Journal of Information Systems Engineering and Management
2025, 10(10s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

AI-Driven Skin Cancer Detection: A Deep Learning Perspective

Yashwant S. Ingle[1], Dr. Nuzhat Faiz Shaikh[2]
[1]*PhD Scholar, Dept. of Computer Engineering, SKNCOE SPPU, Pune, India*
yashwant.ingle218@gmail.com
[2]*PhD Research Supervisor, Dept of Computer Engineering, MES Wadia COE, SPPU, Pune, India*
nfshaikh76@gmail.com

ARTICLE INFO

Received: 04 Nov 2024

Revised: 22 Dec 2024

Accepted: 05 Jan 2025

ABSTRACT

The rising incidence of skin cancer necessitates the development of accurate and efficient diagnostic methods. This study introduces a novel approach for detecting skin cancer using an encoding-decoding technique with convolutional neural networks (CNNs) to enhance feature extraction from skin lesion images. Advanced CNN architectures, including DenseNet201, VGG16, and Xception, are utilized to classify skin lesions into seven categories. A thorough evaluation on a large dataset confirms the effectiveness of the proposed method in accurately identifying various types of skin cancer. Furthermore, a comparative study of multiple CNN models provides key insights into their relative strengths and limitations for diagnostic purposes. This research contributes to improving computer-aided skin cancer detection, paving the way for more reliable and accessible screening solutions.

Keywords: Skin Cancer Detection, Deep Learning, Convolutional Neural Networks (CNNs), Medical Image Analysis

Introduction

Skin cancer is a growing global public health concern, with its incidence steadily rising over the past decades. According to the World Health Organization (WHO), it accounts for nearly one-third of all cancer diagnoses, making it the most common form of cancer in humans. Various factors, including genetic susceptibility and prolonged exposure to ultraviolet (UV) radiation, contribute to its development. However, early detection remains crucial for effective treatment and improved patient outcomes. In this context, advancements in artificial intelligence (AI) and machine learning (ML) offer significant potential to enhance the accuracy and efficiency of skin cancer detection.

Identifying and categorizing skin lesions provide distinct challenges owing to the variety of lesion types and their nuanced visual attributes. Conventional skin cancer diagnostic techniques predominantly depend on dermatologists' visual assessments, which may be subjective and susceptible to inaccuracies, especially when lesions display confusing characteristics. Computer-aided diagnostic (CAD) systems utilizing machine learning algorithms have become essential instruments for dermatologists to detect and classify skin cancer. These systems evaluate digital images of skin lesions to identify pertinent aspects and categorize them as benign or malignant, assisting doctors in making educated diagnosis choices.

In recent years, deep learning techniques, especially convolutional neural networks (CNNs), have exhibited significant success in numerous picture identification tasks, including medical image analysis. Convolutional Neural Networks (CNNs) are well designed for skin cancer detection as they can autonomously learn distinguishing characteristics from unprocessed image data, eliminating the necessity for manual feature extraction. Extensively annotated datasets, such as the International Skin Imaging Collaboration (ISIC) dataset, facilitate the development and evaluation of CNN-based computer-aided diagnostic systems for skin cancer detection.

Notwithstanding advancements in automated skin cancer detection, some obstacles remain, such as categorizing skin lesions into various subtypes and elucidating model predictions in clinical environments [8]. Researchers have investigated innovative methods integrating feature extraction techniques with CNN architectures to enhance

classification performance. We propose a comprehensive approach for detecting skin cancer that integrates an encoding-decoding algorithm with the latest CNN architectures, including DenseNet201, VGG16, and Xception.

The encoding-decoding approach is a feature extraction tool, capturing hierarchical representations of skin lesion images that CNNs utilize for classification. This hybrid methodology utilizes the synergistic advantages of feature-based and data-driven techniques, thereby augmenting the discriminative capability of the CAD system. We broaden the classification job beyond determining if a lesion is benign or malignant to encompass seven distinct forms of skin cancer, facilitating a more precise diagnosis.

Literature Reviews

Dorj et al. utilized an online dataset of 3,753 images across four categories. [1] Utilizing AlexNet for feature extraction and an ECOC SVM for classification, they attained an astounding accuracy of 94.2 percent. They collected the internet dataset according to the standard benchmark protocols. Rezvantalab et al. [2] utilized a different methodology, including 120 images from eight categories within the HAM10000 dataset. The authors employed multiple pre-trained models to present their findings, including DenseNet 201, ResNet 152, InceptionV3, and InceptionResNetV2. The most precise model, DenseNet 201, attained an accuracy of 86.59%. We computed and documented AUC values for each model and class to convey the findings. Each class comprised several test specimens. The AUC values for the three courses in the PH2 dataset varied from 93.80% to 99.30%. Hosny et al. [3] achieved an impressive accuracy of 98.61 percent via a customized variant of AlexNet. They implemented this customized model to enhance the original dataset and obtained 4,400 photos. Dascalu and David's [4] analysis of the ISIC 2017 dataset, which categorized 5161 photos into two classes, attained an AUC of 81.40 percent. Their distinctive methodology employs sonification and K-means clustering to evaluate the impact of image quality on diagnostic precision.

In another study, Pham et al. [5] utilized the ISIC 2016 dataset (172 photographs) and the HAM10000 dataset (1113 images), both derived from a singular class. Their approach achieved an accuracy of 74.75 percent utilizing LBP balanced random forest, HSV, and linear normalization. The primary objective of this work was to compare the color, texture, and morphology of melanoma skin cancer cells. Hekler and his co-authors [6] achieved an accuracy of 82.95% by integrating physician assessments with predictions generated by a Convolutional Neural Network (CNN) utilizing the HAM10000 and ISIC datasets, including a total of 11,444 images categorized into five groups. This study yielded results for binary and multi-class classifications with the XGBoost algorithm.

Several groups have studied the HAM10000 dataset, with Emara et al. [7] using a modified InceptionV4 model yielding the best accuracy results. Their approach yielded a performance of approximately 94.7% accuracy. The primary purpose of the InceptionV4 modifications was to address the imbalanced class ratios present in that dataset. The study by Chaturvedi et al. could have been more successful, with a result of 83.1% for the same dataset. It used a pre-trained MobileNet architecture (and could use some transfer learning techniques because the original melanoma dataset was large, containing 38,569 photographs).

The study by Mohapatra et al. [8] utilized the seven classes within the HAM10000 dataset. They attained an accuracy of 80% using an unmodified, pre-trained MobileNet model. Conversely, Chen et al.'s 10 N/A dataset had nine skin lesions. Chen et al. attained an accuracy of 83.74% by employing a pre-trained ResNet50 model on the N/A dataset. Furthermore, they demonstrated the effective classification of nine separate categories of skin lesions with the ResNet50 model.

The National Cancer Center in Tokyo supplied Jinnai et al. [9] with a dataset including 5,846 pictures categorized into six distinct groups. The precision of their FRCNN, BCD, and TRN methodologies was 86.2 percent, 79.5 percent, and 75.1 percent, respectively. They utilized a custom dataset of the two primary categories, benign and malignant, to evaluate classifiers. Using ResNetXt101, Chaturvedi et al. [10] achieved an astonishing accuracy of 92.83% by analyzing the seven-class HAM10000 dataset. An in-depth analysis by Chaturvedi and colleagues revealed the best hyperparameter settings for identifying histopathology images, and their results showed that the ResNetXt101 model was the top-performing model for this task.

Mehwish Dildar et al. [11] present the approaches to assess the appearance of lesions, compare two characteristics (such as symmetry), or evaluate the contrast of color, size, or shape against the standard. Certain specialists predict with certainty the outcome of a boxing match between benign and malignant skin cancers. Nevertheless, these efforts

seldom result in effective software for early skin cancer detection; most programs rely on skilled individuals to analyze histological sections' septal (sandwich) imaging.

Nourelhoda M. Mahmoud et al.histological sections' septal (sandwich) imaging [12] develops and delineates an automated diagnosis system for early skin cancer. Artificial intelligence enables interaction with dermoscopic images of cutaneous lesions. The segmentation part of the system utilizes snake and region-expanding algorithms tailored to prevailing conditions. The findings indicate that adaptive snakes surpass expanding regions in accuracy and efficiency. Support vector machines and artificial neural networks are the primary techniques employed in the categorization phase, with the conclusion that artificial neural networks outperform support vector machines. The method using artificial neural networks (ANNs) achieves 94% accuracy, 96% precision, 95.83% specificity, 92.30% sensitivity, and an F1 score of 0.94. The device is intuitive, time-efficient, and effectively delivers a prompt determination of "skin cancer or not" for patients.

Umesh Kumar Lilhore et al. [13] substituted the model's excitation and squeezing components with the pragmatically beneficial channel attention component to decrease the parameter count. They proposed the implementation of cross-layer connections among mobile modules to optimize the utilization of synthetic features. They utilized dilated convolutions to enhance the receptive field. They concentrated on improving the model's performance by fine-tuning hyperparameters, an essential aspect of optimization efforts. They use sophisticated optimization methods, such as Bayesian optimization, to identify the optimal hyperparameters for the pre-trained MobileNet-V3. We evaluated our enhanced MobileNet-V3 compared to ResNet-152v2, VGG-19, MobileNet, VGG-16, and MobileNet-V2 (utilizing the HAM-10000 dataset for training and testing) for the detection and differentiation of melanoma. The metrics employed to evaluate the efficacy of these techniques in identifying and accurately diagnosing melanomas, compared to human pathologists' findings, are precision, sensitivity, accuracy, and specificity. This research indicates that the MobileNet-V3 model achieves 97.84% precision, 96.35% sensitivity, 98.86% accuracy, and 97.32% specificity with optimal hyperparameters. This research not only produced results but also proved beneficial. Patients who stood to benefit the most received enhanced medical care that was potentially lifesaving and economical.

Balaha, H. M. et al. [14] advocate for an autonomous, threshold-oriented methodology for segmenting, classifying, and detecting skin malignancies. This method utilizes a meta-heuristic optimizer called the Sparrow search algorithm (SpaSA). The procedure employs five distinct configurations of the U-Net model (U-Net, U-Net++, Attention U-Net, V-net, and Swin U-Net) for the segmentation phase. The pre-trained models utilized by the authors in this work comprise VGG16, VGG19, MobileNet, MobileNetV2, MobileNetV3Large, MobileNetV3Small, NASNetMobile, and NASNetLarge. The authors employed the meta-heuristic SpaSA to optimize the hyperparameters of eight CNN models. The dataset was sourced from five public entities. They generated two datasets from the segmented images: one with two classes and the other with ten classes. U-Net++, utilizing DenseNet201 as its backbone architecture, attained the highest results documented for the "skin cancer segmentation and classification" dataset. It employed an alternative variant of the cosine loss function, resulting in a loss of 0.104 on the test set. It performed exceptionally on various metrics, achieving 94.16% accuracy, 91.39% F1-score, 99.03% AUC, 96.08% and 96.41% IoU for the two classes specified in the dataset. The authors indicated that U-Net++ attained 77.19% and 75.47% on two weakly supervised training dataset instances. The Attention U-Net utilizing DenseNet201 achieved the highest performance in the "PH2" dataset. It experienced a loss of 0.137, with precision, accuracy, AUC, and other metrics varying from 92.74% to 68.04%, while the "squared hinge" and "hinge" loss configurations ranked lowest. Our convolutional neural network (CNN) experiments attained a maximum accuracy of 98.27% when applied to the "ISIC 2019 and 2020 Melanoma" dataset. Our optimal model was derived from a pre-trained MobileNet architecture. The pre-trained MobileNet model, with an alternative dataset, attained a second-place accuracy of 98.83% in our skin cancer classification models. The lowest accuracy recorded (85.87%) was achieved using a MobileNetV2 pre-trained model on an alternative skin cancer dataset. Our accuracy rates for each dataset were competitive when juxtaposed with the results of 13 analogous research.

Tabrizchi H. et al. [15] developed a novel technique for diagnosing early skin cancer. The processing of dermoscopic images constitutes the foundation. The model's design is founded on the VGG-16 network, a well-established convolutional neural network (CNN) framework. Instead of employing the conventional configuration of the VGG-16 network, we opted for an enhanced version of the architecture as the primary framework in our model. Most nearly skin cancer diagnosis cements are influenced by modifications to the model's image data processing pipeline. Our objective is to improve the accuracy of skin cancer diagnosis to a level deemed operationally viable in real-world

scenarios. The results indicate that our proposed model exhibits greater accuracy than the alternative methods evaluated.

Adla, D. et al. [16] present a strategy for enhancing skin cancer categorization and lesion segmentation. A dynamic graph cut algorithm was employed to achieve this. This methodology rectifies the prevalent issues of over-segmentation and under-segmentation in cut algorithms by precisely segmenting skin lesions, including minuscule ones. We additionally illustrate the efficacy of data augmentation. In a recent skin cancer competition, our model attained an exceptional performance metric of 97.986% across six categories, primarily due to a substantial reduction in false positives relative to the nearest competitor. The outcomes of two separate transfer models indicate that the accuracy of this model is primarily due to the faults it circumvented rather than its utilization of novel training images.

Skin Cancer Detection Using Deep Learning Algorithm

Fig. 1 depicts the system's block diagram of skin cancer detection using SegNet.

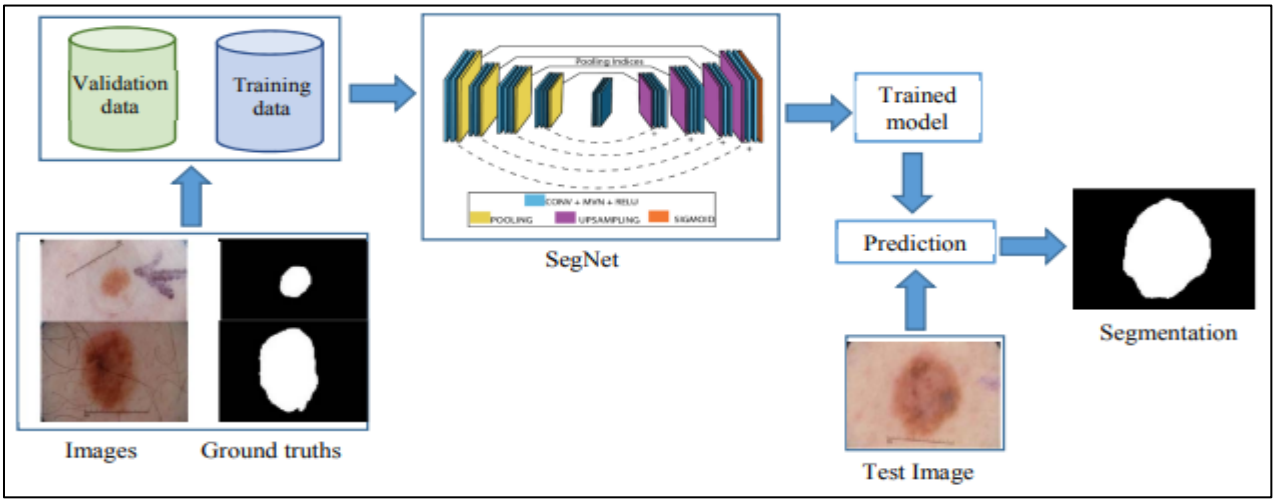


Fig. 1. Block diagram of the Skin lesion segmentation using SegNet

A. *Skin lesion Image Database*

The IPH2 dermoscopic collection comprises 200 images and their respective label masks. Each image possesses fixed dimensions of 572 by 765 pixels and is formatted in RGB. [14] The dataset is accessible to the public for research and experimental purposes. Each image was initially resized to 192 × 256 for training purposes before being input into the network. It diminishes the network's training parameters, time, and complexity without substantially affecting the results.

B. *SegNet*

SegNet is a convolutional neural network (CNN) architecture designed explicitly for pixel-wise picture segmentation tasks. It delineates medical image characteristics, such as cell nuclei in dermatological images from the PH2 collection.

SegNet is an advanced framework for deep semantic pixel-wise segmentation. The computer vision and robotics group at Cambridge University proposes it. SegNet comprises a pixel-wise classification layer, a corresponding decoder network, and an encoder network. Before the non-overlapping max pooling and subsampling layers in each encoder network, batch normalization, a ReLU activation, and one or more non-linear convolutional layers are used. Decoders are fundamentally akin to encoders; the primary distinction is in their linearity. The decoders utilize max-pooling for non-linear upsampling. These preserve high-frequency values and diminish the quantity of trainable parameters in segmented images. The softmax layer accepts the output from the final decoder and generates the ultimate result [9]. This is a summary of the essential elements and procedures:

- Keras implements the SegNet architecture. It consists of an encoding and a decoding stage.
- The encoding stage includes convolutional layers with batch normalization and activation functions, followed by max-pooling layers for down-sampling.

- Upsampling and transposed convolutional layers with batch normalization and activation functions are used in the decoding stage.
- Skip connections combine low-level and high-level features, allowing for precise object localization.
- The final layer outputs a binary segmentation mask.

We employ the binary cross-entropy as the loss function here. The cross-entropy function calculates the deviation of each class's prediction from the true value. We calculate the ultimate loss by averaging the classwise errors. Depending on the mask, this problem has two classes: black or white (0 or 1). Therefore, we employ binary cross-entropy as the loss function rather than the initially proposed categorical cross-entropy. The binary cross-entropy takes the following form:

$$L(y, \bar{y}) = -\frac{1}{N}\sum_{i=0}^{N}(y * \log(\bar{y}_1) + (1-y) * (1-\bar{y}_1)) \quad (1)$$

The network's SGD (Stochastic Gradient Descent) is the optimizer. One of the numerous values commonly used for the learning rate parameter, the learning rate, is set to 0.001, making it a crucial hyperparameter in the optimization.

Momentum also provides an updated rule inspired by physical optimization. The benefits of utilizing momentum with SGD include significantly accelerating the learning process from tiny changes. Similarly, the update process saves and uses the velocities of all parameters. For optimization, a momentum value of 0.9 is used.

C. Performance Evaluation

Metrics such as Intersection over Union (IoU), Dice Coefficient (DI), Precision, Recall (Sensitivity), and Accuracy were used to assess the performance of the proposed system.

- Intersection over Union (IoU): This metric, which gauges the overlap between expected and ground truth segmentation, is also called the Jaccard Index. The computation involves dividing the union of the anticipated and ground truth regions by the intersection of these regions.

$$IoU = \frac{Intersection\ of\ predicted\ and\ GT}{Union\ of\ predicted\ and\ GT} \quad (2)$$

  IoU values range from 0 to 1, with higher values indicating better overlap.

- Dice Coefficient (DI): The similarity between the predicted and ground truth segmentations is also measured by the Dice Coefficient. It is derived by dividing the total of the areas of the anticipated and ground truth regions' intersections twice.

$$DI = \frac{2*Intersection}{\text{Area of Predicted + Area of Ground Truth}} \quad (3)$$

  Similar to IoU, the Dice Coefficient has a range of values from 0 to 1, where larger values correspond to more accurate segmentation.

II. SKIN CANCER RECOGNITION USING DEEP LEARNING ALGORITHM

This approach was created to categorize various forms of skin cancer. Fig. 2 displays the proposed system's block diagram.

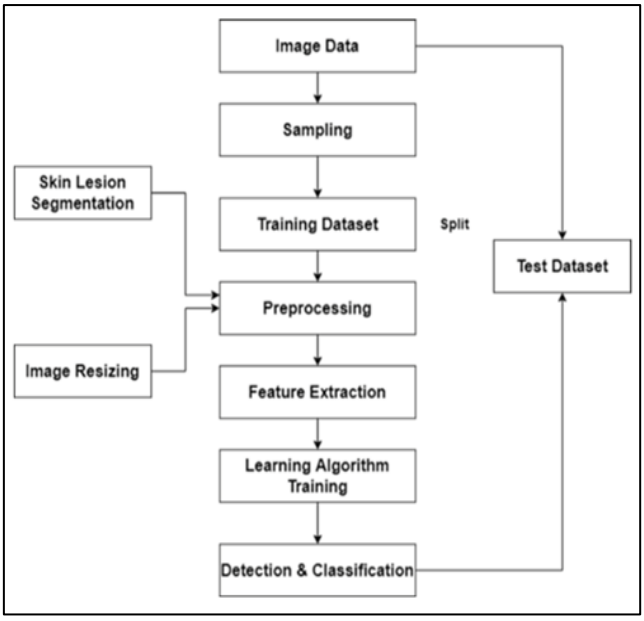


Fig. 2. FigFigBlock diagram of Skin cancer recognition

A. Dataset Preparation

The dataset for segmentation and classification of Skin lesions is HAM1000, and it is available at https://www.kaggle.com/datasets/surajghuwalewala.

The comprehensive HAM10000 dataset substantially contributes to computer-aided skin cancer diagnosis research. It includes 10,000 excellent images of skin lesions. A dermatoscope was used to capture each image, with a 3000 × 2000 pixels resolution. The dermatoscopes are calibrated to 100x, the same magnification as a light microscope. The dataset classifies the skin lesions into multiple categories, such as basal cell carcinoma (BCC), nevus, and melanoma. The examples include images of benign and malignant lesions, giving the dataset the depth and breadth necessary for accurate AI training.

B. Dataset preprocessing

The median filter is a popular method for image processing that reduces noise in skin lesion photos and improves the features' sharpness. To use it, first, choose a pixel to work on and then calculate the neighborhood's median pixel values. We select that neighborhood based on size and specific shape, typically square or circular. Once determined, we change the pixel value to the median value. The median filter only allows for the neighborhood's size and shape changes. Nonetheless, it works well on photos with edges and eliminates impulsive noise.

Noise reduction using a median filter in skin lesion photos can help increase the precision of later image analysis activities, such as feature extraction or machine learning-based categorization. During the image capture, noise of various kinds, such as speckles or random fluctuations in pixel intensity, can appear in skin lesion images.

Although median filtering is effective in some circumstances, the features of the picture noise and the objectives of the image processing must be considered when selecting a filtering method. Moreover, carefully applying any image processing method is crucial, as it can impact the interpretation of medical images.

C. Dataset Splitting

Splitting the dataset is a critical step in creating a machine-learning model. This method divides the dataset into two subsets: training (80%) and validation (20%). The training set is a collection of examples with known labels used to build a model. During training, the model uncovers patterns and connections among the labels in the training set. By evaluating the model on untested data, the validation set helps with hyperparameter adjustment and model evaluation during training, avoiding overfitting and guaranteeing optimal performance. Lastly, the testing set objectively gaits the model's performance in real-world situations by evaluating the model's generalization abilities on entirely new data. This thorough dataset separation technique makes it easier to build reliable machine-learning models to make precise predictions and perform well in generalization.

D. Classification of the skin lesion using a deep learning algorithm

This method employed CNN, Densenet201, Xception, and hybrid CNN-SVM algorithms. This section presents the detailed architecture of all the algorithms.

1) CNN: Image recognition and classification are the specialties of convolutional neural networks (CNNs). They are multi-layered feed-forward neural networks composed of filters and filter banks that extract image features. By adjusting filter weights, CNNs can detect specific features, such as edges or curves. A typical CNN architecture alternates between convolutional and pooling layers, followed by one or more fully connected layers. Fig. 3 depicts the architecture of CNN.

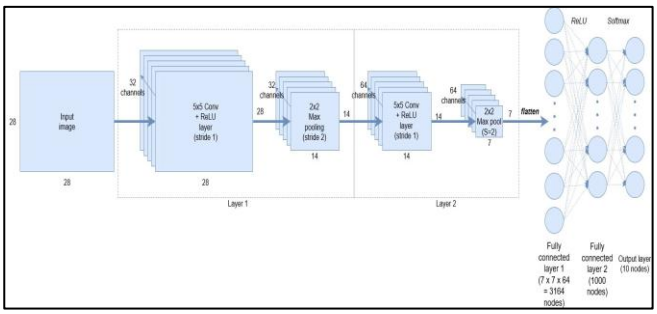


Fig. 3. Architecture Diagram of CNN algorithm

a) Convolutional Layer: The convolutional layer in a CNN is essential for feature extraction from images, identifying key features like edges and textures. It reduces the number of parameters using a sparse connection method, where filters with fixed weights move over the image to detect specific features. This process, known as "feature mapping," uses multiple filters (channels) to produce distinct outputs. Each filter learns different features, enhancing the network's image understanding. For grayscale images, the convolution layer's output is three-dimensional, representing the number of channels and their outputs.

b) ReLU Layer: ReLU (Rectified Linear Unit) is a non-linear activation function that operates element-wise on a feature map. It transforms values less than or equal to zero to zero, effectively highlighting only the positive values. This can be mathematically expressed as $F(x) = \max(0, x)$), where values below zero are replaced with zero, emphasizing the positive aspects of the feature map.

c) Pooling Layer: Each activation map experiences a decrease in dimensionality. However, the pooling layer retains the most critical data. The images provided are intended to create several non-overlapping rectangles. Moreover, now, what is pooling? Like many others, pooling is a sliding window approach, but instead of using tunable weights, it applies some statistical function to the contents of its window. Max pooling is the most commonly used form of pooling; it uses the max() function on the contents of its window. We occasionally use other variations, such as mean pooling, which calculates the statistical mean of the contents. In this chapter, we will focus on max pooling. The following diagram illustrates the max pooling process.

d) Flattening Layer: Convolutional neural networks (CNNs) effectively create object representations from high-resolution data. However, to achieve a final classification, a traditional classifier needs to be added to interpret the network's rich output. This involves flattening the CNN output into a one-dimensional vector before applying the classifier. The pooled feature map, derived from pooling operations, is essential for understanding and processing the CNN output, providing more than just a human-readable summary.

e) Fully Connected Layer: The fully connected layer (FCL) is used to classify images based on features extracted from the CNN. A Softmax activation function classifier processes the signals from the FCL, producing a list of probabilities for each class label. These probabilities indicate the network's confidence in each class, essentially rating the class labels. Adding the FCL to a CNN can effectively classify images into various categories.

2) Vgg16: Each year, ImageNet organizes the Large Scale Visual Recognition Challenge (ILSVRC), a significant computer vision competition with numerous participating teams. Competitors first tackle the challenge of localizing objects in images, then progress to the more complex task of categorizing elements within those images. One notable

entry came from researchers at the University of Oxford in the UK, which generated significant excitement in the field. The architecture of the Vgg16 model is presented in Fig. 4.

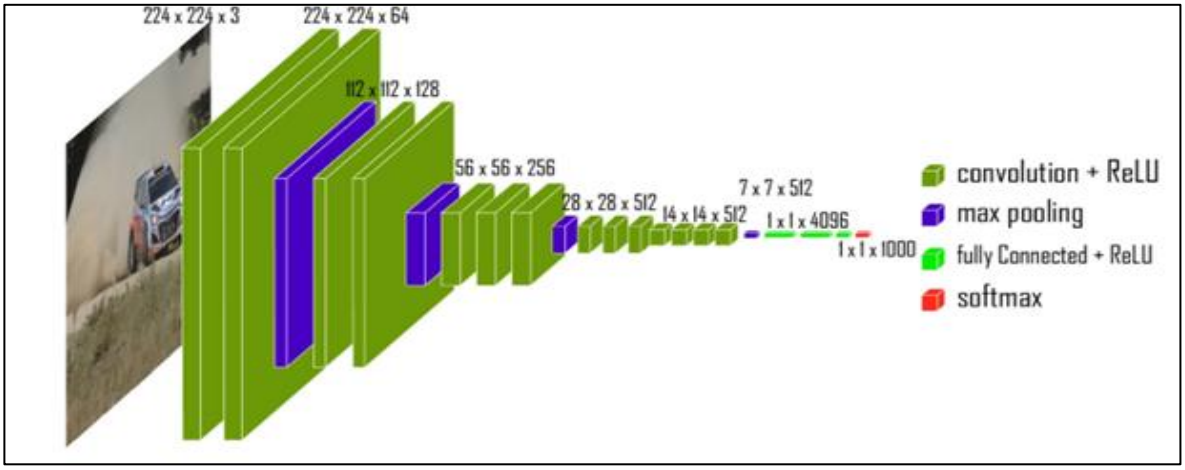


Fig. 4. VGG-16 model architecture

The ImageNet dataset, with 14 million images across 1,000 categories, enables the model to attain an impressive top-5 test accuracy of 92.7%. The model analyzes input images measuring 224 × 224 × 3 pixels to achieve this precision. The initial two layers of the network have 64 channels and employ a 3 × 3 filter size with "SAME" padding in TensorFlow. This padding method preserves the spatial dimensions of the image throughout the layers.

The model employs 256 filters, measuring 3 × 3, in the subsequent two convolutional layers. Subsequently, it comprises two groups of three convolutional layers, each succeeded by a max-pooling layer. Each set of 512 filters of dimensions 3 × 3, with padding values of (3, 3). Following the final max-pooling layer, the image undergoes processing through two convolutional layers utilizing 3 × 3 filters instead of the bigger 7 × 7, ZF-11 × 11, or AlexNet filters.

The stack compresses a (7, 7, 512) feature map into a feature vector 25,088, incorporating a convolutional and max-pooling layer. Three completely connected layers manipulate this vector: the first produces a vector of dimensions (1, 4096), the second similarly produces a vector of dimensions (1, 4096), and the third generates a vector of size (1, 1000), corresponding to the 1,000 classes from the ILSVRC challenge. A softmax layer accepts the output from the third layer to normalize the classification vector. All concealed layers employ ReLU as their activation function.

The objective of the ImageNet research project is to create a comprehensive library of photos accompanied by annotations, including labels. Prior computer vision tests have illustrated the efficacy of models including InceptionV1, InceptionV2, VGG-16, and VGG-19, all of which were trained on ImageNet. We developed them from inception and trained them on an extensive dataset comprising over 14 million images over about 20,000 categories. The models are vast and profound because of the substantial amount of image data, which renders them particularly effective in feature extraction from images. The pre-trained models of the image annotation project can refine computer vision tasks designated for diverse pictures across many categories.

3) Xception: In 2016, François Chollet proposed the Xception architecture, an enhanced iteration of the Inception network designed to overcome the limits of conventional CNNs and improve their functionality. "Xception" denotes "Extreme Inception," signifying a more profound iteration of the Inception module. Xception substitutes conventional convolutional layers with depthwise separable convolutions. This method bifurcates convolution into two stages: depthwise convolution, which utilizes a singular filter for each input channel, and pointwise convolution, which executes a 1x1 convolution to amalgamate data across channels. Figure 5 illustrates the design of the Xception algorithm.

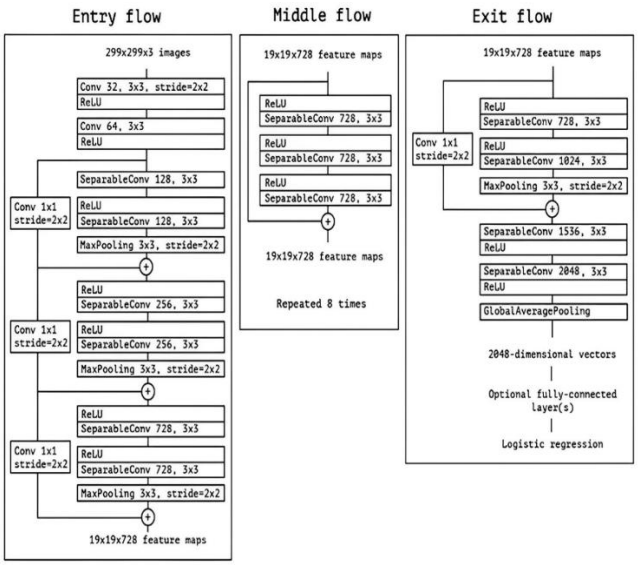


Fig. 5. The architecture of the Xception model

The Xception module, seen in Figure 5.12, comprises three primary components: the entry flow, the middle flow, and the exit flow. The input flow commences with two layers of depthwise separable convolutions, succeeded by ReLU activations and residual connections. We incorporate max pooling layers and diverse forms of separable convolutions, specifying the precise strides for each layer. Utilize 'ADD' operations for skip connections instead of concatenating tensors. The input flow transforms an image from dimensions 299 × 299 × 3 to 19 × 19 × 728. The graphic delineates image dimensions, layer configurations, filter quantities and shapes, pooling procedures, and the optional fully linked layer in the intermediate and exit pathways.

4) DenseNet201: Transfer learning is exceptionally efficient for classification problems with limited datasets, and Deep Transfer Learning (DTL) can further improve outcomes. This research presents a DTL model founded on DenseNet201, utilizing a convolutional neural network architecture with ImageNet weights for feature extraction. DenseNet interlinks each layer with all preceding layers in a feed-forward configuration, mitigating the vanishing gradient issue by guaranteeing each layer has input from all prior layers. This approach expands each layer's input and output space while preventing the model from becoming excessively huge by concatenating outputs from all preceding layers at each step. Figure 6 depicts the architecture of the Densenet method.

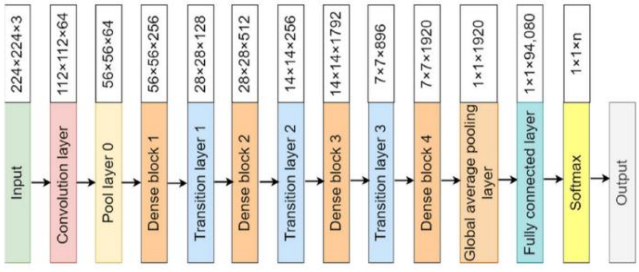


Fig. 6. Architecture of Densenet201

DenseNet201 is a convolutional neural network belonging to the DenseNet model family, recognized for its dense connectivity architecture. In DenseNet, every layer obtains inputs directly from all prior layers, alleviating the vanishing gradient issue, enhancing feature reuse, and promoting feature propagation. This design improves the network's efficiency and performance. A thorough examination of the DenseNet201 architecture is available for evaluation.

- DenseNet201: This iteration of the DenseNet architecture comprises 201 layers, including activation, batch normalization, convolutional, and additional layers. The "201" denotes the network's complexity, rendering it appropriate for intricate jobs and extensive datasets.

- Pre-trained Weights: DenseNet201 is frequently pre-trained on massive datasets like ImageNet. This pre-training allows the model to acquire hierarchical features that may be refined for individual tasks, improving its performance across diverse applications.
- Architecture: DenseNet201 utilizes a deep convolutional network characterized by dense connectivity, bottleneck layers, and transition blocks. This design enables the recording of complicated data patterns, rendering it effective for picture categorization and many computer vision tasks. Employing pre-training and transfer learning enhances its functionalities across various applications.

E. Performance Evaluation

The proposed approach is evaluated using several performance metrics, including accuracy, recall, F1-score, and precision. These are conventional metrics employed to assess classifiers. Their merit is in providing divergent viewpoints on the efficacy of a classifier. Furthermore, they may be easily derived from the confusion matrix, an additional advantageous characteristic.

- Accuracy: By computing the ratio of correct predictions to total predictions, accuracy assesses the overall correctness of the classifier's predictions. It has the following definition:

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (4)$$

- where the numbers represent the number of accurate positive predictions (TP), true negative predictions (TN), false positive predictions (FP), and false negative predictions (FN). Although accuracy gives a broad picture of the classifier's performance, imbalanced datasets might not fit it well.
- Precision: The percentage of accurately anticipated positive cases among all positively predicted instances is the subject of precision. It is computed as follows:

$$Precision = \frac{TP}{(TP + FP)} \quad (5)$$

A classifier's precision tells us how well it avoids producing false positive results. A better precision shows a reduced rate of misclassifying negative cases as positive.

- Recall, also known as True Positive Rate or Sensitivity, quantifies the percentage of accurately anticipated positive cases among all positive cases. It is computed as follows:

$$Recall = \frac{TP}{(TP + FN)} \quad (6)$$

Recall highlights the classifier's ability to identify positive instances correctly, and it is advantageous when the goal is to minimize false negatives.

- F1 score: The F1 score is a metric that balances recall and precision by combining both into one. It is computed as follows and is the harmonic mean of recall and precision:

$$F1\ Score = \frac{2 * (Precision * Recall)}{(Precision + Recall)} \quad (7)$$

We equilibrate precision and recall by evaluating false positives and negatives in the F1 score. It is beneficial when class distribution is imbalanced or recall and precision are equivalent.

These metrics are essential for binary classification problems due to the presence of positive and negative classes. These metrics can also be utilized for multi-class classification problems by calculating them independently for each class and subsequently averaging the results (e.g., micro-averaging, macro-averaging).

Prioritizing a set of metrics necessitates meticulously evaluating your categorization task's specific requirements and characteristics. For example, memory may be vital in medical diagnosis to minimize false negatives, while precision may be more essential in spam email classification to avert false positives.

Result

This section presents the results of skin cancer detection and recognition.

F. Results of Skin cancer segmentation

The training loss and accuracy obtained for the SegNet for 100 epochs are presented in Fig. 7.

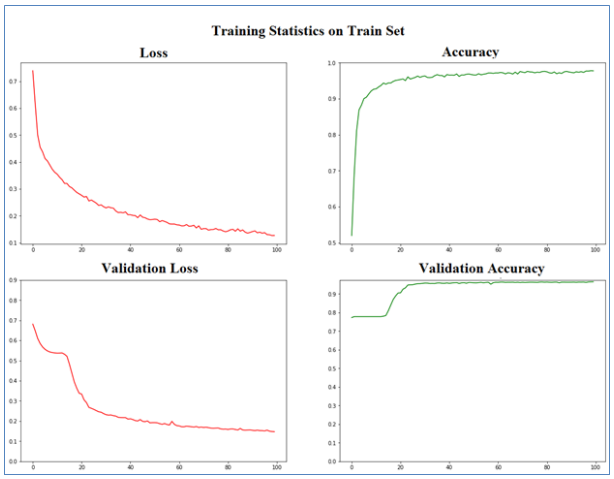


Fig. 7. Training progress graph of SegNet for skin cancer segmentation

From Fig. 7, it is observed that the training loss of the algorithm decreases as the epoch increases while the accuracy of detection increases. The proposed system achieved good detection accuracy with lower losses. Table I tabulates the performance of the SegNet algorithm on various testing images.

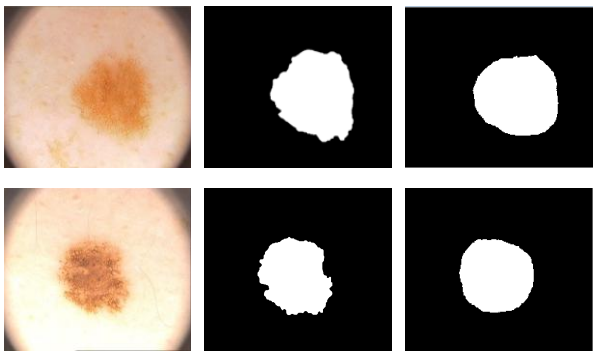
TABLE I. PERFORMANCE OF THE SEGNET FOR SKIN CANCER SEGMENTATION

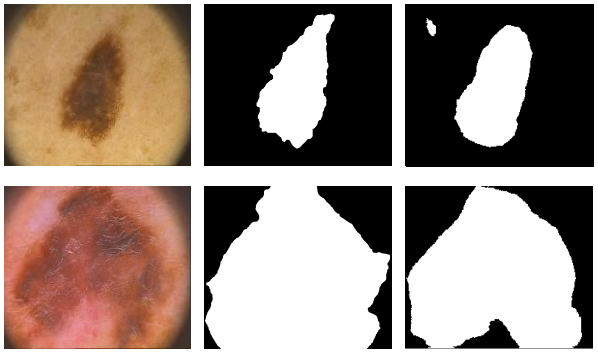
Image Name	IOU	DI	Precision	Recall	Accuracy
IMD390	88.6487	93.6959	94.1563	93.3030	97.5891
IMD392	91.3809	93.10577	91.7550	94.4533	97.9594
IMD393	82.5091	91.7462	96.8662	87.0751	91.6137
IMD394	81.4715	88.3025	99.0446	79.4143	95.4935
IMD395	84.5796	93.0471	88.3356	98.1681	95.2107
IMD396	78.6634	89.8264	88.3665	91.2851	95.7885
IMD397	77.4382	90.6989	95.4223	86.3694	93.7947
IMD398	73.4582	85.7910	97.364	76.6427	80.8492
IMD399	83.2597	90.0542	83.1518	98.0891	95.8943
IMD400	70.5491	88.2608	99.5692	79.2122	90.3401
IMD402	86.0697	93.4688	96.6370	90.4492	96.8892
IMD403	86.0531	93.5515	98.5809	88.9933	91.0624
IMD404	72.8725	88.1839	80.7587	97.0557	87.9130
IMD405	81.5398	88.6892	83.671	94.1985	96.0306
IMD406	79.9282	91.7546	94.6498	88.9359	92.2932
IMD407	79.8268	90.9848	96.39	86.0291	93.4773
IMD408	92.1096	95.8868	97.3147	94.4909	92.8751
IMD409	85.8480	92.1197	96.4400	88.1463	90.1570

IMD410	74.6085	86.7594	97.7357	77.9552	86.9954
IMD411	96.4960	98.3061	99.3681	97.2624	96.7285
IMD413	80.7056	89.5501	100.0	81.0572	82.0800
IMD417	89.1924	94.4246	97.3593	91.6499	89.8010
IMD418	86.7094	93.0006	92.5390	93.4416	92.0043
IMD419	91.7258	96.3433	93.1384	99.7670	93.9066
IMD420	79.8968	88.3287	88.1336	88.4920	82.8938
IMD421	85.4635	92.8264	91.8733	93.7829	87.0442
IMD423	74.4015	89.6570	82.1417	98.6506	85.0484
IMD424	76.5431	86.9446	100.0	76.8805	77.2705
IMD425	50.2752	67.0301	100.0	50.3504	57.9203
IMD426	50.7592	70.3199	89.8994	57.6574	71.9563
IMD427	89.4674	95.4487	92.1210	98.9965	96.8892
IMD429	86.8115	91.9125	87.0902	97.2628	96.8404
IMD430	90.1406	94.4177	91.0153	98.2227	97.6155
IMD431	85.4778	94.7782	96.5293	93.0510	95.3369
IMD432	84.3751	91.3896	91.7192	91.0015	96.3277
IMD433	86.1530	84.3162	76.5241	93.6214	95.7946
IMD434	84.3383	89.3746	86.8090	91.9834	95.6685
IMD435	82.9337	91.1022	89.9630	92.2472	85.9456
IMD436	88.5409	94.9120	95.7019	94.0830	94.3074
IMD437	83.0516	94.5458	92.9558	96.1681	94.5271

Table I presents the performance metrics of a SegNet model for skin cancer segmentation over multiple testing images. The metrics include intersection over union (IoU), dice coefficient (DI), precision, recall, and accuracy. Table I reveals a promising detection accuracy.

Fig. 8 presents the qualitative analysis of the proposed system.





(a) (b) (c)

Fig. 8. Qualitative analysis of the proposed system (a) Input image, (b) Groundtruth, and (c) Output of the proposed system

We developed the proposed system using Python. The qualitative analysis of the developed system focuses on assessing the correctness of skin lesion segmentation, and the results indicate a notable achievement in minimizing false positives. Qualitative analysis involves a thorough visual examination of the segmentation outcomes, comparing them to ground truth or reference images to ensure accuracy. The emphasis on lower false positives is crucial in medical image analysis, as it signifies a reduced likelihood of incorrectly identifying non-lesion areas as lesions.

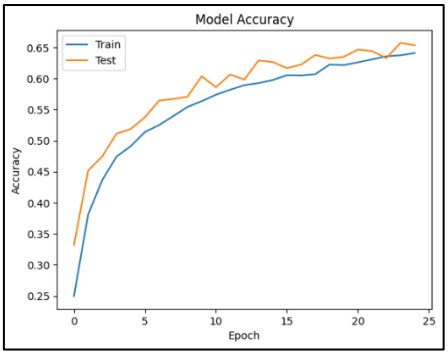
G. Skin cancer Recognition

The proposed system is developed to classify skin cancer into different types. The results of different deep-learning algorithms for classifying skin lesions into seven categories are presented in this section.

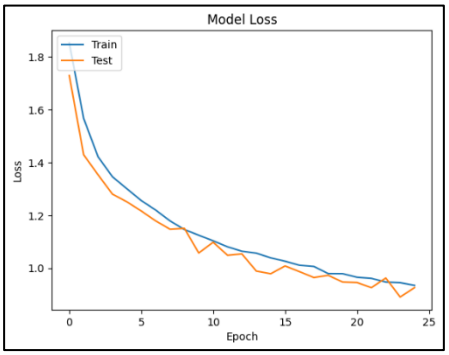
1) Results of Skin Cancer Recognition Using Deep Learning Algorithm

The proposed system is developed to classify skin cancer into different types.

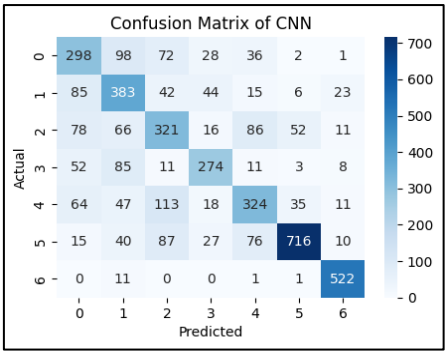
a) CNN: The CNN algorithm's results for classifying skin cancer into seven distinct types are shown in Fig. 9 below.



(a)



(b)



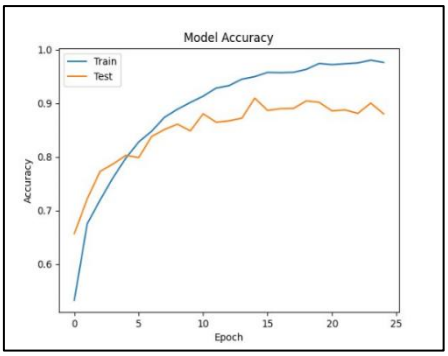
(c)

Classification Report	precision	recall	f1-score	support
akiec	0.50	0.56	0.53	535
bcc	0.52	0.64	0.58	598
bkl	0.50	0.51	0.50	630
df	0.67	0.62	0.64	444
mel	0.59	0.53	0.56	612
nv	0.88	0.74	0.80	971
vasc	0.89	0.98	0.93	535
accuracy			0.66	4325
macro avg	0.65	0.65	0.65	4325
weighted avg	0.67	0.66	0.66	4325

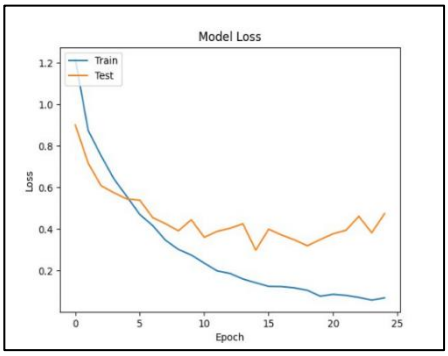
(d)

Fig. 9. Training performance of CNN on HAM10000 Dataset (a) Accuracy (b) Loss (c) Confusion Matrix (d) Classification report

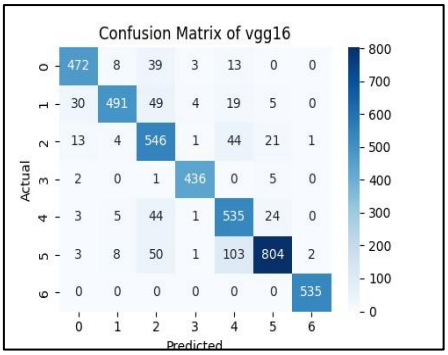
b) Vgg16: The results of the Vgg16 algorithm for classifying skin cancer into 7 different types are presented below in Fig.10.



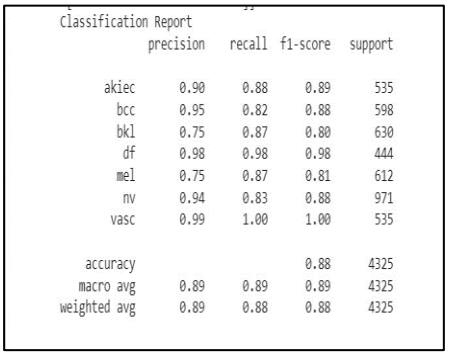
(a)



(b)



(c)



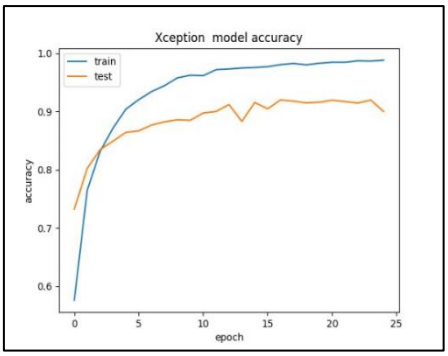
Classification Report

	precision	recall	f1-score	support
akiec	0.90	0.88	0.89	535
bcc	0.95	0.82	0.88	598
bkl	0.75	0.87	0.80	630
df	0.98	0.98	0.98	444
mel	0.75	0.87	0.81	612
nv	0.94	0.83	0.88	971
vasc	0.99	1.00	1.00	535
accuracy			0.88	4325
macro avg	0.89	0.89	0.89	4325
weighted avg	0.89	0.88	0.88	4325

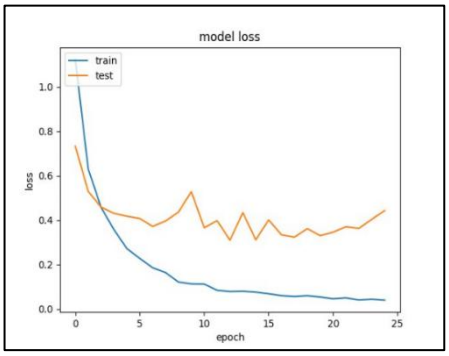
(d)

Fig. 10. Training performance of Vgg16 on HAM10000 Dataset (a) Accuracy (b) Loss (c) Confusion Matrix (d) Classification report

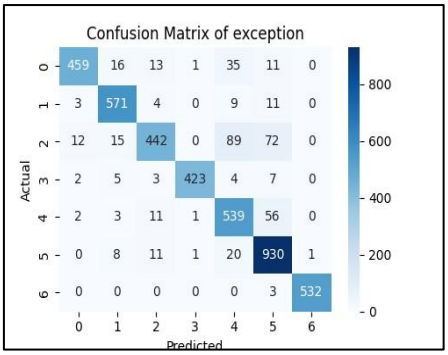
c) Xception: The results of the Xception algorithm for classifying skin cancer into 7 different types are presented below in Fig.11.



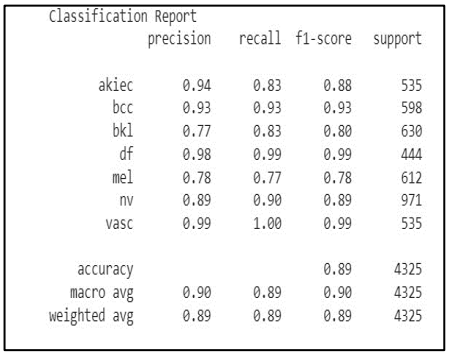
(a)



(b)



(c)



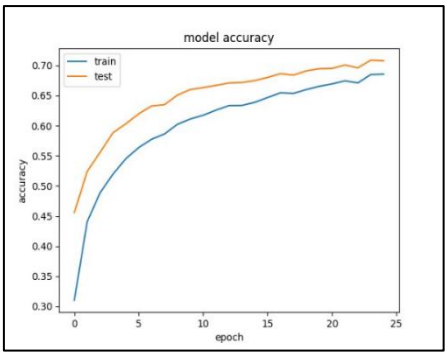
Classification Report

	precision	recall	f1-score	support
akiec	0.94	0.83	0.88	535
bcc	0.93	0.93	0.93	598
bkl	0.77	0.83	0.80	630
df	0.98	0.99	0.99	444
mel	0.78	0.77	0.78	612
nv	0.89	0.90	0.89	971
vasc	0.99	1.00	0.99	535
accuracy			0.89	4325
macro avg	0.90	0.89	0.90	4325
weighted avg	0.89	0.89	0.89	4325

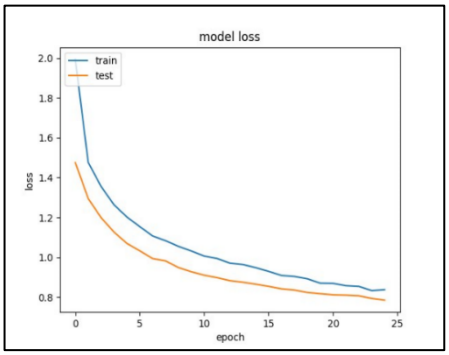
(d)

Fig. 11. Fig. 6.5 Training performance of Xception on HAM10000 Dataset (a) Accuracy (b) Loss (c) Confusion Matrix (d) Classification report

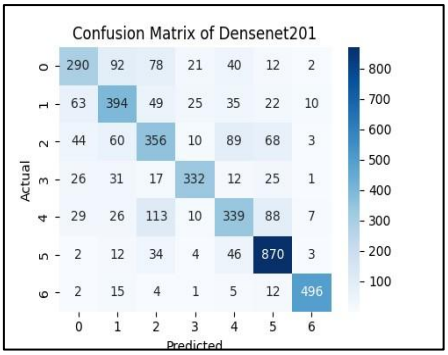
d) Densenet201: The results of the Densenet201 algorithm for classifying skin cancer into 7 different types are presented below in Fig.12.



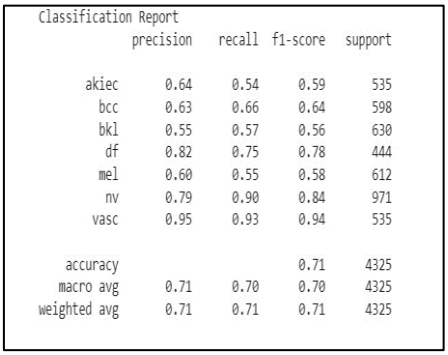
(a)



(b)



(c)



Classification Report

	precision	recall	f1-score	support
akiec	0.64	0.54	0.59	535
bcc	0.63	0.66	0.64	598
bkl	0.55	0.57	0.56	630
df	0.82	0.75	0.78	444
mel	0.60	0.55	0.58	612
nv	0.79	0.90	0.84	971
vasc	0.95	0.93	0.94	535
accuracy			0.71	4325
macro avg	0.71	0.70	0.70	4325
weighted avg	0.71	0.71	0.71	4325

(d)

Fig. 12. Training performance of Densenet201 on HAM10000 Dataset (a) Accuracy (b) Loss (c) Confusion Matrix (d) Classification report

The comparative analysis of the three algorithms for the classification of skin cancer into seven different types is presented in Table II.

TABLE II. COMPARATIVE ANALYSIS OF DIFFERENT CLASSIFIERS ON THE HAM10000 DATASET

Algorithm	**Precision**	**Recall**	**F1-Score**	**Accuracy**
CNN	0.65	0.66	0.66	0.66
Vgg16	0.89	0.88	0.88	0.88
Xception	0.89	0.89	0.89	0.89
Densenet	0.71	0.71	0.71	0.71

Table 6.1 compares different classifiers applied to the HAM10000 dataset, measuring performance in terms of precision, recall, F1 score, and total accuracy. We evaluate three algorithms: Xception, DenseNet, and CNN. CNN's accuracy is 0.66, recall is 0.66, F1-score is 0.66, and precision is 0.65. Although CNN does quite well across various parameters, it could be more decisive in any area. Xception, on the other hand, performs better, with accuracy, recall, and F1-score all at 0.90 and precision at 0.90. This shows that Xception maintains a balanced performance in terms of precision and recall while achieving high levels of accuracy. Densenet201's precision, recall, F1-score, and accuracy are all 0.72, placing it in the middle of the other two classifiers. It suggests consistent performance across the assessed measures but falls below Xception. Concerning precision, recall, and F1-score metrics, Xception seems to be the most successful classifier for the HAM10000 dataset, according to this analysis, which also shows a high degree of accuracy.

A comparative analysis of the proposed system, along with recent work, is presented in Table III.

TABLE III. COMPARATIVE ANALYSIS OF THE PROPOSED SYSTEM WITH RECENT WORK

Algorithm	**Accuracy (%)**
Shifted MobileNetV2 [17]	81.90
Shifted GoogLeNet [17]	80.50
9 layered CNN [18]	80.00
Vgg16+googLeNet ensemble [19]	81.50
Our CNN	66
Our Vgg16	88
Our Xception	89
Our Densenet	71

The table presents a comparative analysis of a proposed system for skin lesion classification against recent work in the field, using accuracy as the evaluation metric. Various algorithms are listed alongside their respective accuracy percentages. The Shifted MobileNetV2 and GoogLeNet algorithms, referenced from a prior study [17], achieved an accuracy of 81.90% and 80.50%, respectively. Another recent model, a 9-layered CNN [18], reported 80.00% accuracy. Additionally, an ensemble method combining Vgg16 and GoogLeNet [19] resulted in 81.50% accuracy.

In comparison, the proposed system includes several models with varying performance. The custom CNN developed in the system reached a lower accuracy of 66%, while Vgg16 and Xception achieved higher accuracy rates of 88% and 89%, respectively, outperforming the previous models. The Densenet in the proposed system attained an accuracy of 71%, showing moderate performance. Overall, the highest accuracy in the proposed system was observed with the Xception model, demonstrating its superior performance for skin lesion classification in this analysis.

Conclusion

The proposed system effectively enhances skin cancer detection and recognition using the PH2 dataset, applying augmentation techniques like rotation and flipping for better generalization. The SegNet architecture carefully balances the extraction of features and spatial information. Stochastic gradient descent (SGD), batch normalization, and activation functions are optimization methods that make the model work better. Detailed explanations of the training process ensure transparency and reproducibility. The study demonstrates how CNNs improve skin cancer detection, particularly VGG16, Xception, and Densenet201. It highlights Xception's superior accuracy, precision, recall, and F1-score performance. Densenet201's connectivity and VGG16's simplicity are vital considerations, with a thorough evaluation suggesting areas for future improvement. The research advances diagnostic tools for skin cancer, emphasizing the importance of ongoing innovation in medical image processing to improve patient outcomes and public health.

The study provides a strong foundation for advancing automated skin cancer detection and categorization. Future improvements could focus on developing customized models tailored to skin cancer classification, potentially surpassing the accuracy of existing architectures like Xception and Densenet201. Expanding datasets to include diverse populations and clinical scenarios could enhance the generalizability and robustness of these models. Additionally, integrating patient demographics, histopathology, and clinical data could significantly improve diagnostic accuracy beyond visual inspection, further advancing the field of dermatological image analysis.

Competing Interests

The authors declare that there are no competing interests related to this study.

Funding Information

This research was conducted without external funding. The authors utilized publicly available datasets and resources for the analysis.

Author contribution

Yashwant S. Ingle conceptualized the study and supervised the project. Yashwant S. Ingle performed the data analysis and algorithm development. Yashwant S. Ingle and Dr. Nuzhat Faiz Shaikh contributed to the literature review and manuscript writing. All authors reviewed and approved the final manuscript.

Data Availability Statement

The dataset used in this study is publicly available and can be accessed at [insert URL or repository name]. All relevant data and code used for the analysis are available upon reasonable request from the corresponding author.

Research Involving Human and /or Animals

This study did not involve any human participants or animals. The analysis was conducted using existing publicly available datasets.

Informed Consent

As this research utilized publicly available datasets, informed consent was not required. All data were anonymized and ethically sourced in accordance with relevant guidelines.

References

[1] U.-O. Dorj, K.-K. Lee, J.-Y. Choi, and M. Lee, "The skin cancer classification using deep convolutional neural network," Multimedia Tools and Applications, vol. 77, no. 8, pp. 9909–9924, 2018.

[2] A. Rezvantalab, H. Safigholi, and S. Karimijeshni, "Dermatologist level dermoscopy skin cancer classification using different deep learning convolutional neural networks algorithms," 2018.

[3] K. M. Hosny, M. A. Kassem, and M. M. Foaud, "Skin Cancer Classification using Deep Learning and Transfer Learning," in 2018 9th Cairo International Biomedical Engineering Conference (CIBEC), pp. 90–93, Cairo, Egypt, 2018. [Online].

[4] A. Dascalu and E. O. David, "Skin cancer detection by deep learning and sound analysis algorithms: a prospective clinical study of an elementary dermoscopy," eBioMedicine, vol. 43, no. May, pp. 107–113, 2019.

[5] T. C. Pham, G. S. Tran, T. P. Nghiem, A. Doucet, C. M. Luong, and V.-D. Hoang, "A Comparative Study for Classification of Skin Cancer," in 2019 International Conference on System Science and Engineering (ICSSE), pp. 267–272, Dong Hoi, Vietnam, 2019.

[6] A. Heller, J. S. Utikal, A. H. Enk, et al., "Superior skin cancer classification by the combination of human and artificial intelligence," European Journal of Cancer, vol. 120, pp. 114–121, 2019.

[7] T. Emara, H. M. Afify, F. H. Ismail, and A. E. Hassanien, "A Modified Inception-v4 for Imbalanced Skin Cancer Classification Dataset," in 2019 14th International Conference on Computer Engineering and Systems (ICCES), pp. 28–33, Cairo, Egypt, 2019.

[8] S. Mohapatra, N. V. S. Abhishek, D. Bardhan, A. A. Ghosh, and S. Mohanty, "Skin Cancer Classification Using Convolution Neural Networks," in Advances in Distributed Computing and Machine Learning, A. Tripathy, M. Sarkar, J. Sahoo, K. C. Li, and S. Chinara, Eds., vol. 127 of Lecture Notes in Networks and Systems, Springer, Singapore, 2021.

[9] S. Jinnai, N. Yamazaki, Y. Hirano, Y. Sugawara, Y. Ohe, and R. Hamamoto, "The development of a skin cancer classification system for pigmented skin lesions using deep learning," Biomolecules, vol. 10, no. 8, p. 1123, 2020.

[10] S. S. Chaturvedi, J. V. Tembhurne, and T. Diwan, "A multi-class skin cancer classification using deep convolutional neural networks," Multimedia Tools and Applications, vol. 79, no. 39-40, pp. 28477–28498, 2020.

[11] Mehwish Dildar, Shumaila Akram, Muhammad Irfan, Hikmat Ullah Khan, Muhammad Ramzan, "Skin Cancer Detection: A Review Using Deep Learning Techniques," Int J Environ Res Public Health. 2021

[12] Nourelhoda M. Mahmoud & Ahmed M. Soliman, "Early automated detection system for skin cancer diagnosis using artificial intelligent techniques," Scientific Reports, 2024

[13] Umesh Kumar Lilhore, Sarita Simaiya, Yogesh Kumar Sharma, "A precise model for skin cancer diagnosis using hybrid U-Net and improved MobileNet-V3 with hyperparameters optimization", Scientific Reports, 2024
[14] Balaha, H. M. & Hassan, A.E.-S. Skin cancer diagnosis is based on deep transfer learning and a sparrow search algorithm. Neural Comput. Appl, 2023
[15] Tabrizchi, H., Parvizpour, S. & Razmara, J. An improved VGG model for skin cancer detection. Neural Process. Lett. 2023.
[16] Adla, D. et al. A full-resolution convolutional network with a dynamic graph cut algorithm for skin cancer classification and detection. Healthcare Analytics (2023).
[17] Thurnhofer-Hemsi, K., López-Rubio, E., Domínguez, E. & Elizondo, D. A. Skin lesion classification by ensembles of deep convolutional networks and regularly spaced shifting. IEEE Access 9, 112193–112205 (2021).
[18] Nugroho, A. A., Slamet, I. & Sugiyanto,. Skins cancer identification system of HAMl0000 skin cancer dataset using convolutional neural network. AIP Conf. Proc. 2202(1), 020039 (2019).
[19] Mobiny, A., Singh, A. & Van Nguyen, H. Risk-aware machine learning classifier for skin lesion diagnosis. J. Clin. Med. 8(8), 1241 (2019).