressive enhancement in accuracy, accompanied by a concurrent decline in Jaccard loss and improvement in Dice Coefficient, IoU, precision, and recall over time. These visuals represent the iterative refinement and optimization of the model, showcasing a positive trend in performance as training progresses.

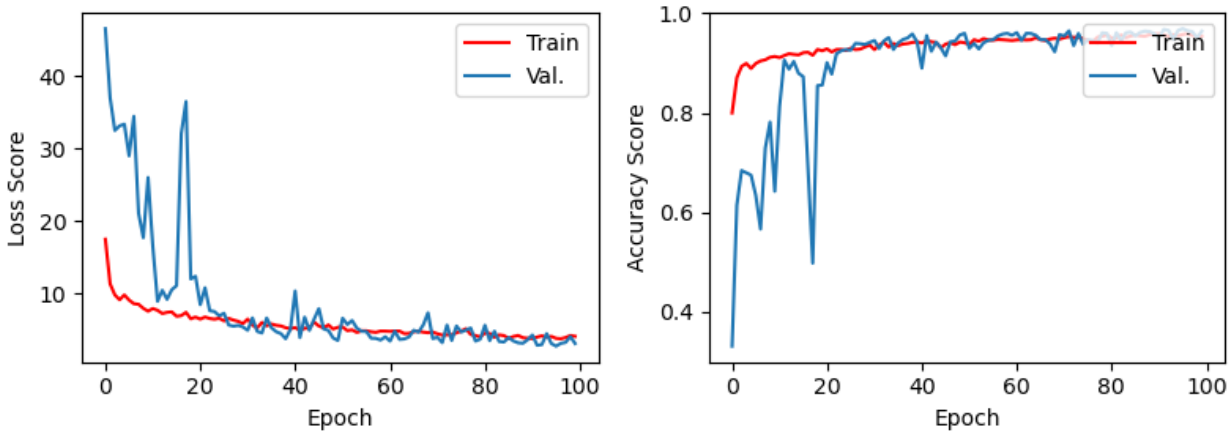

Figure12: Tracking Progress: Evaluation of Jaccard Loss and Accuracy for the Model Architecture.

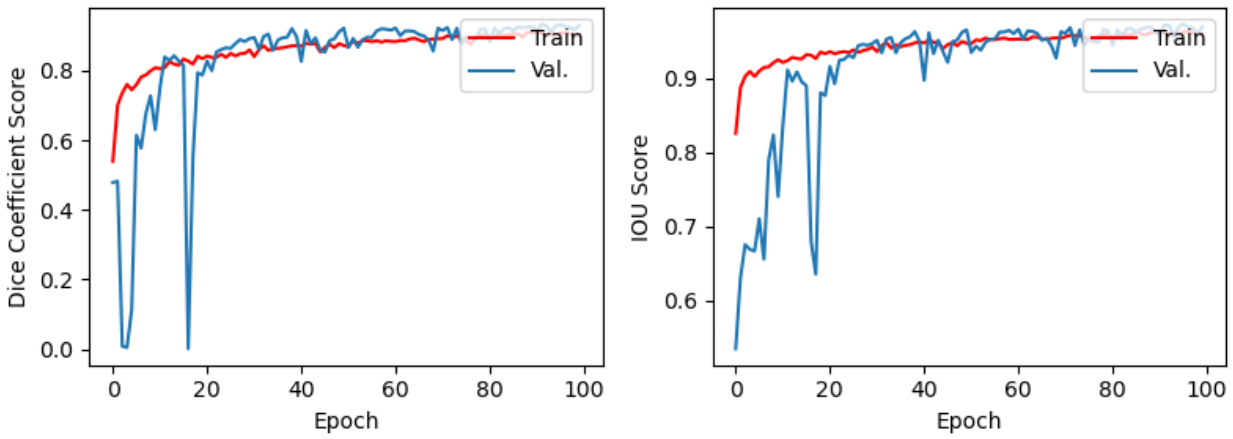

Figure 13: Tracking Progress: Evaluation of Dice Coefficient and IoU for Proposed Model Architecture.

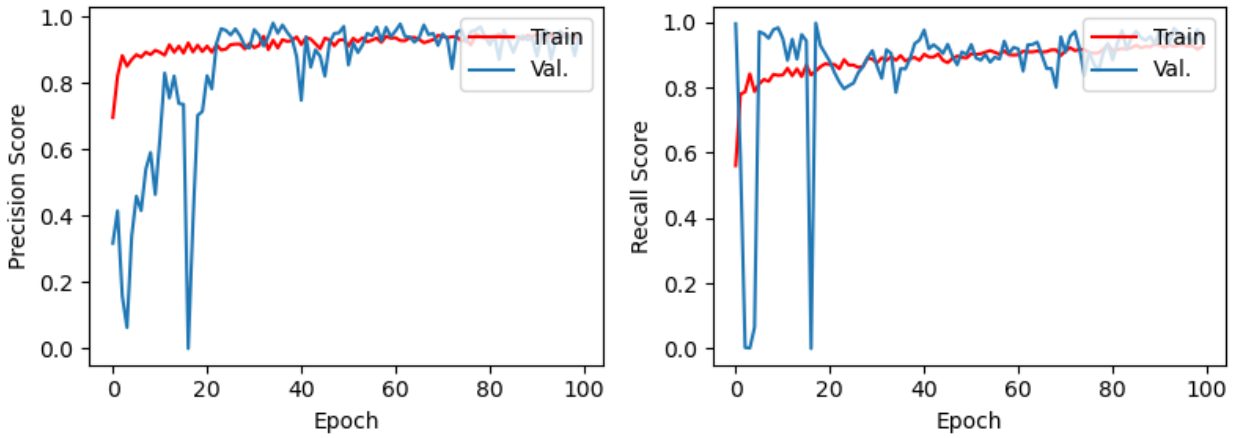

Figure 14: Tracking Progress: Evaluation of precision and recall for Proposed Model Architecture.

5.3 Proposed Model Prediction

In Fig 15, several predictions from the suggested model are showcased. This visual representation offers a model's performance by displaying examples of its predictions for skin cancer lesions. These predictions provide insights into how the model categorizes and classifies different types of lesions. The suggested model performs well across all seven classes, despite the imbalance in data distribution as illustrated in Figure 3. When the model predicts accurately, the confidence level typically falls within the 97%–100% range. Figures 16 and 17 show several predictions from the U-net segmentation model.

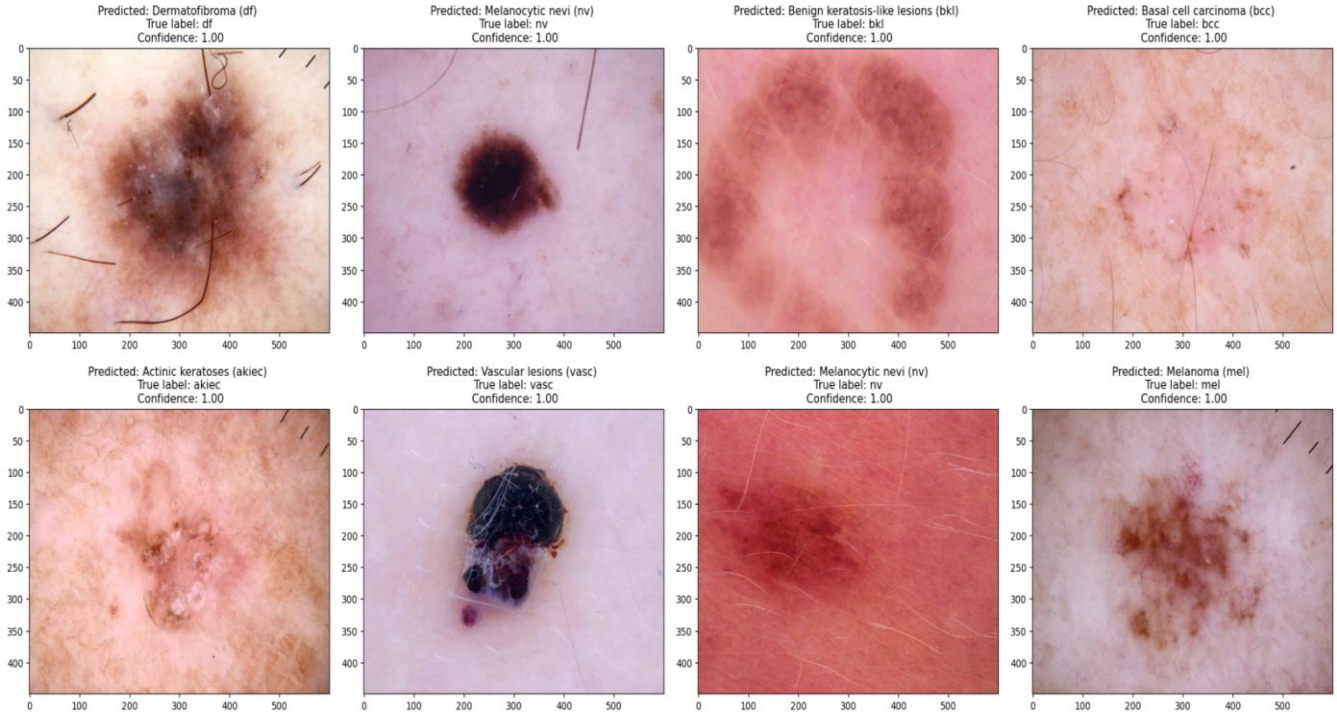

Figure 15: Sample classification predictions from the suggested model.

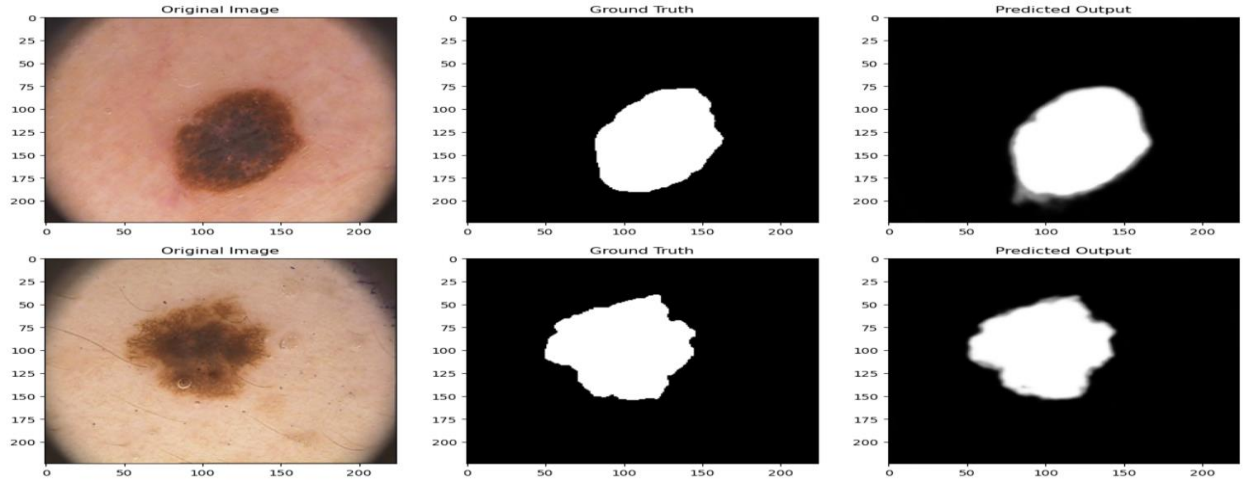

Figure 16: Sample predictions from the segmentation model.

Combining segmentation and classification models represents a promising approach to improving skin cancer diagnostic systems. In this integrated framework, the segmentation model precisely defines the lesion boundaries of skin cancer, effectively isolating the affected areas. These segmented regions are then localized or cropped and sent to the classification model for further analysis and diagnosis. Figure 17 illustrates this process by depicting the sequential workflow in which segmented lesions are accurately identified and then classified to determine the specific type of skin cancer. By combining these two methodologies, we can leverage the strengths of segmentation for precise delineation and classification for accurate diagnosis, ultimately improving the efficiency and reliability of skin cancer detection systems.

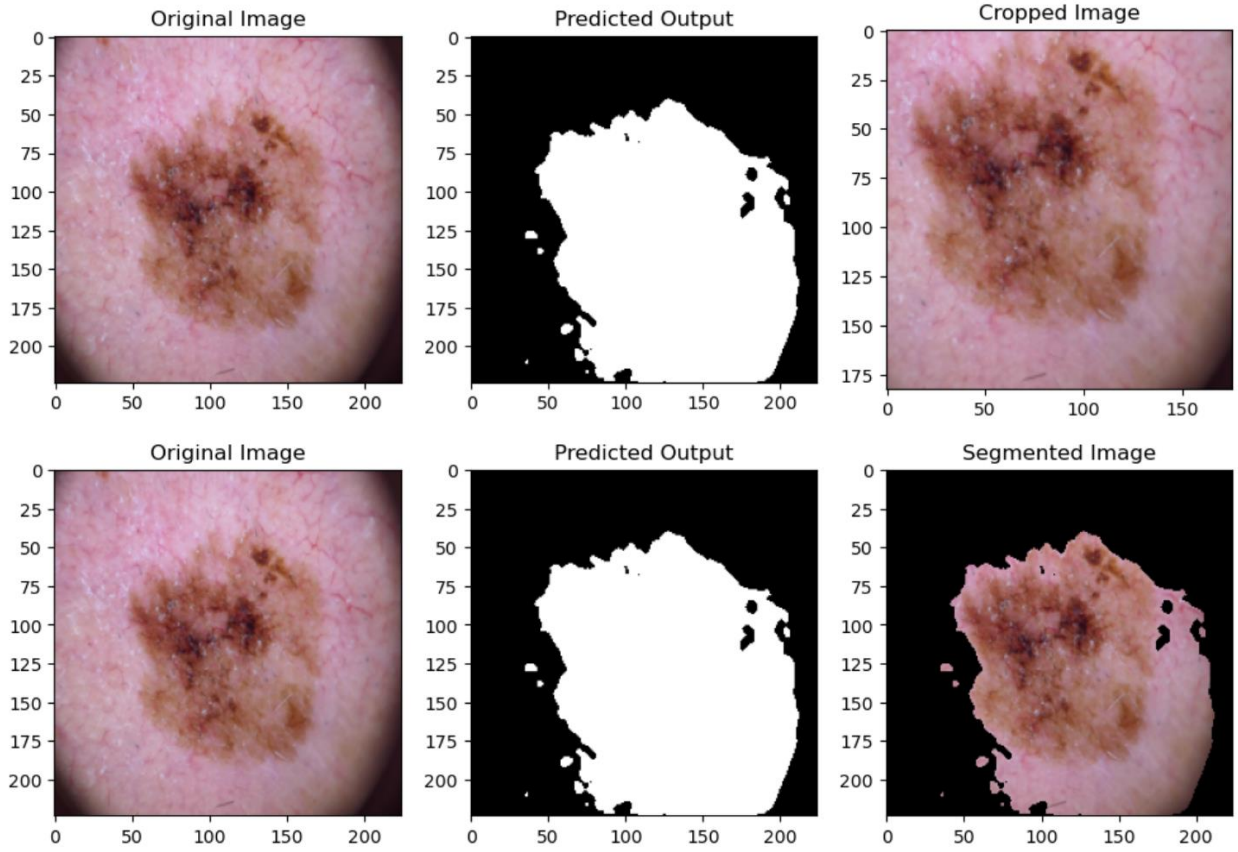

Figure 17: Segmentation-Driven Skin Cancer Diagnosis Model

7. Conclusion

The research highlights the significant advances made at the intersection of medical science and artificial intelligence, particularly in skin cancer diagnosis. The study found that integrating convolutional neural networks (CNNs) and leveraging the HAM10000 and PH2 datasets improved the accuracy and reliability of skin cancer classification and segmentation. The research proposed DeepConvNet architecture emerged as a front-runner, outperforming existing algorithms in accurately identifying several types of skin cancer lesions. With accuracy, precision, and recall scores of 99.5%, our model demonstrated its ability to accurately diagnose skin cancer lesions with unprecedented precision. The results obtained from the segmentation phase of our study underscore its pivotal role in advancing skin cancer diagnostics. Through segmentation, we achieved precise delineation of lesion boundaries, enabling accurate localization and isolation of affected areas. This level of precision not only enhances the efficiency of subsequent diagnostic processes but also facilitates targeted analysis by focusing exclusively on relevant regions of interest.

Finally, our research adds significantly to the ongoing efforts in dermatology healthcare by providing an innovative and accessible solution for skin cancer diagnosis. By leveraging the power of CNNs and advanced deep learning techniques, our proposed method paves the way for efficient, precise, and scalable solutions to reduce the global burden of skin cancer.

References

1. S Panda, AS Tiwari, MR Prusty. Comparative study on different Deep Learning models for Skin Lesion Classification using transfer learning approach. International Journal of Scientific and Research Publications. 2021;11(1):219-32.

2.Y Wang, Y Wang,J Cai, TK Lee, C Miao, ZJ Wang. Ssd-kd: A self-supervised diverse knowledge distillation method for lightweight skin lesion classification using dermoscopic images. Medical Image Analysis. 2023 Feb 1;84:102693.

3. K Sirotkin, M Escudero-Viñolo, P Carballeira, JC SanMiguel. Improved skin lesion recognition by a self-supervised curricular deep learning approach. arXiv preprint arXiv:2112.12086. 2021 Dec 22.

4. Aldhyani TH, Verma A, Al-Adhaileh MH, Koundal D. Multi-class skin lesion classification using a lightweight dynamic kernel deep-learning-based convolutional neural network. Diagnostics. 2022 Aug 24;12(9):2048.

5. S Maqsood, R Damaševičius. Multiclass skin lesion localization and classification using deep learning-based features fusion and selection framework for smart healthcare. Neural Networks. 2023 Mar 1;160:238-58.

6. AR Baig, Q Abbas, R Almakki, ME Ibrahim, L AlSuwaidan, AE Ahmed. Light-Dermo: A Lightweight Pretrained Convolution Neural Network for the Diagnosis of Multiclass Skin Lesions. Diagnostics. 2023 Jan 19;13(3):385.

7. B Shetty, R Fernandes, AP Rodrigues, R Chengoden, S Bhattacharya, K Lakshmanna. Skin lesion classification of dermoscopic images using machine learning and convolutional neural network. Scientific Reports. 2022 Oct 28;12(1):18134.

8. MS Ali, MS Miah, J Haque, MM Rahman, MK Islam. An enhanced technique of skin cancer classification using deep convolutional neural network with transfer learning models. Machine Learning with Applications. 2021 Sep 15;5:100036.

9. F Zhuang, Z Qi, K Duan, D Xi, Y Zhu, H Zhu, H Xiong, Q He. A comprehensive survey on transfer learning. Proceedings of the IEEE. 2020 Jul 7;109(1):43-76.

10. A Hosna, E Merry, J Gyalmo, Z Alom, Z Aung, MA Azim. Transfer learning: a friendly introduction. Journal of Big Data. 2022 Oct 22;9(1):102.

11. A Mohammed, R Kora. A comprehensive review on ensemble deep learning: Opportunities and challenges. Journal of King Saud University-Computer and Information Sciences. 2023 Feb 1.

12. Y Nie, P Sommella, M Carratu, M Ferro, M O'nils, J Lundgren. Recent Advances in Diagnosis of Skin Lesions Using Dermoscopic Images Based on Deep Learning. IEEE Access. 2022 Aug 17.

13. D Popescu, M El-Khatib, L Ichim. Skin lesion classification using collective intelligence of multiple neural networks. Sensors. 2022 June 10;22(12):4399.

14. MA Khan, MY Javed, M Sharif, T Saba, A Rehman. Multi-model deep neural network based features extraction and optimal selection approach for skin lesion classification. In2019 international conference on computer and Information Sciences (IC-CIS) 2019 Apr 3 (pp. 1-7) Aljouf, Kingdom of Saudi Arabia. IEEE.

15. V Anand, S Gupta, A Altameem, SR Nayak, RC Poonia, AK Saudagar. An enhanced transfer learning-based classification for diagnosis of skin cancer. Diagnostics. 2022 Jul 5;12(7):1628.

16. TM Alam, K Shaukat, WAKhan, IAHameed, LA Almuqren, MA Raza, M Aslam, S Luo. An efficient deep learning-based skin cancer classifier for an imbalanced dataset. Diagnostics. 2022 Aug 31;12(9):2115.

17. S Aladhadh, M Alsanea, M Aloraini, T Khan, S Habib, M Islam. An effective skin cancer classification mechanism via medical vision transformer. Sensors. 2022 May 25;22(11):4008.

18. S Jain, U Singhania, B Tripathy, EA Nasr, MK Aboudaif, AK Kamrani. Deep learning-based transfer learning for classification of skin cancer. Sensors. 2021 Dec 6;21(23):8142.

19. MA Al-Masni, DH Kim, TS Kim. Multiple skin lesions diagnostics via integrated deep convolutional networks for segmentation and classification. Computer methods and programs in biomedicine. 2020 Jul 1;190:105351.

20. A Panthakkan, SM Anzar, S Jamal, W Mansoor. Concatenated Xception-ResNet50—A novel hybrid approach for accurate skin cancer prediction. Computers in Biology and Medicine. 2022 Nov 1;150:106170.

21. RD Seeja, A Suresh. Deep learning-based skin lesion segmentation and classification of melanoma using support vector machine (SVM). Asian Pacific journal of cancer prevention: APJCP. 2019;20(5):1555.

22. A Tajerian, M Kazemian, M Tajerian, A Akhavan Malayeri. Design and validation of a new machine-learning-based diagnostic tool for the differentiation of dermatoscopic skin cancer images. PLoS One. 2023 Apr 14;18(4):e0284437.
23. P Tschandl, C Rosendah, H Kittler. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Scientific Data. 2018; 5: 180161. Search in. 2018;2.
24. SA Hicks, I Strümke, V Thambawita, M Hammou, MA Riegler, P Halvorsen, S Parasa. On evaluation metrics for medical applications of artificial intelligence. Scientific reports. 2022 Apr 8;12(1):5979.
25. SQ Nisa, AR Ismail. Comparative Performance Analysis of Deep Convolutional Neural Network for Gastrointestinal Polyp Image Segmentation.