A NOVEL ATTENTION-BASEDSKIN CANCER DETECTION USING DERMATOLOGICAL IMAGES

VIDHYASHREE B [1], Dr PRASANNA LAKSHMI G S [2], CHETLA CHANDRA MOHAN [3], G KIRAN KUMAR [4], Dr CH BHAVANI [5], ANNAMARAJU THANUJA [6], PAVINDRA REDDY PONDUGALA [7], MELAM NAGA RAJU [8]

[1] Research Scholar, Department of Computer Science and Engineering, Nagarjuna College of Engineering and Technology, Visvesvaraya Technological University, Belagavi - 590018
[2] Research Supervisor, Associate Professor, Department of Computer Science and Engineering, Nagarjuna College of Engineering and Technology, Visvesvaraya Technological University, Belagavi - 590018
[3] Assistant Professor, Department of Information Technology, Prasad V Potluri Siddhartha Institute of Technology, Vijayawada, Andhra Pradesh, India
[4] Assistant professor, Department of Computer Science and Engineering (data science), Madanapalle Institute of Technology & Science , Madanapalle, Andhra Pradesh ,India
[5] Associate Professor, Department of Computer Science and Engineering, CVR College of Engineering, Hyderabad, Telengana, India.
[6] Assistant Professor, Department of Computer Science and Engineering, Koneru Lakshmaiah Education foundation ,Vaddeswaram, Andhra Pradesh ,India
[7] Assistant Professor, Department of Information Technology, R.V.R & J.C College of Engineering, Chowdavaram, Guntur Andhra Pradesh ,India
[8] Associate Professor, Department of Information Technology, Seshadrirao Gudlavalleru Engineering College, Gudlavalleru,Andhra Pradesh ,India
Emails: [1]vshree56@gmail.com, [2]prasannalakshmigs@gmail.com ,[3]chetlachandramohan123@gmail.com, [4]kirankumarg@mits.ac.in, [5]saravanan.ba.mct@snsce.ac.in, [6]a.thanuja@kluniversity.in
[7] pavindra.reddyppr@gmail.com [8] melam.nagaraju5810@gmail.com

ABSTRACT

Skin cancer is characterized by the abnormal growth of skin cells, caused by various factors. Recently, the number of people affected by skin cancer has been steadily increasing. Early diagnosis plays a vital role in improving patient outcomes, while late detection can lead to complications and, in some cases, mortality. Traditionally, skin cancer detection relies on visual inspection by dermatologists, a process that can be time-consuming, requires significant expertise, and is prone to errors. These challenges have attracted more researchers to develop autonomous systems for skin cancer detection. However, several issues remain unresolved, such as achieving high accuracy, which is still difficult due to minimal variation between benign and malignant cells, as well as the dermatological images being obscured by hair, moles, or skin texture. While some models achieve higher accuracy, they are highly complex and unsuitable for deployment on edge devices. To address this, the research proposed a novel Attention-based MobileNet-Transformer for skin cancer detection. The processed images are given to MobileNet for effective feature extraction with minimal parameters, and the extracted features are refined using a Convolutional Block Attention Module (CBAM) attention mechanism. The refined features are then given to a transformer module equipped with a fully connected layer (FCL) for capturing temporal dependencies and performing classification. The proposed model effectively extracts all the features from the skin images to achieve better accuracy. When compared with existing models like MobileNet, Xception, and DenseNet, the proposed approach achieved the highest accuracy of 94%, whereas the others remained below 90%. The ablation research and state-of-the-art comparison highlight the model's effectiveness. The experimental analysis shows that the proposed network is suitable for real-time skin cancer detection because of its reliable output and lower model complexity.

Keywords—*Skin Cancer, Dermatological Images, Deep Learning, MobileNet, Transformer, Attention Mechanism, Confusion Matrix*

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

1. INTRODUCTION

Recently, skin cancer has become a common type of cancer and has a significant impact on patient survival [1]. Early identification of the disease is important for preventing the progression of skin cancer and reducing cancer-related deaths. As per data from the American Cancer Society, skin cancer accounts for a 75% mortality rate, highlighting its severity as a major global health concern [2]. Furthermore, the WHO estimates that one-third of the world's population has been diagnosed with skin cancer. Moreover, the frequency of skin cancer has steadily increased over the past few decades. According to medical data from 2020, 1.5 million people were diagnosed with skin cancer, and 57,000 died from the disease worldwide [3]. Melanoma cases were documented in the United States in 2022, totalling 97,920 [4]. According to scientific studies, being exposed to the sun's UV rays elevates the risk of developing skin cancer by destroying skin cells. Other factors that influence the proliferation of malignant cells include alcohol consumption, tobacco usage, and environmental pollution.Skin cancer is caused by the proliferation of benign or malignant tumors in the epidermis. Malignant and pre-malignant tumors are the most fatal types of skin cancer, but any of them can kill a patient. In a malignant context, the rapid expansion of cancer cells affects nearby tissues. Furthermore, malignant tumors can divide and spread to other sections of the body via the lymphatic system. Benign tumors, unlike malignant tumors, never metastasize or spread to neighboring tissues. Ignoring malignant tumors eventually leads to metastasis, which culminates in the patient's death. Early detection significantly improves the chances of successful treatment for all skin cancer types [5].

A physician goes through a series of steps to identify skin cancer. First, lesions are examined with the naked eye. Dermoscopy is utilizedto investigate the skin lesion's pattern. In this approach, a gel is applied to the affected areaand examined under a microscope. For a more thorough investigation, a portion of the affected area is removed and undergoes microscopic testing; this procedure is known as a biopsy [6]. Some specialists use the ABCDE technique to diagnose skin lesions, which considers a variety of factors such as colour, border, asymmetry, lesion diameter, and progression over time [7]. However, the diagnosis is entirely dependent on the expertise of skin specialists and the availability of healthcare infrastructure. Cancer, particularly skin cancer, can be prevented from spreading and effectively treated if detected and diagnosed early. Furthermore, early identification of skin cancer reduces mortality and medical costs. Additionally, the manual approach to skin cancer inspection is tedious and prone to errors throughout the diagnostic process[8].

The use of computer-aided systems in medicine has become more common during the past decade. Such technologies can be used to detect skin cancer. Historically, several parameters such as skin color, shape, texture, and so on have been used. Extracting various features is a time-consuming and complicated operation. However, recent advances have founda way for feature extraction using DL architecture. Recently, DL-based computer-aided systems have been used to diagnose a variety of diseases, with excellent results. There is a great opportunity to use computer-aided technology to help medical staff diagnose illnesses in their early stages.Several approaches have recently been developed for the classification of skin cancer using machine learning (ML) and DL frameworks. Table I discusses some of the most recent methods and their limitations for classifying skin images

Table 1: Istribution Ofultrasonic Skin Image Dataset

Ref	Model Used	Insights	Limitations
[9]	Support Vector Machine (SVM)	The features ofLocal Binary Pattern (LBP) plus Histogram of Oriented Gradients (HOG), and Gray-Level Co-occurrence Matrix (GLCM) with the SVM approach are compared to identify the optimal feature extraction technique for skin cancer categorization. The test results demonstrate that GLCM combined with SVM produces higher accuracy than LBP+HOG.	Feature extraction and classification are handled separately and are not fully autonomous. Important feature selection is not discussed. The achieved accuracy is only 74%, likely due to improper feature selection.
[10]	SVM	Pre-processing techniques such as grayscaling, smoothing, and image	The success of the method significantly depends on the quality of

		segmentation are applied. Texture features are retrieved using GLCM, and categorization is performed using SVM.	dermoscopic images. Pre-processing techniques require improvement.
[11]	DenseNet+ Random Forest (RF)	A two-step hierarchical categorization strategy is proposed using DenseNet and RF. It demonstrates superior performance compared to end-to-end DenseNet.	Variations in image quality, lighting conditions, and patient demographics affect performance. There is no comparison with other models.
[12]	Custom CNN	A custom end-to-end CNN is developed from dermoscopic images, utilizing only pixels and disease labels as inputs.	The model achieves a minimum accuracy of 72.73%. Performance needs to be improved through hybrid models.
[13]	Transfer Learning	AlexNet, ResNet-18, SqueezeNet, and ShuffleNet are constructed and evaluated for binary skin cancer classification.	While effective, the selected models may limit exploration. Newer or hybrid methods could yield better outcomes.
[14]	SVM, K-Nearest Neighbour, RF	Watershed segmentation is applied. Features such as shape, the ABCD rule, and GLCM are extracted from the segmented images. Three classifiers are compared, with SVM delivering the best performance.	Feature extraction requires human involvement. The performance of classifiers is entirely dependent on the quality of the extracted features.
[15]	Combine Multiple features	The technique incorporates both high-level features (DenseNet, ResNet, DarkNet) and low-level features (local binary pattern and discrete wavelet transform). Minimum Redundancy Maximum Relevance (mRMR) is used for feature selection. Results indicate that integrating multiple feature types improves classification accuracy.	Although mRMR is employed, the final feature set (2,500 features) remains large. Simplifying the model or adopting more advanced feature selection techniques could further enhance performance.
[16]	Transfer Learning	VGG-16, MobileNet LeNet, and Xceptionare compared for skin cancer detection. Xception delivers the best performance.	The accuracy is still not satisfactory. Performance could be improved by combining the strengths of different models or by applying fine-tuning and optimization techniques.
[17]	Wide Neural Network (WNN)	CNNs such as MobileNet, AlexNet, Inception, and ResNetare used for feature extraction. Gray Wolf Optimizer (GWO) is applied for feature reduction. Six ML classifiers are used for final classification. The AlexNet-GWO-WNN combination demonstrates excellent performance.	DL-based feature extraction combined with GWO-based feature selection increases model complexity, making it unsuitable for deployment on edge devices.

The literature survey on skin cancer reveals several interesting points. Traditional feature extraction methods such as GLCM, LBP, GLRM, and HOG have been used in combination with machine learning techniques to detect skin cancer. However, their accuracy is generally low. This limitation arises because these feature extraction methods are insufficient to capture all the important information from skin images. Additionally, they are time-consuming and heavily dependent on the programmer's expertise. Several studies have shown that utilizing a single deep learning technique is not capable of accurately classifying skin images. In contrast, combining multiple models has proven to

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

achieve better accuracy. Furthermore, hybrid deep learning approaches yield improved performance. However, most existing deep learning models are highly complex, making them unsuitable for real-time deployment.

To address the aforementioned issue, the research proposes a novel attention-based MobileNet-Transformer architecture for skin cancer detection. The key contributions of the research are as follows:

- An attention-based MobileNet is employed for feature extraction. MobileNet captures both local and global features from the input skin images efficiently. The CBAM is used to refine the extracted features by focusing on important regions and emphasizing critical features through a dual attention mechanism. A Transformer block is integrated to extract long-range dependencies between pixels, enabling a deeper understanding of spatial relationships in the image. The final set of features is passed through anFCL for classification. The proposed network effectively captures comprehensive features, resulting in significantly improved classification accuracy for skin cancer detection.
- The proposed network is compared with traditional deep learning models. An ablation study is conducted on the proposed network, and the results are also compared with recent research works from the year 2022. The evaluation is carried out using the standard performance metrics.

The structure of this research is organized as follows: Section I discusses skin cancer, the importance of early detection, and recent developments in skin cancer detection. Section II presents the entire research methodology, from data collection to model development. Section III analyzes the proposed network through various experiments. Section IV concludes the research and outlines its limitations and directions for future work.

2. RESEARCH METHODOLOGY

The research methodology for skin cancer detection is discussed in this section. The study begins with collecting dermatology images from openly accessible sources. The collected images undergo several preprocessing steps to enhance detection accuracy. These preprocessing steps include image resizing, grayscale conversion, hair removal, and data balancing. The processed images are then fed into the proposed novel Deep Learning (DL) model for skin cancer classification. Finally, the proposed network is evaluated against various existing DLnetworks and state-of-the-art research using standard performance metrics. Figure 1 shows the workflow of the proposed methodology, from data collection to model validation.

The proposed Attention-based MobileNet-Transformer consists of three important modules. First, MobileNet is used to extract both local and global features from the processed skin images. The reason for choosing MobileNet is its ability to provide high accuracy with a lightweight architecture. Second, the CBAM is employed. This attention mechanism focuses on important regions using spatial attention, and it further refines the extracted features through channel attention. CBAM helps highlight the most relevant characteristics while suppressing irrelevant ones. Third, the refined features undergo global average pooling to convert them into a sequential format, which acts as the input to the Transformer. The Transformer extracts long-range characteristics, enabling the model to identify the spatial relationships and structural patterns within the image. Finally, the Transformer's output is flattened and passed to an FCL to categorize the skin images as benign or malignant.

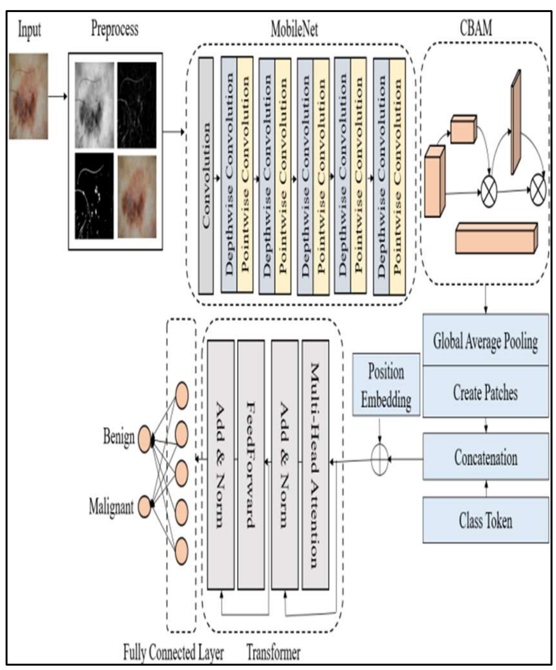
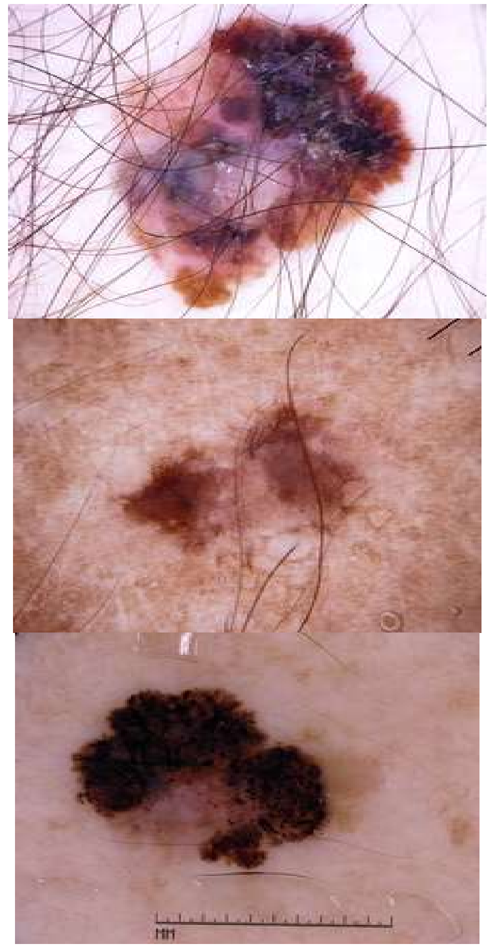


Figure 1:

a. Data Collection and Processing

This research utilizes publicly available datasets from the Kaggle platform [18]. The data is sourced from the ISIC database and includes two kinds of skin cancer images: benign and malignant. This study contains 3,297 skin images, of which 1,800 are benign and 1,497 are malignant. All RGB format images are 224 × 224 pixels. To balance the dataset, geometric augmentation is applied. The final balanced dataset consists of 3,000 data points in each benign and malignant class. The sample images from each category are shown in Figure 2.

a) Benign Images



b) Malignant Images

Figure 2: Sample Skin Cancer Images from the Kaggle Dataset

Resizing images to standard dimensions is an important preprocessing step in DL because it ensures input size consistency, which is required for good model training and inference. Most DL architectures, such as CNNs, require predetermined input dimensions (e.g., 224×224 for ResNet or 299×299 for InceptionV3) to properly evaluate data batches. Without resizing, variable image sizes would cause computational inefficiencies, memory issues, and errors during batch processing. Furthermore, scaling helps to maintain compatibility with pre-trained networks, which are often trained on homogeneous datasets [19].Because artifacts may be the underlying reason for unsatisfactory results, hair removal from the skin images is important [20]. In this research, the Digital Hair Removal (DHR) algorithm is used to remove the hairs [21]. The algorithm has four stages: Gray conversion, BlackHat transformation (BHT), Mask creation, and InPainting.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Gray conversion, as the name implies, converts the skin images to gray color. Grayscale refers to an image composed of varying gray shadesfrom black to white. The pixels in a gray image represents the luminance (brightness) of the corresponding point, typically expressed on a scale from 0 to 255 in 8-bit images.

The BHT approach is used to determine hair forms in terms of direction. Morphological treatments such as opening, closing, erosion, and dilation are used to reduce or enhance specific image regions [22]. The BHT computes the difference between the input image's closing and its original image. BlackHat operators can be mathematically described as shown in Equations (1 and 2):

$$BlackHat\ (A) = (A \odot B) - A \quad (1)$$

$$BlackHat\ (A) = [(A \oplus B) \ominus B] - A \quad (2)$$

Where, Arepresents the input image matrix and B representsthe kernel matrix, $\oplus$ represents dilation, $\ominus$ represents erosion, and $\odot$ represents closing. This change is used to improve image components with larger structural elements and those that are darker than their surroundings.The BlackHat morphological technique provides an image with fluctuations. To improve the contrast of the hair sections, binary thresholding is used as shown in Equation (3):

$$I(x,y) = \begin{cases} 1, if I(x,y) > threshold \\ 0, \quad otherwise \end{cases} \quad (3)$$

The output image highlights the hair contours.After creating the mask to highlight the hair regions, the InPainting step [23] is used to remove them from the original image. Using data from nearby pixels, InPainting fills in undesirable or missing sections of an image. The idea is to recreate the removed region so that it merges with its surroundings and is visually unnoticeable.Let Ω be the region to be inpainted, and $\partial\Omega$ its boundary. A diffusion-based technique is used to fill missing pixel values in Ω . This involves iteratively averaging neighboring pixels from $\partial\Omega$, driven by a diffusion kernel.

$$I(p) = \frac{1}{|N(p)|} \sum_{q \in N(p)} I(q) \quad (4)$$

Where, $I(p)$ represents the intensity at pixel$p \in \Omega$, $N(p)$ represents the neighborhood of p on $\partial\Omega$, and $I(q)$ represents the intensity of neighboring pixel$q$. This technique ensures that the filled region is visually consistent with the surrounding texture and color patterns. As a result of inpainting, the final image is free of hair artifacts, making it a cleaner and more trustworthy input for feature extraction and classification in the skin cancer detection pipeline. Figure 3 depicts the outputs of each stage of the DHR process.

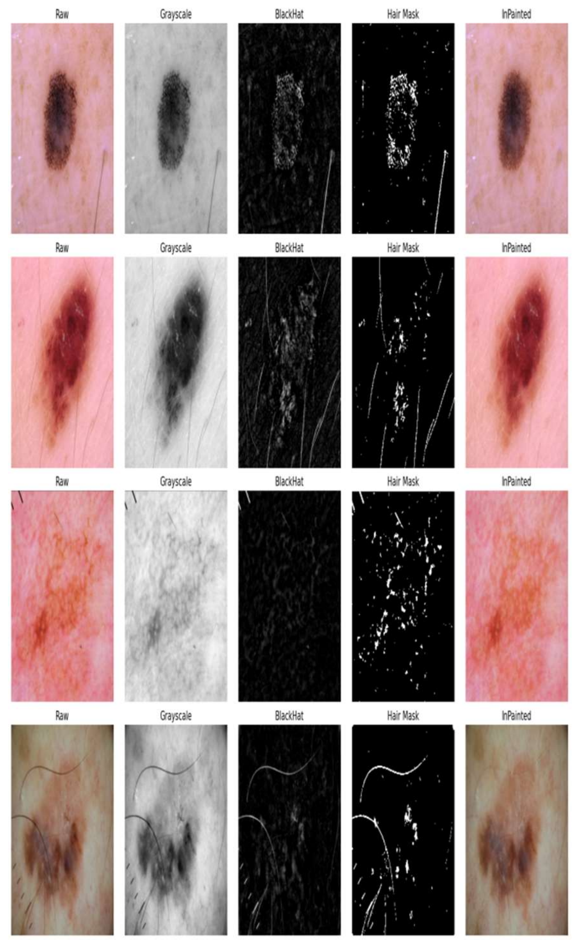


Figure 3: Outcome of the DHR technique on skin images at each stage.

b. Feature Extraction - MobileNet

AlexNet, VGGNet, and GoogLeNet are among the complex networks frequently used in skin cancer classification. However, complex networks impact both network size and processing performance. When using such networks for automatic detection, constraints such as platform computation speed, real-time visual tasks, and other practical limitations must be considered. MobileNet [24] offers an efficient,lightweight architecture, enabling fewer network parameters

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

and faster processing. Figure 4 shows the MobileNet network structure.

The core component of MobileNet is the depthwise separable convolution (DWSC). DWSC is a type of decomposed convolution [25]. The standard convolution method is split into two stages: (1) extracting feature maps and (2) superimposing those extracted maps. The DWSC algorithm separates this into two distinct layers. First, the depthwise convolution (DWC) extracts features from each channel individually. Then, the pointwise convolution (PWC) utilizes a 1×1 convolution to integrate the outcome from the first stage. This decomposition is a simple and effective method for reducing redundant computations and optimizing the network structure.

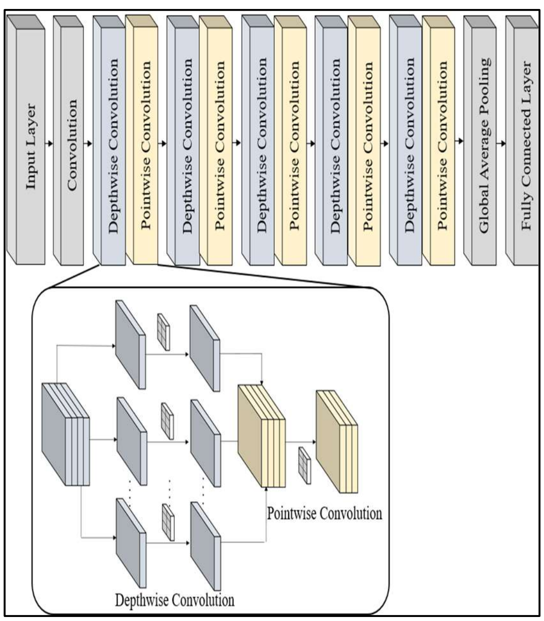


Figure 4: MobileNet Architecture

The dimension of the input feature map is $D_F * D_F * M$, where M represents the total input channels, D_F representsthe feature map dimension, and N representsthe totaloutput channels. The parameters of a conventional convolutional layer are $D_k * D_k * M * N$, where D_k is the dimensionof the convolution kernel. If the dimensionof the feature map's output remains unchanged, the computation cost of conventional convolution (C) becomes substantial and is given in Equation (5).

$$C = D_k * D_k * M * N * D_F * D_F \tag{5}$$

MobileNet employs DWSCs to decouple the relationship between output channels and kernel size, thus reducing redundant computations. The computational cost of DWC (C_1) is shown in Equation (6):

$$C_1 = D_k * D_k * M * D_F * D_F \tag{6}$$

Since DWC filters input channels without generating new features, an additional PWC is required to merge these features and create new multi-channel outputs. Therefore, DWC is faster than standard convolution. The computational cost of PWC (C_2) is given in Equation (7):

$$C_2 = M * N * D_F * D_F \tag{7}$$

The total cost for DWSC is calculated using Equation (8):

$$\frac{C_1+C_2}{C} = \frac{D_k*D_k*M*D_F*D_F+M*N*D_F*D_F}{D_k*D_k*M*N*D_F*D_F} = \frac{1}{N} + \frac{1}{D_k^2} \tag{8}$$

Using a 5×5 convolution kernel, DWSC can be up to 25 times less computationally intensive than standard convolution when N is large. This demonstrates that MobileNet significantly reduces processing cost and model size compared to conventional convolutional neural networks, thereby improving computation speed [26]. MobileNet's transfer learning approach, coupled with subsequent fine-tuning of network parameters, has yielded promising results. The earlier derivations and application cases demonstrate that, for the same feature map size, DWSC substantially reduces both computation and parameter count. MobileNet is a simple yet effective network. Therefore, this study employs MobileNet for feature extraction to enhance the classification accuracy of skin cancer.

c. CBAM

In computer vision, attention mechanisms refer to the capacity to focus on specific areas of an image while disregarding other portions. To interpret complex visual input quickly and effectively, the human brain's visual cortex uses attention methods. Later on, this technique was used in computer vision to improve functionality. Attention methods are dynamic procedures that use variable feature weighting to extract meaningful information from animage. Attention approaches have significantly improved computer vision performance [27]. DL-based attention mechanisms are used in a variety of tasks. They filter a large number of DL samples, removing unnecessary details and selecting information that is more

Journal of Theoretical and Applied Information Technology
15st July 2026. Vol.104. No.13
© Little Lion Scientific

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

useful for the task at hand.

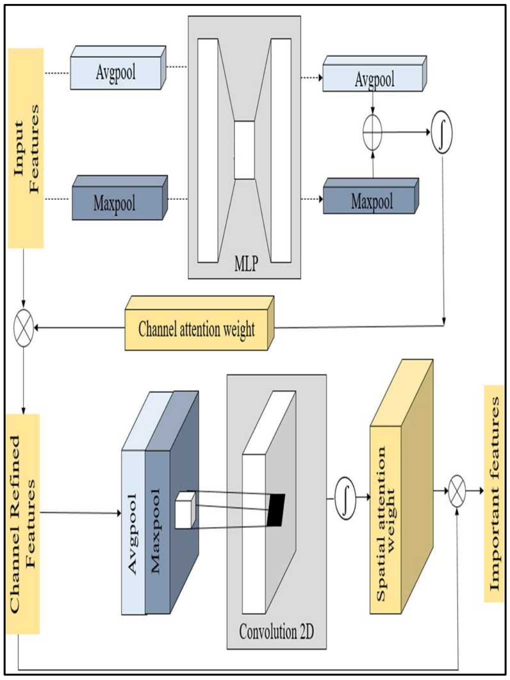


Figure 5: Working of CBAM module

The CBAM is a lightweight attention module used in this work. CBAM is a one-of-a-kind module that combines numerous attention methods, making it simple to integrate into CNN networks [28]. This module expands CNN's feature representation capabilities, making it a simple yet effective update. Furthermore, CBAM is a low-cost solution that can be swiftly integrated into a variety of network configurations while requiring minimal processing resources. The CBAM module can gradually create an attention pattern map inside the channel and spatial boundaries [29]. The final map is then created by multiplying both sets of attention map data with the initial input map, allowing for flexible pattern changes. Figure 5 illustrates the working of the CBAM technique. The approach uses the Channel Attention Module (CAM) to generate the feature map F' and the Spatial Attention Module (SAM) to construct the feature map F'', starting with the intermediate activation map F. Equation (9)is used to describe the computation process.

$$\begin{cases} F' = M_C(F) * F \\ F'' = M_S(F') * F' \end{cases} \tag{9}$$

The symbol ⊗ represents the multiplicative operation within the matched components. The input characteristic map is represented as $F(\in R^{C*H*W})$. The CA represents the resultant weight of F' as $M_C(\in R^{C*1*1})$ and $M_S(\in R^{1*H*W})$ represents the weight F'' obtained using SA. During the initial step of the CAM, a few $C \times 1 \times 1$ feature maps are created using average and max pooling based on width and height. The ultimate CA weights, M_C, are then produced by sending them to the specified Multi-Layer Perceptron (MLP) layer for summing, where they are activated using the sigmoid activation function. Equation (10) can be used to demonstrate the CA computation approach.

$$M_C(F) = \sigma\big[MLP\big(AvgPool(F)\big) + MLP\big(MaxPool(F)\big)\big] \tag{10}$$

MLP refers to a multilayer perceptron, whereas σ represents a sigmoid function. F′ is the feature map used in the SA process. The channel generates $1 \times H \times W$ activation maps through average and max pooling techniques. This feature map is then subjected to the concatenation procedure. The SA weights M_C are created by applying a 7×7 convolution filter to reduce the amount of the map's features. Equation (11) depicts the process of calculating SA.

$$M_S(F) = \sigma\{f^{7*7}[AvgPool(F); MaxPool(F)]\} \tag{11}$$

The convolution method, denoted by f^{7*7}, uses a 7×7convolution filter to gather spatial patterns of the target.

d. Transformer

The Transformer network contains three important elements: a transformer encoder, an embedding layer, and an MLP head [30]. The feature extraction model comprises a convolutional, a max pooling, a residual, and a patch embedding layer. For extracting long-range features, the kernel and the convolutional stride in patch embedding were merged and used to shift the channel. Before entering the transformer encoder, ensure that it contains both the position embedding and the class token. Both are trainable parameters. The transformer encoder elements are made up of encoder blocks, and it has three main components: multi-head attention (MHA), layer norm, and an MLP block [31].AnMLP is made up of three layers: FCL, dropout, and GELU. The architecture of the transformer network is given in Figure 6.

Unlike classical attention, the chosen MHA feature maps a key-value (K, Q) pairings and queries (V) to output vectors, with the three learnable matrices having dimensions d_k, d_q, and

d_v, respectively. To obtain the values of Q, K, and V, multiply R by three learned matrices, $P_{QKV} \in \mathbb{R}^{D*3d_k}$ as shown in Equations (12)

$$[Q, K, V] = RP_{QKV} \tag{12}$$

The MHA feature, which is a core module of the transformer, allows the network to respond to input from several representation subspaces at variouspositions[32]. Before running the scaled dot-product attention computation, it executes three linear projections to convert the Q_S, K_S, and V_S to more selective representations. The MHA's final outputs are then obtained by concatenating each head and passing it through another linear projection. To create attention weights, dot products are produced, and the Q_S with all K_S are linked with a scaling factor $\left(1/(d_k)^{1/2}\right)$. The softmax function is then applied to the V_S. Thus, the generated attention weights that denote the key component of image patches are defined as shown in Equation (13):

$$Attention(V, K, Q) = Softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{13}$$

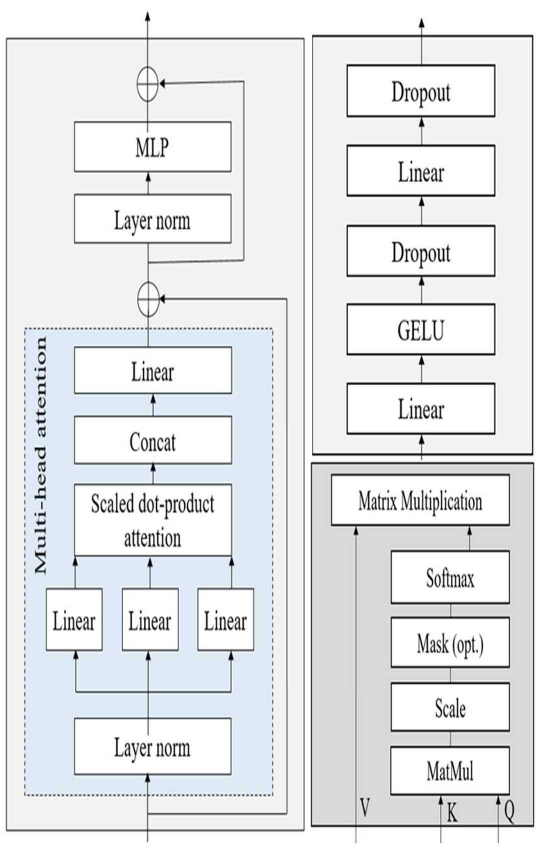


Figure 6 : Structure of Transformer's Encoder

To minimize the sequence computation, map the Q_S, K_S, and V_S, h times using distinct linear projections before running Equation (13) simultaneously. The resulting h vectors are then integrated using the concatenation function, with each vector denoted as a head.After concatenation, the output of the MHA mechanism is derived using Equation (14)

$$MHA(V, K, Q) = Concat(head_1, \dots, head_h)W^o \tag{14}$$

Where,Concat $(\cdot)$represents the concatenation function, $head_i = Attention\left(QW_i^Q, KW_i^K, VW_i^V\right)$, and $W^o \in \mathbb{R}^{h*d_v*D}$represents learnable weights for a feedforward layer. When the MHA module's resultsare determined, the MLP module processes the resulting data. The MLP block combines dual FCL and the GELU to generate the outcome of the transformer encoder's single layer. The formula for the GELU is shown in Equation (15):

$$GELU(x) = x\phi(x) = x.\frac{1}{2}\left[1 + erf\left(x/\sqrt{2}\right)\right] \tag{15}$$

Where, $\phi(x)$ represents the normal Gaussian cumulative distribution. Additionally, the layer norm in the transformer is determined using Equations (16, 17):

$$\mu^f = \frac{1}{H}\sum_{i=1}^{H} a_i^f \tag{16}$$

$$\sigma^f = \sqrt{\frac{1}{H}\sum_{i=1}^{H}\left(a_i^f - \mu_f\right)^2} \tag{17}$$

Where, H represents the total hidden units, a_i^f represents the aggregated input of the f -th layer'si-th hidden unit. All hidden units within a layer share the same normalization parameters μ and σ. The normalized activation$\hat{a}^f$ can then be created using Equation (18):

$$\hat{a}^f = \frac{a^f - \mu_f}{\sqrt{\left(\sigma^f\right)^2 + \varepsilon}} \tag{18}$$

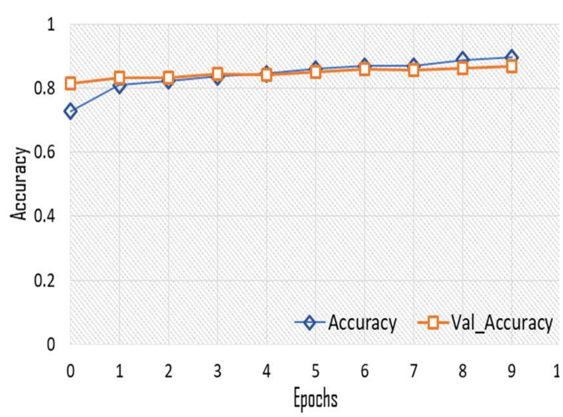
The constant ε provides numerical stability.

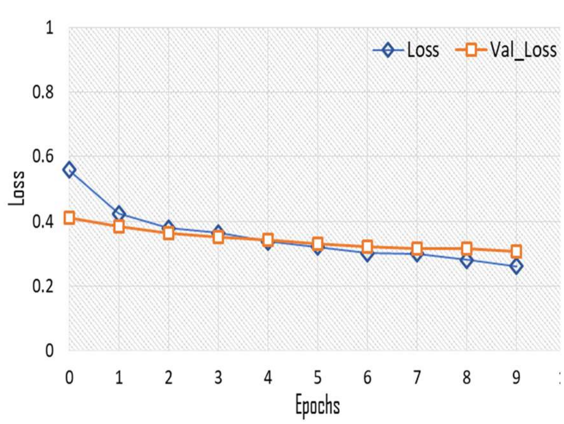
3. RESULTS AND DISCUSSION

The skin cancer classification using a novel Attention-based MobileNet-Transformer network was done in the Google Colab environment. The TPU (8 cores) was selected as the hardware accelerator and used to train and test the DL model efficiency. The proposed network is compared with MobileNet, DenseNet, and Xception models. For a fair comparison, all models were trained on 5,400 images and tested with 600 images. The other parameters, such as the number of epochs, were set to 10, with the categorical cross-entropy function, Adam optimizer, and the 0.0001 learning rate.

First, all three DL models and the proposed network were evaluated in the training and validation phases using accuracy and loss functions. The performance plots of MobileNet, Xception, DenseNet, and the proposed network are shown in Figures 7-10. By analyzing the figures, it can be observed that the accuracy plot of all three conventional DL networks reaches a maximum of 0.85–0.89 for training and 0.84–0.87 for validation. The minimum loss value for both training and validation is in the range of 0.3–0.4. However, in the proposed network, the maximum training and validation accuracy values are 0.96 and 0.95, respectively, and the minimum loss values are 0.2 and 0.23. The comparison of outcomes in the training and validation phases indicates that the proposed network outperforms the conventional DL models.

a) Accuracy plot

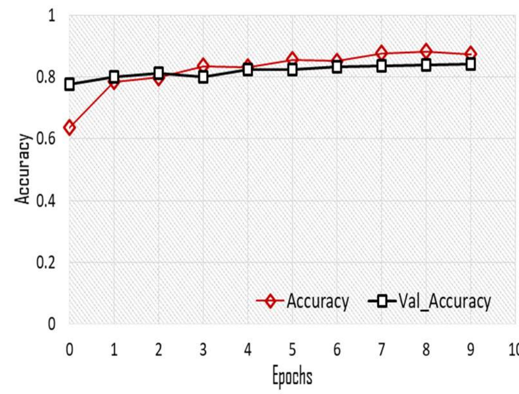


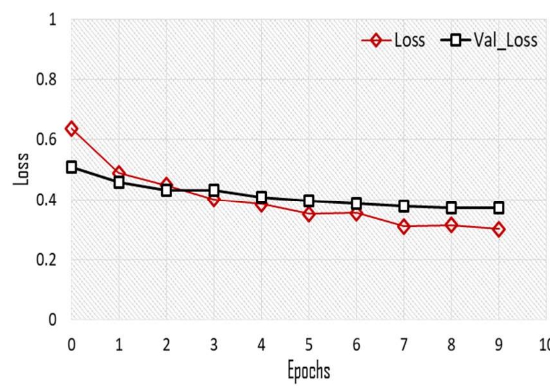


b) Loss plot

Figure 7 : Performance plot of MobileNet

a) Accuracy plot

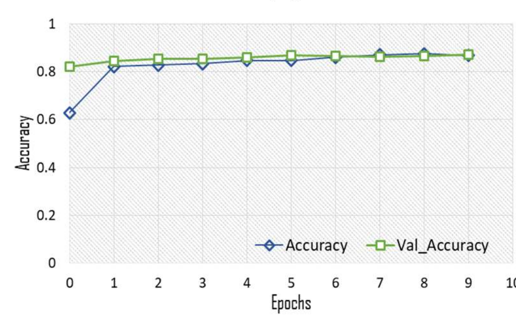


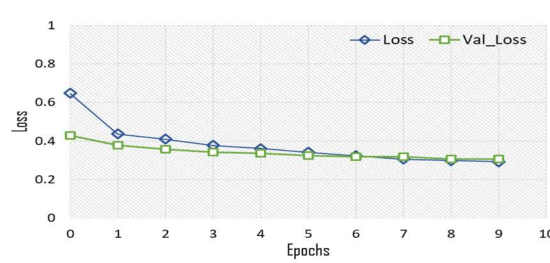


b) Loss plot

Figure 8: Performance plot of Xception

a) Accuracy plot

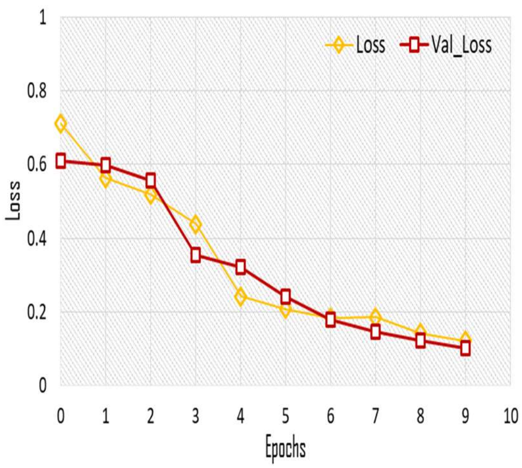


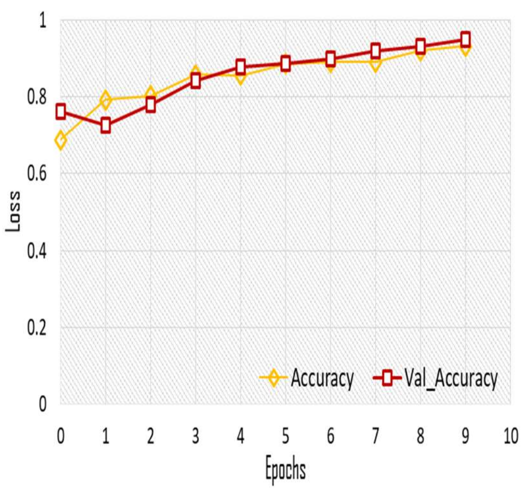


b) Loss plot

Figure 9: Performance plot of DenseNet

a) Accuracyplot





b) Loss plot

Figure 10: Performance plot of Attention-based MobileNet-Transformer

After training and validating the networks, the models were tested on 300 benign and 300 malignant images. The confusion matrix elements (True Positive-TP, True Negative-TN, False Positive-FP, False Negative-FN) and performance metrics for the DL networks are provided in Table II. The MobileNet model correctly identified 268 malignant and 267 benign images but incorrectly predicted 65 images. Using these values, the performance metrics were calculated. MobileNet achieved an accuracy of 89.17%, with precision, recall, and F1 scores of 89%. Next, the Xception model correctly classified 260 malignant and 245 benign images, but incorrectly detected 55 as malignant and 40 as benign. The performance metrics for Xception are: 84.17% accuracy, 83% precision, 87% recall, and 85% F1-score. DenseNet correctly identified 526 images and wrongly predicted 74. The accuracy of DenseNet was 87.67%. Finally, the proposed network correctly identified 286 benign images from 300 and 278 malignant images from 300 samples, incorrectly identifying only 9 samples. The proposed network achieved the highest accuracy of 94%, with precision, recall, and F1 scores of 93%, 95%, and 94%, respectively.

Table 2: Performance Analysis of Proposed Network with Conventional DL Networks

DL Model	TP	TN	FP	FN	Accuracy	Precision	Recall	F1-Score
MobileNet	268	267	33	32	89.17	89	89	89
Xception	260	245	55	40	84.17	83	87	85
DenseNet	253	273	27	47	87.67	90	84	87
Proposed Network	286	278	22	14	94	93	95	94

Next, an ablation reseacrh was done on the proposed network to analyze the importance of each module in the network. First, the attention mechanism was removed, and the MobileNet features were directly passed to the transformer, referred to as Ablation Study 1. Next, the transformer was removed, and the refined features were given to the FCL for classification, referred to as Ablation Study 2. Lastly, only the transformer was used for feature extraction and classification, with both MobileNet and CBAM removed, referred to as Ablation Study 3. The confusion matrix elements (TP, TN, FP, FN) for each ablation study are given in Table III. Using the confusion matrix, the performance metrics are calculated and shown in Figure 11. The figure displays the performance analysis of the ablation studies and the proposed network. It clearly shows that the removal or modification of any module in the network reduces the accuracy of the skin cancer detection system, emphasizing the importance of each feature extracted by MobileNet, CBAM, and the transformer for better accuracy.

Table 3:Confusion Matrix of the Ablation Studies

Ablation Study	TP	TN	FP	FN
AS-1	274	269	28	29
AS-2	271	266	31	32
AS-3	268	259	34	39
Proposed Network	286	278	22	14

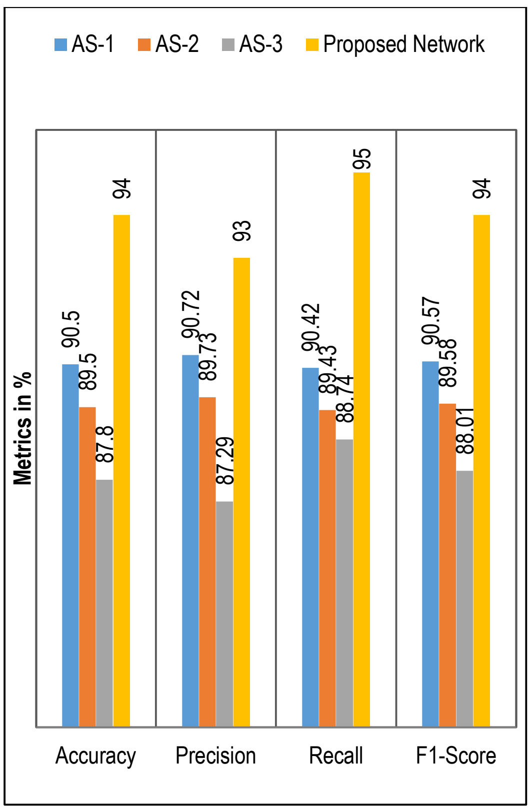


Figure 11: Performance Analysis of Proposed Network Modules Through Ablation Experiments

Finally, the proposed network was compared with recent research works, as shown in Table IV. The table mentions the research, models used, and the metrics attained. The research papers were selected using keywords such as "skin cancer," "Kaggle dataset," and "binary classification." Around 15 research papers from the years 2022 to 2025 were taken for comparison. All the research used the same dataset of skin cancer as in this work. In the comparison, only a few models proposed in research [36, 38, 39, 41, 42, 43] achieved an accuracy greater than 90%. Among these, the highest accuracy was attained by ResNet [36], with an accuracy of 93.5%, while the proposed network achieved an accuracy of 4%.

Table 4: Summary of Classification Performance: Proposed Model vs. Existing Research

Ref	Model	Accuracy	Recall	Precision	F1-Score
[33]	Fine-tuning ResNet50	87	89	84	87
[34]	Modified VGG16	89.09	-	-	-
[35]	ResNet	86.57	86	86.48	86.23
[36]	ResNet	93.5	77	94	86
[37]	VGG-16	84.242	-	-	-
[38]	Ensemble ML Model	92	91	92	92
[39]	InceptionV3	91.26	92	89	92
[40]	Fine-tuned DenseNet	87	87	86	86
[41]	Fusion of Inception and DenseNet	92.27	92.33	90.80	91.57
[42]	CNN	91.67	91.5	91.5	91
[43]	CNN	92	92	92	92
[44]	Compact-DCNN Model	84.8	-	80.5	81.6
[45]	Fine-tuned EfficientNetB3	87.12	87	87	87
[46]	RESNET101 with Logistic Regression	87.89	87.29	88.56	87.92
Proposed Network	Attention-based MobileNet - Transformer	94	95	93	94

4. CONCLUSION

An effective skin cancer classification network is proposed in this research. The dermatological images from Kaggle are collected. The collected images are imbalanced; a geometrical augmentation method is used to solve this. To improve the DL network's accuracy, hair from the images is removed using the DHR technique. The processed images are fed into the proposed network and three other conventional techniques—MobileNet, Xception, and DenseNet—for evaluation. The test outcome shows that the proposed network achieves an excellent accuracy of 94%, while the other models achieve 89.17%, 84.17%, and 87.67%, respectively. Ablation research is done on the proposed network to justify the importance of each module. Even minimal changes to the proposed network result in a drop in accuracy for skin cancer detection. Finally, the proposed network is compared with around 15 existing research works. The proposed network outperforms them and stands as the best. This research provides a promising solution for reliable skin cancer detection using a lightweight, novel DL model. It will be helpful for physicians to detect skin cancer in individuals with high accuracy and minimal time.

The limitation of the research is that the proposed network is evaluated using only one dataset, raising concerns about generalizability. To make the model real-time and robust, it should be tested across various datasets under different environmental conditions. In the future, the proposed network will be tested against several other public datasets and real-time datasets. Additionally, the proposed network will be deployed on FPGA hardware and analyzed using several metrics such as power consumption and inference time.

COMPETING INTERESTS
The authors declare that there are no conflicts of interest regarding the publication of this manuscript.

FUNDING
The authors declare that no funds, grants or any other support were received during the preparation of this manuscript

AUTHOR CONTRIBUTION
All authors contributed equally.

DATA AVAILABILITY STATEMENT
Not applicable.

RESEARCH INVOLVING HUMAN AND /OR ANIMALS
No experimentation involving humans or animals was conducted.

INFORMED CONSENT
Not applicable.

REFERENCES

[1] Duarte, Ana Filipa, Bernardo Sousa-Pinto, Eckart Haneke, and Osvaldo Correia. "Risk factors for development of new skin neoplasms in patients with past history of skin cancer: a survival analysis." *Scientific reports* 8, no. 1 (2018): 15744.

[2] Roky, Amdad Hossain, Mohammed Murshedul Islam, Abu Mohammed Fuad Ahasan, Md Saqline Mostaq, Md Zihad Mahmud, Mohammad Nurul Amin, and Md Ashiq Mahmud. "Overview of skin cancer types and prevalence rates across continents." *Cancer pathogenesis and therapy* 3, no. 02 (2025): 89-100.

[3] Kowalski, Szymon, Julia Karska, Maciej Tota, Katarzyna Skinderowicz, Julita Kulbacka, and Małgorzata Drąg-Zalesińska. "Natural compounds in non-melanoma skin cancer: Prevention and treatment." *Molecules* 29, no. 3 (2024): 728.

[4] Siegel, Rebecca L., Kimberly D. Miller, Hannah E. Fuchs, and Ahmedin Jemal. "Cancer statistics, 2022." *CA: a cancer journal for clinicians* 72, no. 1 (2022): 7-33.

[5] Carter, Emily. "Identifying types of skin cancer, risk factors, and effective treatments." *International Journal of Advanced Engineering Technologies and Innovations* 10, no. 2 (2024): 79-98.

[6] Greenwood, Jason D., Stephen P. Merry, and Christopher L. Boswell. "Skin biopsy techniques." *Primary Care: Clinics in Office Practice* 49, no. 1 (2022): 1-22.

[7] Alsararatee, Hasan H. "Dermatological assessment: history-taking and examination." *British Journal of Nursing* 33, no. 21 (2024): 1026-1036.

[8] Khattar, Sonam, and Ravinder Kaur. "Computer assisted diagnosis of skin cancer: A survey and future recommendations." *Computers and Electrical Engineering* 104 (2022): 108431.

[9] Khoirunisa, Friska Eka, and Novrido Charibaldi. "Effect of Using GLCM and LBP+ HOG Feature Extraction on SVM Method in Classification of Human Skin Disease Type." *Inform: Jurnal Ilmiah Bidang Teknologi Informasi dan Komunikasi* 9, no. 2 (2024): 145-150.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[10] Hasanah, Riyan Latifahul, and Dwiza Riana. "Classification of dermoscopic image of skin cancer using the GLCM method and multi-SVM algorithm." *Rekayasa* 14, no. 3 (2021): 407-415.

[11] Suleiman, Taofik Ahmed, Daniel Tweneboah Anyimadu, Andrew Dwi Permana, Hsham Abdalgny Abdalwhab Ngim, and Alessandra Scotto di Freca. "Two-step hierarchical binary classification of cancerous skin lesions using transfer learning and the random forest algorithm." *Visual Computing for Industry, Biomedicine, and Art* 7, no. 1 (2024): 15.

[12] Oumoulylte, Mariame, Ahmad El Allaoui, Yousef Farhaoui, Fatima Amounas, and Youssef Qaraai. "Deep Learning Algorithms for Skin Cancer Classification." In *The International Conference on Artificial Intelligence and Smart Environment*, pp. 345-351. Cham: Springer International Publishing, 2022.

[13] Hussein, Hadeer, Ahmed Magdy, Rehab F. Abdel-Kader, and Khaled Abd Elsalam Ali. "Binary classification of skin cancer using pretrained deep neural networks." *Suez Canal Engineering, Energy and Environmental Science* 1, no. 2 (2023): 10-14.

[14] Murugan, A., S. Anu H. Nair, and KP Sanal Kumar. "Detection of skin cancer using SVM, random forest and kNN classifiers." *Journal of medical systems* 43, no. 8 (2019): 269.

[15] Attallah, Omneya, and Maha Sharkas. "Intelligent dermatologist tool for classifying multiple skin cancer subtypes by incorporating manifold radiomics features categories." *Contrast media & molecular imaging* 2021, no. 1 (2021): 7192016.

[16] Jatmiko, Rizky Hafizh, and Yoga Pristyanto. "Investigating The Effectiveness of Various Convolutional Neural Network Model Architectures for Skin Cancer Melanoma Classification." *MATRIK: Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer* 23, no. 1 (2023): 1-16.

[17] Saleh, Neven, Mohammed A. Hassan, and Ahmed M. Salaheldin. "Skin cancer classification based on an optimized convolutional neural network and multicriteria decision-making." *Scientific Reports* 14, no. 1 (2024): 17323.

[18] Skin Cancer: Malignant vs. Benign dataset from Kaggle. Available online: **https://www.kaggle.com/fanconic/skin-cancer-malignant-vs-benign** (accessed on 1 May 2025).

[19] Talebi, Hossein, and Peyman Milanfar. "Learning to resize images for computer vision tasks." In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 497-506. 2021.

[20] Salido, Julie Ann, and Conrado Ruiz. "Hair artifact removal and skin lesion segmentation of dermoscopy images." *Asian Journal of Pharmaceutical and Clinical Research* 11, no. 3 (2018).

[21] Li, Wei, Alex Noel Joseph Raj, Tardi Tjahjadi, and Zhemin Zhuang. "Digital hair removal by deep learning for skin lesion segmentation." *Pattern Recognition* 117 (2021): 107994.

[22] Said, Khairul Anuar Mat, and Asral Bahari Jambek. "Analysis of image processing using morphological erosion and dilation." In *Journal of Physics: Conference Series*, vol. 2071, no. 1, p. 012033. IOP Publishing, 2021.

[23] Guillemot, Christine, and Olivier Le Meur. "Image inpainting: Overview and recent advances." *IEEE signal processing magazine* 31, no. 1 (2013): 127-144.

[24] Howard, Andrew G. "Mobilenets: Efficient convolutional neural networks for mobile vision applications." *arXiv preprint arXiv:1704.04861* (2017).

[25] Hsiao, Shen-Fu, and Bo-Ching Tsai. "Efficient computation of depthwise separable convolution in MoblieNet deep neural network models." In *2021 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW)*, pp. 1-2. IEEE, 2021.

[26] Sinha, Debjyoti, and Mohamed El-Sharkawy. "Thin mobilenet: An enhanced mobilenet architecture." In *2019 IEEE 10th annual ubiquitous computing, electronics & mobile communication conference (UEMCON)*, pp. 0280-0285. IEEE, 2019.

[27] Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. "Attention is all you need." *Advances in neural information processing systems* 30 (2017).

[28] Woo, Sanghyun, Jongchan Park, Joon-Young Lee, and In So Kweon. "Cbam: Convolutional block attention module." In *Proceedings of the European conference on computer vision (ECCV)*, pp. 3-19. 2018.

[29] Wang, Yifan, Wu Wang, Yang Li, Yaodong Jia, Yu Xu, Yu Ling, and Jiaqi Ma. "An

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

attention mechanism module with spatial perception and channel information interaction." *Complex & Intelligent Systems* 10, no. 4 (2024): 5427-5444.

[30] Khan, Salman, Muzammal Naseer, Munawar Hayat, Syed Waqas Zamir, Fahad Shahbaz Khan, and Mubarak Shah. "Transformers in vision: A survey." *ACM computing surveys (CSUR)* 54, no. 10s (2022): 1-41.

[31] Lu, Siyuan, Meiqi Wang, Shuang Liang, Jun Lin, and Zhongfeng Wang. "Hardware accelerator for multi-head attention and position-wise feed-forward in the transformer." In *2020 IEEE 33rd International System-on-Chip Conference (SOCC)*, pp. 84-89. IEEE, 2020.

[32] Naz, Hina, Zuping Zhang, Mohammed Al-Habib, Fuad A. Awwad, Emad AA Ismail, and Zaid Ali Khan. "Posture Detection of Heart Disease UsingMulti-Head Attention Vision Hybrid (MHAVH) Model." *Computers, Materials & Continua* 79, no. 2 (2024).

[33] Bechelli, Solene, and Jerome Delhommelle. "Machine learning and deep learning algorithms for skin cancer classification from dermoscopic images." *Bioengineering* 9, no. 3 (2022): 97.

[34] Anand, Vatsala, Sheifali Gupta, Ayman Altameem, Soumya Ranjan Nayak, Ramesh Chandra Poonia, and Abdul Khader Jilani Saudagar. "An enhanced transfer learning based classification for diagnosis of skin cancer." *Diagnostics* 12, no. 7 (2022): 1628.

[35] Agarwal, Kartikeya, and Tismeet Singh. "Classification of skin cancer images using convolutional neural networks." *arXiv preprint arXiv:2202.00678* (2022).

[36] Islam, Md Sirajul, and Sanjeev Panta. "Skin Cancer Images Classification using Transfer Learning Techniques." *arXiv preprint arXiv:2406.12954* (2024).

[37] Ibrahim, Adam M., Mohammed Elbasheir, Somia Badawi, Ashraf Mohammed, and Amir F. Mohammed Alalmin. "Skin cancer classification using transfer learning by VGG16 architecture (case study on Kaggle dataset)." *Journal of Intelligent Learning Systems and Applications* 15, no. 3 (2023): 67-75.

[38] Alfi, Iftiaz A., Md Mahfuzur Rahman, Mohammad Shorfuzzaman, and Amril Nazir. "A non-invasive interpretable diagnosis of melanoma skin cancer using deep learning and ensemble stacking of machine learning models." *Diagnostics* 12, no. 3 (2022): 726.

[39] Lembhe, Ashutosh, Pranav Motarwar, Rudra Patil, and Susan Elias. "Enhancement in skin cancer detection using image super resolution and convolutional neural network." *Procedia Computer Science* 218 (2023): 164-173.

[40] Bello, Abayomi, Sin-Chun Ng, and Man-Fai Leung. "Skin cancer classification using fine-tuned transfer learning of DENSENET-121." *Applied Sciences* 14, no. 17 (2024): 7707.

[41] Akter, Maksuda, Rabea Khatun, Md Alamin Talukder, Md Manowarul Islam, Md Ashraf Uddin, Md Khabir Uddin Ahamed, and Ansam Khraisat. "An Integrated Deep Learning Model for Skin Cancer Detection Using Hybrid Feature Fusion Technique." *Biomedical Materials & Devices* (2025): 1-15.

[42] Pillai, Rudresh, Neha Sharma, Sheifali Gupta, Deepali Gupta, Sapna Juneja, Saurav Malik, Hong Qin, Mohammed S. Alqahtani, and Amel Ksibi. "Enhanced skin cancer diagnosis through grid search algorithm-optimized deep learning models for skin lesion analysis." *Frontiers in Medicine* 11 (2024): 1436470.

[43] Wang, Xikun. "Deep learning-based and machine learning-based application in skin cancer image classification." In *Journal of Physics: Conference Series*, vol. 2405, no. 1, p. 012024. IOP Publishing, 2022.

[44] Fofanah, Alhaji Balla, Emre Özbilge, and Yonal Kırsal. "Skin cancer recognition using compact deep convolutional neural network." *Çukurova Üniversitesi Mühendislik Fakültesi Dergisi* 38, no. 3 (2023): 787-797.

[45] Salian, Saumya R., and Sudhir D. Sawarkar. "Melanoma skin lesion classification using improved efficientnetb3." *Jordanian Journal of Computers and Information Technology* 8, no. 1 (2022).

[46] Qayyum, Hafza, Syed Tahir Hussain Rizvi, Muddasar Naeem, Umamah bint Khalid, Musarat Abbas, and Antonio Coronato. "Enhancing Diagnostic Accuracy for Skin Cancer and COVID-19 Detection: A Comparative Study Using a Stacked Ensemble Method." *Technologies* 12, no. 9 (2024): 142.