FMDB *Transactions on Sustainable Computing Systems*

MFETok-MLP Based Lesion Segmentation and Marine Predictor Based FRNN for Classifying the Dermoscopic Image

V. Asha[1,*], K. Aravinda[2], M. Arunadevi Thirumalraj[3], S. Gopikha[4], Amarilys González García[5]

[1]Department of Computer Application, New Horizon College of Engineering, Bengaluru, Karnataka, India.
[2]Department of Electronics and Communication Engineering, New Horizon College of Engineering, Bengaluru, Karnataka, India.
[3]Department of Computer Science and Engineering, Karunya Institute of Technology and Science, Coimbatore, Tamil Nadu, India.
[3]Department of Computer Science and Business Management, Saranathan College of Engineering, Tiruchirappalli, Tamil Nadu, India.
[4]Department of Information Technology, St. Joseph's College of Engineering, Chennai, Tamil Nadu, India.
[5]Department of Research and Development, Placental Histotherapy Center, Havana, Cuba.
asha.gurudath@gmail.com[1], aravindake@gmail.com[2], aruna.devi96@gmail.com[3], gopikha.in@gmail.com[4], agonzalez@miyares-cao.cu[5]

Abstract: Millions of individuals worldwide have deadly skin cancer. Prognosis and treatment improve with early identification and correct diagnosis. CNNs have improved medical image processing. This study presents a CNN-based method for skin cancer detection. This study employed two publicly available benchmark datasets: ISIC 2020 and HAM10000. The recommended method involves feature classification, segmentation, picture preprocessing, and hyperparameter tweaking for classification. Image brightness data was used to create a unique contrast enhancement method. UNet architecture is segmented next. The feature classification model used the Tok-MLP with dynamic sparse attention (MFETok-MLP) and Multi-scale Feature Extraction modules. This classification improved the model's ACC and training resilience. The Fast Recurrent Neural Network classified lesion images. The MPOA fixed the FRNN classifier hyperparameters to develop an excellent skin cancer prediction ACC. The suggested skin cancer detection model achieved excellent ACCs of 99.78% and 99.81% on the ISIC-2020 and HAM10000 datasets, respectively. These data demonstrate that the proposed method is more effective than current methods, helping dermatologists and other medical professionals detect skin cancer.

Keywords: Skin Cancer; Contrast Enhancement; Tokenised Multilayer Perceptron (Tok-MLP); Fast Recurrent Neural Network; Marine Predators' Optimisation Algorithm (MPOA); Convolutional neural networks (CNNs).

Received on: 14/12/2024, **Revised on:** 09/03/2025**, Accepted on:** 24/04/2025**, Published on:** 09/12/2025

Journal Homepage: https://www.fmdbpub.com/user/journals/details/FTSCS

DOI: https://doi.org/10.69888/FTSCS.2025.000524

Cite as: V. Asha, K. Aravinda, M. A. Thirumalraj, S. Gopikha, and A. G. García “MFETok-MLP Based Lesion Segmentation and Marine Predictor Based FRNN for Classifying the Dermoscopic Image,” *FMDB Transactions on Sustainable Computing Systems*, vol. 3, no. 4, pp. 215–232, 2025.

Copyright © 2025 V. Asha *et al.*, licensed to Fernando Martins De Bulhão (FMDB) Publishing Company. This is an open access article distributed under CC BY-NC-SA 4.0, which allows unlimited use, distribution, and reproduction in any medium with proper attribution.

1. Introduction

---

[*] Corresponding author.

The largest organ in the body, the skin, is essential for regulating body temperature, shielding internal organs, and providing a sensory link to the outside world [1]. Because Langerhans cells are present, they also help the immune system function by assisting in detecting and defending against infections [2]. Uncontrolled cell growth is the hallmark of skin cancer, which is a frequent and occasionally fatal illness [3]. It is most frequently found in areas that receive UV radiation from the sun, whether through tanning facilities or direct sunlight. There are various forms of skin cancer, the most prevalent of which are squamous-cell carcinoma, basal cell and melanoma. Of these, basal cell carcinoma is the most treatable and least aggressive [4]; [5]. Squamous cell carcinoma, on the other hand, is more aggressive and has the potential to spread if left untreated. Because melanoma has a strong propensity to spread, it is the deadliest kind of skin cancer [6]. For skin cancer to be properly managed, early identification and treatment are essential [7]. Changes in the size, form, or colour of moles, as well as the emergence of new, odd growths, are warning indicators [8]. Identifying suspicious moles is made easier by the ABCDE criteria, which stand for Asymmetry, Border irregularity, Colour change, Diameter greater than 6 mm, and evolving [9].

CNNs and deep learning (DL) have revolutionised skin cancer diagnostics thanks to advances in medical technology [10]. CNNs are becoming more and more used for the diagnosis of skin cancer because of their superior interpretation of medical images, such as histology slides and dermoscopic images [11]. They can identify subtle characteristics and patterns in skin lesions that may indicate cancer. To eliminate diagnostic subjectivity, CNNs are very useful for classifying skin lesions such as melanoma, basal cell carcinoma, and squamous cell carcinoma [23]. They do this by offering an objective and consistent assessment [12]. It's crucial to remember that although CNNs and DL have enormous potential for detecting skin cancer, they shouldn't take the position of medical experts. Rather, they are useful instruments to improve the ACC and efficiency of diagnosis [13]. CNNs can help dermatologists diagnose patients more quickly and accurately, reduce the risk of misdiagnosis, and provide a crucial second opinion. The use of CNNs for the diagnosis of skin cancer is consistent with the larger trend of AI in healthcare, which promises to revolutionise medical procedures [14]. This work's main contributions are:

- Presents a novel feature selection method for skin lesion detection using the MFETok-MLP-based UNet.
- It offers a new integrated approach to FRNN classification, with hyperparameter tuning via MPOA, greatly improving effectiveness and ACC in skin cancer diagnosis.
- Results are evaluated using the ISIC-2020 and HAM10000 datasets, with performance metrics including Accuracy (ACC), Precision (PR), Recall (RC), F-measure (F1), Dice Similarity Coefficient (DSC), and Jaccard Similarity Coefficient (JSC).

2. Related Works

DSCC_Net, a novel CNN-based DL framework for the detection of skin cancers, was presented by Tahir et al. [15]. To fully evaluate this approach, three popular benchmark datasets (DermIS, HAM10000, and ISIC 2020) were chosen. With an impressive area under the curve (AUC) of 99.43%, a high ACC of 94.17%, an RC rate of 93.76%, a remarkable PR of 94.28%, and an F1-score of 93.93%, DSCC_Net did very well in discriminating between the four forms of skin cancer. The most often used models were Inception-V3, ResNet-152, Vgg-19, MobileNet, Vgg-16, EfficientNet-B0, and Vgg-19, with corresponding ACC ratings of 91.82%, 92.51%, 91.46%, 89.12%, and 89.46%. The findings show that the proposed DSCC_Net model outperforms the current baseline models. The results of this research should help dermatologists and other medical professionals recognise skin cancer. Ichim et al. [16] developed two distinct models to help identify different skin abnormalities, including potentially lethal melanoma. These models generated group judgements by combining individual assessments with a range of neural network development strategies. Among the systems, one distinguished itself as a creative solution with an amazing 91.04% ACC rate. Numerous intermediate binary classification subsystems were present. This outperformed the traditional approach, which relied on the weighted voting of component networks. Furthermore, the study showed that the effectiveness of these models varied depending on the kind of skin lesion. The F1 score, for instance, ranged from 81.4% to 94.1%. A unique CNN model based on transfer learning (TL) was developed by Baig et al. [17] to distinguish among seven forms of pigmented skin lesions (PSLs). To improve processing performance, a lightweight CNN model and channel-wise attention (CA) are combined in a novel method dubbed Light-Dermo. Squeeze-and-Excitation (SE) blocks were used to enhance the ShuffleNet backbone, thereby improving its fundamental architecture.

To assess the Light-Dermo model, a dataset of 14,000 PSL images from each of the seven groups was used. Data augmentation methods were applied to each of the seven PSL classes to address imbalances and increase the dataset size. As a result, 28,000 images were gathered from ISIC-2020, HAM10000, and ISIS-2019 databases. The analysis demonstrated that the recommended approach often outperformed the most advanced approaches. The best-trained model has a 98.1% F1, 98.20% SP, 97.45% ST, and 99.14% ACC. Remarkably, compared to the most intricate DL models, the model required fewer parameters. These findings may improve the accuracy with which physicians use Light-Dermo to detect pigmented skin lesions. Abbas et al. [18] used SqueezeNet with depthwise separable CNN models in a unique manner to produce a Skin Lesion Classification System. The primary objective of this paper was to create a low-parameter deep learning architecture with performance close to state-of-the-art (SOTA) models. Depthwise separable convolutions fundamentally alter the SqueezeNet architecture, improving runtime performance over regular convolutional units. The researchers used a data augmentation strategy to improve the Assist-Dermo system and address the imbalance in Positive Skin Lesion (PSL) samples. Using a colour scheme designed for human eyes, the pre-processing included identifying and highlighting the main areas of skin lesions. By adding more layers (lrs) and adjusting filter sizes, the Assist-Dermo system's efficacy and performance were significantly increased with fewer filters and modifications. Searching across several internet sources, such as Ph2, ISBI-2017, HAM10000, and ISIC, yielded a sizable collection of PSL images suitable for teaching and evaluation. With an area under the curve (AUC) of 0.95, ACC of 95.6%, ST of 96.7%, SP of 95%, and other good performance metrics, this dataset demonstrated the model's effectiveness. The test results show that the suggested Assist-Dermo technique outperformed the most complex algorithms in PSL identification across nine distinct classes. Dermatologists may discover better ways to use dermoscopy to distinguish among various types of skin lesions. Ahmed et al. [19] developed novel hybrid algorithms to classify dermoscopic images of skin lesions using two datasets, HAM10000 and PH2. To guarantee peak performance, they precisely calibrated the pictures and fixed the imbalance in the dataset. The ResNet101 and MobileNet pre-trained models were used to assess the PH2 and HAM10000 datasets. Several hybrid methods, such as SVM-MobileNet, SVM-ResNet101, and SVM-MobileNet-ResNet101, were used to diagnose skin lesions early.

These hybrid methods performed better than CNN models that had previously been trained to capture distinctive features, including colour, texture, and shape. These unique qualities were combined with those from the MobileNet and ResNet101 models to create a very lifelike feature set. An artificial neural network (ANN) was trained using the handmade ResNet101 and upgraded MobileNet feature sets, producing unexpectedly accurate results. The results show that when MobileNet is paired with manually created features, an AUC of 97.53%, an ACC of 98.4%, an ST of 94.46%, a PR of 93.44%, and an SP of 99.43% can be achieved on the HAM10000 dataset. Their method worked very well with the PH2 dataset, achieving perfect scores for all assessed criteria. According to a recent study by El-khatib et al. [20], DL may significantly improve the accuracy of melanoma diagnosis. Because computer-assisted detection can save lives, especially in the medical field, it has attracted significant attention. With applications including routine health examinations, patient prognosis, disease prevention, cost-effective treatment, effective population management, and patient empowerment, this study demonstrates DL as a potentially revolutionary advancement in this field. The development of an automated melanoma detection module, intended for inclusion in an Electronic Health Record (EHR) system, was motivated by these advantages. First, a system that combined evaluations from many neural networks—DenseNet-201, ResNet-101, DarkNet-53, Inception-V3, InceptionResNet-V2, GoogLeNet, and Inception—had to be constructed. To evaluate the system's performance, researchers also used Google Vertex AI, Microsoft Azure ML, Salesforce Einstein Vision, and Google Teachable Machine. It was suggested that one start with the F1 when integrating this system into an EHR platform. DermIS and ISIC 2020 are two different datasets used in the study for training and assessment to evaluate how well the models respond to various visual representations. To verify the feasibility of this idea, the study also assessed the system's performance using recent research and applications.

To address problems with limited characterisation capabilities, underutilisation of channel information in the diagnosis of melanoma skin cancer, and feature information loss during downsampling (ds), Wu et al. [21] introduced a unique skin pathological mirror classification approach. By integrating wavelet downsampling with a multichannel attention mechanism, the technique faithfully reproduces the problematic aspects of both high-frequency and low-frequency components. It is based on discrete wavelet downsampling (ds) for feature reconstruction. By using channel information as efficiently as possible, this method reduces the significant feature information loss that comes with downsampling. The three main input components used in the skin cancer classification model are wavelet ds, 333 traditional convolution, and depth-separable, which minimises the number of components while maintaining the precise perceptual field. The Hard-Swish activation function (AF) and Wavelet ds are used to optimise the model's residual module to improve feature representation. ImageNet transfer learning is used to build and fine-tune the network's weight parameters using the larger HAM10000 dataset. The experiment's results show that the suggested method's ACC has significantly improved, reaching 95.84%. Compared to previous methods, our technique offers superior noise immunity, faster processing, and more accurate skin cancer case categorisation. This study provides a novel and useful categorisation system for skin cancer. The "You Only Look Once" neural network model for smartphones was tested by Liutkus et al. [22] for its ability to recognise three distinct types of skin lesions: seborrheic keratoses, melanocytic nevi, and melanomas. Fifty-nine thousand ninety dermatoscopic images were used to train the neural network from the large dataset.

Thirty-two seborrheic keratoses, thirty-five melanocytic nevi, and thirty melanomas were the lesions used in the research to assess the algorithm's efficacy. This enabled evaluation of the system's ACC. The algorithm's performance was contrasted with that of dermatoscopists with varying levels of experience. With 95% confidence intervals for range, specificity, and ST between 0.71 and 0.96, 0.76 and 0.94, and 0.88 and 0.88, the approach demonstrated good performance in the identification of melanomas. An algorithm that performed better than dermatologists without any prior training had a standard deviation of 0.83, indicating a 95% confidence interval of 0.77-0.87. In terms of melanocytic nevi identification, the approach performed better than other dermatological groups, with an ST of 0.77 (95% confidence interval: 0.60-0.90). With this method, however, the ST for seborrheic keratoses was much lower, at 0.52 (95% confidence interval: 0.34–0.69). In conclusion, the "You Only Look Once" neural network model for smartphones achieved levels of accuracy and sensitivity that even highly trained medical experts could not match in diagnosing melanocytic nevi and melanomas. Conversely, increasing the dataset size may increase the ST for seborrheic keratoses.

2.1. Problem Statement

Skin cancer is a prevalent and possibly fatal illness that affects millions of individuals globally. A better prognosis and effective therapy depend on early discovery and accurate diagnosis. This paper explores the application of CNN, a type of DL technique, to create a novel skin cancer detection system. The paper uses publicly accessible benchmark datasets, including HAM10000 and ISIC 2020, that contain irregular lesion characteristics and image artefacts. The proposed model includes new methods for contrast enhancement and lesion image classification, as well as image preprocessing, feature segmentation, feature classification, and hyperparameter tuning for Classification. With an astounding 99.78% ACC rate on ISIC-2020 and 99.81% on the HAM10000 dataset, the model outperformed previous models and offered dermatologists and other medical practitioners invaluable assistance in identifying skin cancer.

3. Proposed Methodology

Figure 1 depicts the stages for implementing the proposed strategy. Image Preprocessing, segmentation using MFETok-MLP-based UNet, and Classification using an MPOA-based FRNN classifier, with MPOA used for hyperparameter tuning, are all included in this section.

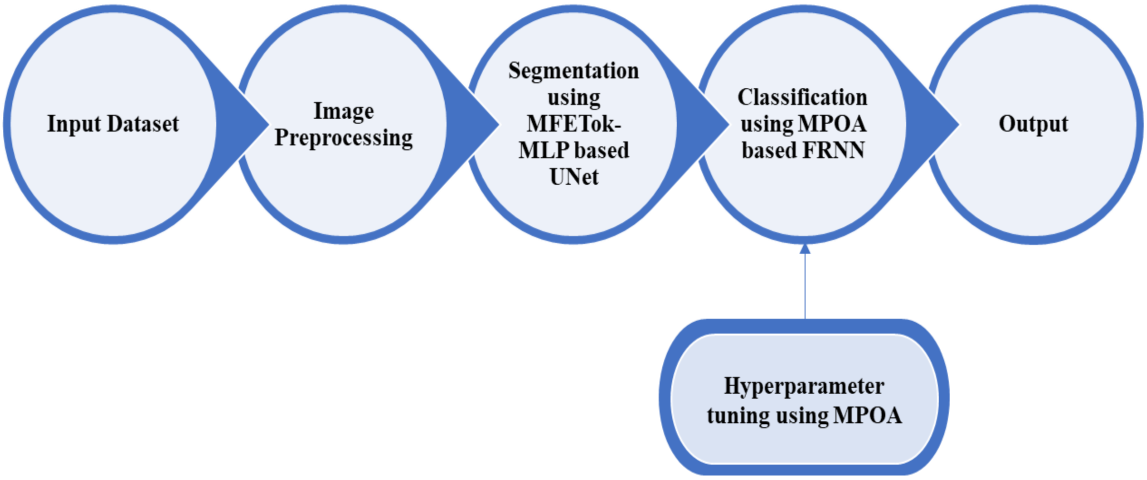


Figure 1: Workflow of the proposed models

3.1. Dataset Description

Dermoscopy images that display the worldwide prevalence of skin cancer are among the many publicly accessible tools on the internet. As shown in Figure 2, this research focuses on photomicroscopy and dermoscopy images, with particular emphasis on four distinct types of skin cancer. To ensure comprehensive training and evaluation of the proposed model, two distinct datasets were used, sourced from different sources. The ISIC-2020 Archive, which has 33,126 dermoscopic images, is one of the largest publicly available collections of skin lesions worldwide. Thirty-two thousand five hundred forty-two of these photos show benign skin lesions, whereas the other 579 depict malignant skin lesions. These photos come from a variety of sources, including patient data from different age groups and donations from various institutions. This dataset, astonishingly, captures the experiences of more than 2,000 patients. Thorough histological examination was performed for the 579 photos in the melanoma class; other images in the non-melanoma group are outside the purview of this study. ISIC, a partnership between the ViDIR Group at the University of Vienna in Austria and Queensland University in Australia, produced the 10,015-image HAM10000 database in 2018. This database classifies various skin lesions into seven distinct groups. Specifically, 2,007 photos from the melanocytic nevi class, 510 images from the basal cell class, and 1,107 images from the melanoma class were included. The study did not include dermoscopic pictures from other demographic groups in the database analysis.

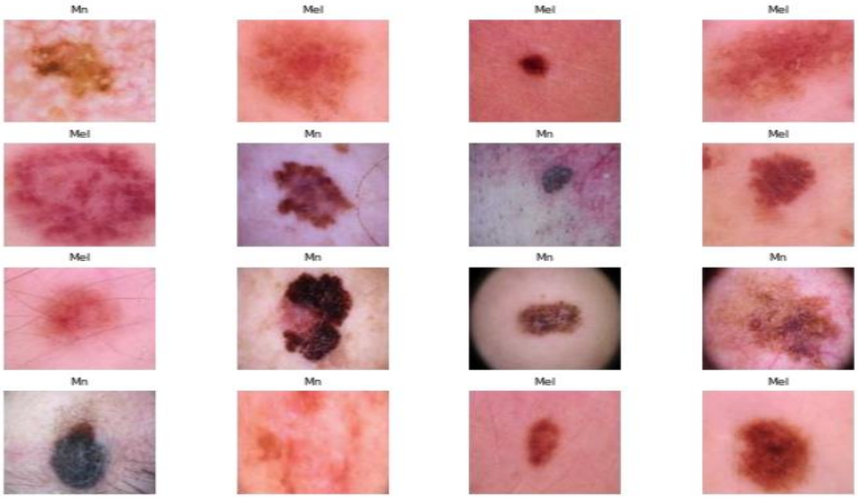


Figure 2: Skin cancer image samples extracted from two datasets

3.2. Image Preprocessing

Contrast enhancement is a key component of the lesion diagnostic method, as was noted in [24]. The research now in publication has investigated several approaches to address low contrast. This research presents a novel method that improves contrast by using texture and colour information. This method uses the unique texture and colour traits that skin lesions often exhibit, in contrast to healthy skin patches. As shown below, equations (1) and (2) are used to calculate the textural information using normalised luminance channels:

$$\varphi_L(u, v) = \left(\lambda \times F(Y)\right) - 16 \tag{1}$$

$$F(Y) = \begin{cases} \sqrt[3]{Y} & \text{for } Y > 0.01 \\ (7.787 * Y) + \frac{16}{\lambda} & \text{elsewhere} \end{cases} \tag{2}$$

$(Y)=\tilde{Y}/100$, $\tilde{Y} = \omega_i \times G, i \epsilon \{0.212,0.715,0.072\}$and $\lambda = 116$. G, the green channel in the original RGB image, is represented by the formula G = 3G. Let's use the provided format to rewrite equation (3):

$$L(u, v) = \varphi_L\left(\sum_{j=1}^{3} \frac{I(u,v)}{3}\right) \tag{3}$$

The original RGB picture in this context is denoted by I (u, v), and the luminance function, as defined by equation (3), is denoted by jL. Equation (4) below defines the Gaussian function that will be used for the luminance picture to examine the textural properties inside the lesion region:

$$\rho(u, v, \sigma) = \frac{L(u,v)}{\varphi(u,v,\sigma)} - L(u, v) \tag{4}$$

In this equation, φ (u, v, σ) is the result of convolving L (u, v) with a filter called the Gaussian using the factor σ (standard deviation). Further information on the σ computation is provided in Equation (5):

$$\sigma = \sqrt{\frac{\sum(uv)^2}{N} - \left(\frac{\sum uv}{N}\right)^2} \tag{5}$$

Equations (6), (7), and (8) simplify the formula ρ (u,v,σ) mentioned before:

$$\rho(u, v, \sigma) \ = \frac{L(u,v) - L(u,v) \times \varphi(u,v,\sigma)}{\varphi(u,v,\sigma)} \tag{6}$$

$$\rho(u, v, \sigma) \ = \frac{L(u,v)[1 - \varphi(u,v,\sigma)]}{\varphi(u,v,\sigma)} \tag{7}$$

$$\rho(u, v, \sigma) = \frac{L(u,v) \times Z}{\varphi(u,v,\sigma)} \tag{8}$$

Z may be written as 1 – φ (u, v, σ). Low-intensity pixels are typically seen in the dermoscopic picture within the lesion region. To differentiate between pixels in the picture that belong to the lesion and those of the surrounding skin, an AF is used. The AF is defined as follows in equation (9):

$$F(A) = \begin{cases} \tilde{\varphi_L}(u, v) & \text{if } \rho(u, v, \sigma) > \varphi(u, v, \sigma) \qquad \text{Lesion Area} \\ \tilde{\varphi_H}(u, v) & \text{otherwise} \qquad\qquad\qquad\quad \text{Healthy Skin Area} \end{cases} \tag{9}$$

The symbols show the damaged lesion and the unaffected healthy skin. $\tilde{\Phi}_H(u, v), \tilde{\varphi_L}(u, v)$, respectively, in the context of the information provided. The Retinex Model was used further to improve the colour properties of the final pixels. Equation (10) explains the function of this model as colour correction, which is as follows:

$$\varphi_{\text{Retinex}}(u, v) = \frac{\tilde{\varphi}_L^i(u,v)}{\tilde{\varphi}_L^i(u,v) \otimes G(\sigma)} \tag{10}$$

When i ϵ the symbol represents L, A, B, the convolution operation ⨂, and the Gaussian filter with a given standard deviation is denoted by G(σ). The preprocessing step yields several noteworthy results, as shown in Figure 3. The figures in this presentation clearly demonstrate how the recommended approach was applied to successfully resolve the contrast issue. These enhanced photos thus become extremely important during the model's learning process.

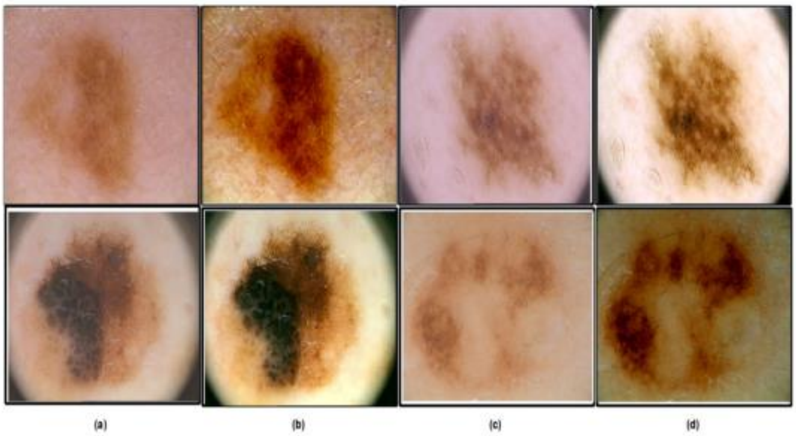


Figure 3: Lesion contrast enhancement results: (**a**,**c**) original image; (**b**,**d**) enhanced image

3.3. MFETok-MLP-Based UNet Segmentation

Building on the UNet architecture, a segmentation network called MFETok-MLP was devised, as shown in Figure 4, to overcome difficulties caused by complex backgrounds and varying illumination conditions. This network's unique elements, the MFE and Tok-MLP modules, are what make it unique. The MFE module is notable for its ability to effectively extract useful and detailed characteristics without using conventional pooling procedures. Using multi-scale and multi-type convolution kernels, it does this. In addition, a frequency attention mechanism (FAM) improves the robustness and the model's capacity to recognise factors associated with cancer. Similarly, the Tok-MLP module efficiently controls model complexity by dynamically using sparse attention. Thus, it enables the algorithm to continuously balance global and local properties, thereby making the best use of global information. Also integrated DSC and cross-entropy loss for model optimisation. This combination ensures accurate lesion diagnosis and effectively addresses concerns about class imbalance. The following are the detailed computation formulae for Equations (11) and (12):

$$L_{CE} = -\sum_{i=1}^{m} t \times \log(y) \tag{11}$$

$$L_{\text{Dice}} = 1 - \frac{2\sum_{i=1}^{n} y \times t + \varepsilon}{\sum_{i=1}^{n} y + \sum_{i=1}^{n} t + \varepsilon} \tag{12}$$

Here, 't' stands for real or true labels, and 'y for model predictions. m ' and 'n' are the variables that represent the total number of classes and pixels, respectively. Furthermore, the hyperparameter 'ε' is included in equation (13) with a value of $1 \times e^{-5}$ to exclude situations in which division by zero may occur expressly:

$$L = \alpha \times L_{CE} + \beta \times L_{Dice} \tag{13}$$

In this formula, the weight coefficients a and b have been set to 1 and 0.5, respectively.

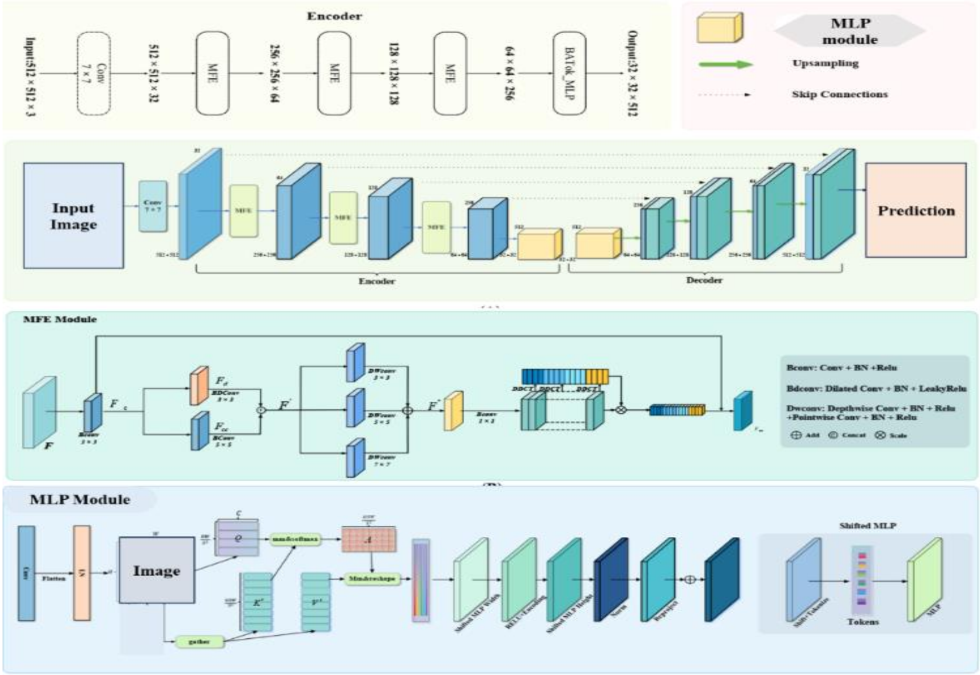


Figure 4: MFETok-MLP-based UNet framework. (**A**) structure of MFETok-MLP-based UNet (**B**) MFE architecture (**C**) Tok-MLP architecture

In contrast to UNet and other cutting-edge networks, the study has shown that these improvements considerably improve the recommended model's overall efficacy and ACC. The main reason for this improvement is the application of new optimisation techniques and feature extraction components. These modifications ensure the model can identify and localise items even when complex backdrops and varying lighting conditions are present.

3.3.1. MFE Module

The convolution technique is a key component of CNNs, which are used to extract local information from pictures. This process skillfully captures precise picture details while maintaining the spatial properties. CNNs frequently use pooling techniques to lessen computing complexity and mitigate overfitting. However, because of the dimensionality reduction, this method may unintentionally decrease ST to smaller targets. A new multiscale module with residual connections has been created to address this issue and improve the ACC of skin cancer localisation and detection. This module operates without pooling, unlike traditional CNNs. It functions across various convolution sizes and levels, as shown in Figure 4B, efficiently maintaining information continuity and enhancing the PR of skin cancer localisation and detection. The MFE module follows the steps listed below: A 3 × 3 convolution layer is used to extract features from the input feature map $F\epsilon. \mathbb{R}^{B\times H\times W\times C}$ to produce a base feature map known as F_c. The fundamental feature map $F_c$is then split into two branches after this first stage. One branch extracts a feature map. F_d that concentrates on fine-grained, local information using a 3 × 3 extended convolution. The alternative branch emphasises more global, semantically rich, and broader information in its feature map, the FCC, generated by a 5x5 convolutional layer. These two branches provide feature maps, F_d F_{cc}, which are then merged along the channel dimension to yield the final feature map, F^', as described by equation (14):

$$F' = \mathrm{Concat}(F_d, F_{cc}) \tag{14}$$

Where Concat denotes concatenation, it uses kernel sizes of 3 × 3, 5 × 5, and 7 × 7 to apply depthwise convolutional operations to this novel feature map. This novel method not only maintains the model's outstanding feature extraction capabilities but also significantly reduces the total number of variables and the computational burden. Convolutional operations with different kernel sizes were used to improve the model's ability to capture information across multiple dimensions. Researchers integrated the outcomes from these three branches, producing a more complex and comprehensive depiction that encompasses a variety of tumour characteristics:

$$F'' = \phi_1\big(D_3(F') + D_5(F') + D_7(F')\big) \tag{15}$$

During the procedure, the study uses a 1×1 convolutional layer, $\phi 1$, to perform dimensionality reduction and channel compression. By combining features of different sizes simultaneously, this method reduces the feature map size and improves the model's overall performance. The symbol Di, which denotes an $i \times i$ $(i = 3, 5, 7)$ dilated convolution layer, is introduced in equation (15). Feature Attention Module (FAM) integration has greatly enhanced the model's ability to focus on key frequency features in images, especially those associated with skin cancer detection. This improvement strengthens the model's resilience and its understanding of global data, enabling it to identify factors associated with skin cancer. Moreover, the FAM effectively removes aberrant textures and ambient noise commonly present in images, thereby improving segmentation accuracy and reliability in the skin cancer detection model. In real-world applications, the input feature map is usually divided into several groups. After partitioning, the final weighted feature map is produced by applying a two-dimensional discrete cosine transform (2DDCT) and a fully connected layer. Equation (16) shows how the final output feature map is produced by smoothly integrating this weighted feature map into the model with the help of a residual structure:

$$F'' = [F''0, F''^{1}, \ldots, F''^{n-1}] \tag{16}$$

$$\mathrm{Freq} = \mathrm{Concat}([2\mathrm{DDCT}(F''), 2\mathrm{DDCT}(F''^{1}), \ldots, 2\mathrm{DDCT}(F''^{n-1})]) \tag{17}$$

$$F_{\mathrm{out}} = \mathrm{Sigmoid}\Big(\mathrm{FC}\big(2\mathrm{DDCT}(F'')\big)\Big) \times F'' + \phi_3(F) \tag{18}$$

"Freq" in Equation (17) refers to the frequency domain properties obtained from the input data by means of a 2DDCT transformation. The pre-divided ratio of these input characteristics is represented by the variable "n" in this equation. The function "Sigmoid," which shows a correlation with "AF," is introduced in Equation (18). The fully connected network layer is represented by "FC". The MFE module applies multiple convolutional processes of different sizes and types to improve feature extraction. Furthermore, a frequency-averaged mask (FAM) is included to direct the model's attention to the image's important frequency-based components. The ACC and model resilience are strengthened by this frequency-domain transformation, leading to improved overall performance.

3.3.2. Tok-MLP Module

To gradually encode and decode input data, the encoder-decoder architecture of the UNet uses a series of convolutional and deconvolutional operations. This enables mapping and extracting features at different levels of the hierarchy. Lower levels of architecture emphasise finer details, such as edges and textures, while higher levels focus on abstract and global elements, such as objects and landscapes. However, in some cases, when conventional convolutional algorithms fail to capture sufficient global contextual information, fine-grained features may be lost. An inventive Tok-MLP module with dynamic sparse attention was introduced to address this problem (Figure 4C). This module greatly improves the model's ability to handle global contextual information, thereby lowering the risk of losing fine-grained features. By examining the feature map's dimensions, the Tok-MLP module effectively collects detailed data and enables the model to focus on particular areas within the convolutional features [25]. However, because it depends on local attention mechanisms that concentrate on specific regions, its ability to retrieve global information is limited. As mentioned in reference, the Tok-MLP module was improved by utilising an adaptive sparse attention mechanism to overcome this restriction [26]. This multi-cycle process allows different input sequence segments to be the focus of each calculation, leading to a more comprehensive understanding of the data. Through continual attentional switching, it is possible to learn about the cosmos and become aware of the intricacies in the here and now. Additionally, a patch embedding layer was added to the Tok-MLP module. This layer preserves spatial information by converting a 2D input to a 1D sequence using a 2D convolutional layer and layer normalisation. Now that two-dimensional spatial data can be integrated into a one-dimensional sequence, the model may facilitate feature extraction and integration. The following stages represent the operational procedure of the Tok-MLP module. First, a patch embedding layer is used to process the input feature map $X \in \mathbb{R}^{B\times H\times W\times C}$. The output sequence Ep is then transformed into $X_{\mathrm{reshaped}} \in \mathbb{R}^{B\times H\times W\times C}$, lining it up with the requirements for the procedures that follow. The exact order of these steps is as follows:

$$X_t = (\ \mathrm{Flatten}\ (\mathrm{Conv}(X, K)))^{T} \tag{19}$$

$$E_p = \big(X_t - \mu(X_t, \dim = 1)/\sqrt{(\sigma^2(X_t, \dim = 1) + \varepsilon)}\big) \tag{20}$$

Where the functions m () and s (), respectively, indicate the mean and variance of the feature dimension, and Conv () denotes convolutional procedures in equation (19). The definition of one hyperparameter, ε, is $1 \times e^{-6}$. The Flatten () function transforms two-dimensional features into one-dimensional sequential features, and the one-dimensional embedded features in equation (20) are generated as $E_p \in \mathbb{R}^{B\times N\times C}$ $N = H \times W$. Following the reshaping process, the feature map that was input $X_{reshaped}$ is divided into many nonoverlapping areas of size S × S. Individual regions are then combined to form feature vectors. Subsequently, to obtain the three tensors Q, K, and V, linear projections were applied to each feature vector in equation (21):

$$F_r = \text{Partition}\left(X_{\text{reshaped}}, S\right) \tag{21}$$

$$Q = F_r \cdot W_q, K = F_r \cdot W_k, V = F_r \cdot W_v \tag{22}$$

where W_q, W_k and W_v are learnable weight matrices. Equation (22) computes the attention ratings for each region using the query (Q), key (K), and value (V) tensors. To create region-level queries and keys ($Q_r$and K_r), Q and K are first averaged across a certain dimension (dim = 1). After that, $Q_r$and K_r were multiplied matrix-wise to create the adjacency matrix X_r, which shows how highly correlated the regions are. The following top-k operation of the assignment gathers the indices of the k most relevant regions, I_r, for each area. The most relevant key-value combinations that are obtained using I_r for each region i is K_g and V_g. The obtained key-value pairs were then subjected to an attention operation, yielding the result X'. Equations (23) through (27), which correspond to the necessary mathematical formulas, are shown below:

$$Q_r = \mu(Q, \text{dim} = 1), K_r = \mu(K, \text{dim} = 1) \tag{23}$$

$$X_r = Q_r(K_r)^T \tag{24}$$

$$I_r = \text{Top}\,K(X_r, k, \text{axis} = 1) \tag{25}$$

$$K_g = \text{Gather}(K, I_r),\ V_g = \text{Gather}(V, I_r) \tag{26}$$

$$X' = \text{attention}\left(Q, K_g, V_g\right) \tag{27}$$

When the most important key-value pairs are represented by $K_g$and V_g, the adjacency matrix X_r represents the correlation degree across regions, and X' is the output generated after the focus operation on the collected key-value pairs. The queries and keys at the region level are Q_r and K_r. The attention map is fed into the TokMLP module after the dynamic sparse attention mechanism has completed collecting information and transformed it into a new tensor X' of dimensions. $X' \epsilon\, \mathbb{R}^{B\times N\times C}$. Several operations are performed by this module, including: The shift operation is applied first to the width dimension. To achieve feature tokenisation, a 3 × 3 convolutional kernel is used to map the channel dimension to the embedding dimension E. Following that, a depthwise convolution layer and a Shifted MLP with a dimension H handle these tokens. The relevant mathematical formulas for these modifications are given by equations (28)-(30):

$$X_{\text{shifted}} = \text{Shift}_{\text{width}}(X) \tag{28}$$

$$T_W = \text{Tokenize}(X_{\text{shifted}}, \text{kernel}_{\text{size}} = 3, \text{channels} = E) \tag{29}$$

$$Y = \text{DWConv}\left(\text{MLP}(T_W, \text{hidden}_{\text{dim}} = H)\right) \tag{30}$$

Where DWConv is a representation of a convolution that is depth-wise separable, and W is the feature map width. First, the nonlinear representation of feature map Y is improved by using the ReLU AF after depthwise convolution. The active feature map, shown by the sign O', is the result of this. $Y_{shifted}$, a changed feature map, is the result of a vertical shift operation that follows. The shifted features are then tokenised again with an embedding size E using a 3×3 convolutional kernel, yielding a new set of tokens, TH. Following that, these TH tokens are sent to a ShiftedMLP module for further handling. This Shifted MLP module produces an output of dimension O, called Z. This is interesting because it establishes a residual connection to support long-range feature queries by adding the initial TH tokens to Z. To ensure feature distribution consistency across LRs,lr normalisation is used. This enhances model training and generalisation. The result of this last phase is indicated by Y_{final}. formulas. The mathematical formulas for these methods are provided in equations (31)-(34):

$$Y_{\text{shifted}} = \text{Shift}_{\text{width}}(Y) \tag{31}$$

$$T_H = \text{Tokenize}\,(Y_{\text{shifted}}, \text{kernel}_{\text{size}} = 3, \text{channels} = E) \tag{32}$$

$$Z = \mathrm{MLP}(\mathrm{ReLU}(T_H),\ \mathrm{output}_{\mathrm{dim}} = O) \quad (33)$$

$$Y_{\mathrm{final}} = \mathrm{LN}(Z + T_H) \quad (34)$$

The term "$\mathrm{Shifted}_{\mathrm{height}}$ " here denotes a shift operation that was performed along the height dimension. The acronym "ReLU" stands for the Rectified Linear Unit, or AF. "Tokenize" and "MLP" refer to a Multilayer Perceptron that has tokenisation and predefined output dimensions, respectively. Finally, the term "LN" denotes lr normalisation. The Tok-MLP module is a major factor in the seamless integration of local and global data. It is included in the last step of the UNet network encoder's feature extraction process. This integration enhances fine-grained detail extraction by varying feature sizes within the Tok-MLP module. This work also addresses the shortcomings of conventional convolutional algorithms in comprehending global contextual data. A dynamic routing sparse attention strategy is used to do this, greatly enhancing the model's comprehension and representation of input. After segmentation, a classification technique is used to improve the model's ACC for skin cancer detection. To achieve Classification, a unique method known as FRNN is utilised. The following section provides a thorough description of the FRNN classification procedure.

3.4. FRNN-Based Classification

To improve the robustness of the proposed model, the FRNN is used in the classification procedure [27]. The recurrent architecture and its unfolding computational graph are shown in Figure 5 for the proposed FRNN model in both the training and testing phases.

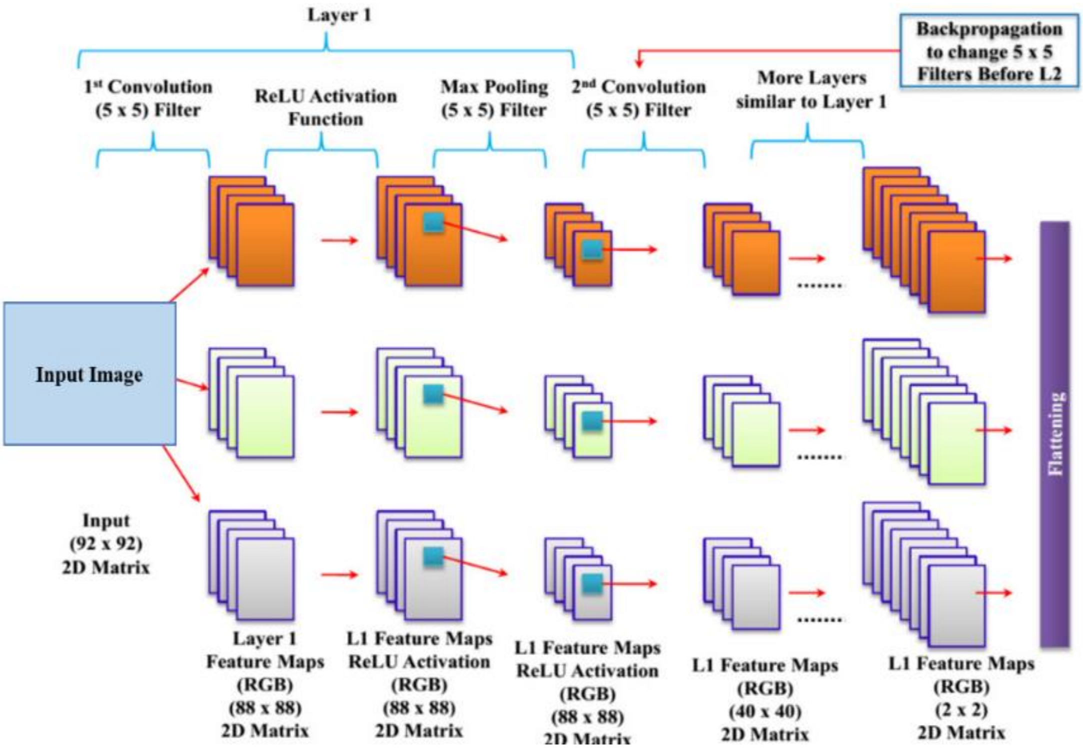
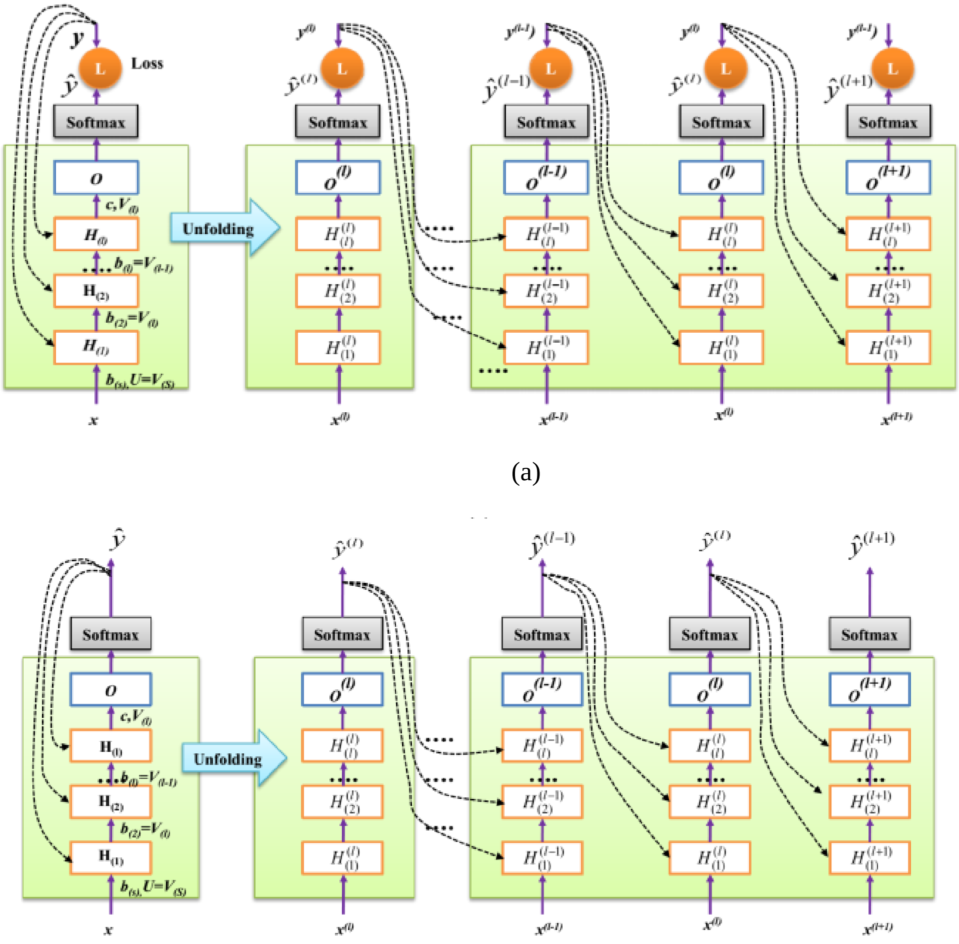


Figure 5: CNN architecture for feature extraction

The idea of defining the connections between the hidden and output layers using the m-dimensional parameter U [e1D] is presented in Figure 5. It also shows the graded matrix V(k) [ek + 1ek], which shows the relationships between the hidden and forward hidden lrs, and the weighted matrix V(k) [Sek], which shows the relationships between the hidden (hid) and output lrs. Notably, depending on the input data distribution and the data streams used during training, the value of ek (where $1 \leq k \leq l$) may change. Next, let's look at the P-FRNN structure. As shown by the dot-filled arrows in Figure 6, it explicitly maintains recurrent connections between hidden layers using hyperplane activity. In contrast to certain other systems, P-FRNN establishes an effective connection between the hidden and output levels without requiring an external set of weights. The input x(t) is represented in Figure 6a as x(t) R D, where D is the input dimension, for each time step t. The initialisation of the HidLR, H(t)(k), where indicates the hidlr associated with k (where $1 \leq k \leq l$). Moreover, the unnormalised output at time t is denoted O(t).



(a)

(b)

Figure 6: FRNN architecture (**a**) training phase and (**b**) test phase

The outcome from the previous time step indirectly affects each learning method in this section when discussing the bias "k" level, represented by "R" with respect to the "k-1" feature plane, during the integration of a new magnitude. This effect is called hyperplane-dependent and does not require any external strength training or parameter adjustments. When the model is being tested, the actual output, denoted "y(t1)," is replaced with the predicted output, "yb(t1)," which is the result of a forward-propagation estimate within the Hyperplane at the hidden level of the P-FRNN model. Equation (35) uses the substitution technique:

$$\mathrm{H(k)} = \mathrm{e}^{\left(-\eta \frac{\mathrm{d}^{\mathrm{i(t)}}_{\mathrm{(k)}}}{\max\left(\mathrm{d}^{\mathrm{i(t)}}_{\mathrm{(k)}}\right)}\right)} \tag{35}$$

The block size, stride and link code are among the common parameters shared by the convolutional layers in both pooling layers in the context as described. In the hyperplane distance equation (37), the representation of the ith feature and the (t - 1) th data at the kth hidden layer is represented as $\mathrm{d}^{\mathrm{i(t)}}_{\mathrm{(k)}}$:

$$\mathrm{d}^{\mathrm{i(t)}}_{(1)} = \frac{|\mathrm{y}^{(\mathrm{t}-1)}|_1 \mathrm{S}\left(\mathrm{b}^{\mathrm{i}}_{(1)} + \mathrm{U}_{\mathrm{i}} \mathrm{x}^{(\mathrm{t})}\right)}{\sqrt{1 + \sum_{\mathrm{j}=1}^{\mathrm{D}} \mathrm{U}_{\mathrm{ij}}^2}} \tag{36}$$

where k is 1, |y|1 is the 1-norm of y, S is the output dimension, and R e1 is the bias at level 1 that corresponds to b (1) concerning the feature plane's extra aspect. Likewise, for 1 < k ≤ l in equation (37):

$$d_{(1)}^{i(t)} = \frac{|y^{(t-1)}|_1 S \cdot \left(b_{(k)}^{i} + V_{(k-1)i}\right) H_{(k-1)}^{(t)}}{\sqrt{1 + \sum_{j=1}^{e_{k-1}} V_{(k-1)ij}^{2}}} \tag{37}$$

The non-normalised output for a given time step, "t", may be computed using equation (38) once the hidden parameters are known. Figure 6b illustrates this method:

$$O^{(t)} = c + V_{(K)} H_{(k)}^{(t)} \tag{38}$$

As a result, the output bias, expressed as 'R S' and symbolised by the symbol 'c,' is equivalent to 'k' (short for 'l'). To arrive at the desired result, yb(t), the first step is to create an unnormalized logarithmic probability, which is then normalised using the softmax activation function, as shown in equation (39). This process is backed by:

$$\widehat{y^{(t)}} = \text{softmax}(O(t)) \tag{39}$$

$$L(y, \hat{y}) = -\sum_i y_i \cdot \log(\hat{y}) \tag{40}$$

In equation (40), L refers to loss, which can be created in several ways. The y band y represents the measured outputs at a given time. The MPOA approach is used to tune the FRNN classifier's hyperparameters, resulting in a significant improvement in skin cancer detection ACC. This development helps with more accurate and trustworthy diagnoses, which is a great benefit to the healthcare industry. An algorithm has been included, along with a thorough explanation of the MPOA procedure used to tune the hyperparameters below.

3.4.1. Hyperparameter Tuning Using MPOA

Hyperparameters must be fine-tuned to achieve the highest possible ACC in diagnosing skin cancer. The research uses the MPOA (Marine Predator Optimisation Algorithm) to achieve this. The population-based metaheuristic method, MPOA, was inspired by the behaviour of marine predators. Equation (41) describes how the initial trial's solution is randomly distributed over the search region:

$$X_0 = X_{\min} + \text{rand}(X_{\max} - X_{\min}) \tag{41}$$

Where the variables' minimum and maximum boundaries are represented by $X_{\min}$ and $X_{\max}$, respectively, and the random value vector rand lies in the range of [0,1]. Based on the "survival of the fittest," equation (42) provides an especially robust approach, labelling the dominant predators and placing them in charge of creating a matrix referred to as the "elite. By utilising information on the prey's locations, the matrices linked to the elite individuals are essential for coordinating the hunt and finding prey:

$$\text{Elite} = \begin{bmatrix} X_{1,1}^{I} & X_{1,2}^{I} & \cdots & X_{1,d}^{I} \\ X_{2,1}^{I} & X_{2,2}^{I} & \cdots & X_{2,d}^{I} \\ \vdots & \vdots & \vdots & \vdots \\ X_{n,1}^{I} & X_{n,2}^{I} & \cdots & X_{n,d}^{I} \end{bmatrix} \tag{42}$$

In this case, $\vec{X}^{I}$ which is replicated repeatedly to generate the most powerful matrix, represents the apex predator. In this case, n denotes the total number of agents, and d denotes the size of each agent. Every time a superior predator replaces an apex predator, the elite matrix is altered. Predators must rearrange themselves according to the prey matrix, as shown below [28]:

$$\text{Prey} = \begin{bmatrix} X_{1,1} & X_{1,2} & \cdots & X_{1,d} \\ X_{2,1} & X_{2,2} & \cdots & X_{2,d} \\ \vdots & \vdots & \vdots & \vdots \\ X_{n,1} & X_{n,2} & \cdots & X_{n,d} \end{bmatrix} \tag{43}$$

where represents jth dimension of ith prey. The velocity ratio determines the three main steps of the optimisation process and is primarily linked to the predator-prey matrices. Equation (43), which describes the life span that involves the predator and its prey, is successfully modelled by this process. Phase 1: Prey is travelling at a higher velocity ratio (i.e., quicker than the predator). Equation (44) presents a mathematical paradigm that is:

$\text{If Iter} < \frac{1}{3}\text{MaxIter}$

$$\begin{aligned}\overrightarrow{\text{stepsize}}_{i} &= \vec{R}_B \otimes \left(\overrightarrow{\text{Elite}}_{\imath} - \vec{R}_B \otimes \overrightarrow{\text{Prey}}_{i}\right)(i = 1,2,\ldots,n)\\ \overrightarrow{\text{Prey}}_{i} &= \overrightarrow{\text{Prey}}_{i} + P \cdot \vec{R} \otimes \overrightarrow{\text{stepsize}}_{\imath}\end{aligned} \tag{44}$$

Where P remains the constant at an assigned value of 0 and ⊗ denotes entry-wise multiplicity. $\vec{R}_L$ is a randomly generated number, the vector with an equal distribution in the range [0, 1], and $\vec{R}_B$ is a vector made up of random integers. Phase 2: The unit velocity ratio indicates that the predator and prey are travelling at the same speed. One way to model this might be as follows [29]:

If $\frac{1}{3}$ Max Iter $<$ Iter $< \frac{2}{3}$ MaxIter

Regarding the initial portion of the population in equation (45):

$$\begin{aligned}\overrightarrow{\text{stepsize}}_{\imath} &= \vec{R}_L \otimes \left(\overrightarrow{\text{Elite}}_{\imath} - \vec{R}_L \otimes \overrightarrow{\text{Prey}}_{i}\right)(i = 1,2,\ldots,n/2)\\ \overrightarrow{\text{Prey}}_{i} &= \overrightarrow{\text{Prey}}_{i} + P \cdot \vec{R} \otimes \overrightarrow{\text{stepsize}}_{\imath}\end{aligned} \tag{45}$$

For the second half of the population in equation (46):

$$\overrightarrow{\text{Prey}}_{i} = \overrightarrow{\text{Elite}}_{\imath_i} + P \cdot CF \otimes \overrightarrow{\text{stepsize}}_{i} \tag{46}$$

Here, CF is a dynamic parameter that controls the predator's gait, $\vec{R}_L$ denotes a set of random numbers that simulates Levy's movement:

$$CF = \left(1 - \frac{\text{Iter}}{\text{MaxIter}}\right)^{\left(\frac{2\times \text{Iter}}{\text{MaxIter}}\right)} \tag{47}$$

Phase 3, the last phase of the optimisation process, is marked by predators moving faster than their prey, resulting in a low velocity ratio. Previous research has indicated a strong correlation between this stage and increased exploitation capability [30]:

$$\begin{aligned}&\text{If Iter} > \frac{2}{3}\ \text{MaxIter}\\ &\overrightarrow{\text{stepsize}}_{\imath} = \vec{R}_L \otimes \left(\vec{R}_L \otimes \overrightarrow{\text{Elite}}_{i} - \overrightarrow{\text{Prey}}_{i}\right)(i = 1,2,\ldots,n)\\ &\overrightarrow{\text{Prey}}_{i} = \overrightarrow{\text{Elite}}_{\imath_i} + P \cdot CF \otimes \overrightarrow{\text{stepsize}}_{\imath}\end{aligned} \tag{48}$$

The procedure for MPOA is highlighted in Table 1 below.

Table 1: MPO algorithm

Objective function $f(x), x = (x_1, x_2, \ldots, x_d)$
Initialize population
Determine the memory − saving, elite matrix and then fitness values.
FOR $t = 1:$ generation of max
IF Iter $< \frac{1}{3}$ Max Iter
$\overrightarrow{\text{stepsize}}_{i} = \vec{R}_B \otimes \left(\overrightarrow{\text{Elite}}_{\imath_i} - \vec{R}_B \otimes \overrightarrow{\text{Prey}}_{\imath_i}\right)(i = 1,2,\ldots,n)$ $\overrightarrow{\text{Prey}}_{i} = \overrightarrow{\text{Prey}}_{i} + P \cdot \vec{R} \otimes\rightleftarrows \text{stepsize}_{i}$
ELSE IF $\frac{1}{3}$ MaxIter $<$ Iter $< \frac{2}{3}$ MaxIter
$\overrightarrow{\text{stepsize}}_{i} = \vec{R}_L \otimes \left(\overrightarrow{\text{Elite}}_{\imath_i} - \vec{R}_L \otimes \overrightarrow{\text{Prey}}_{\imath}^{2}\right)(i = 1,2,\ldots,n/2)$
The populace′s initial half is updated by $\overrightarrow{\text{Prey}}_{i} = \overrightarrow{\text{Prey}}_{i} + P \cdot CF \otimes\rightleftarrows \overrightarrow{\text{stepsıze}}_{\imath}^{2}$

```
The population′s second half is updated by $\overrightarrow{\mathrm{Prey}}_{\mathrm{i}}$
        $\overrightarrow{\mathrm{Elıte}}_{\imath\mathrm{i}} + \mathrm{P} \cdot \mathrm{CF} \otimes \overrightarrow{\mathrm{stepsize}}_{\mathrm{i}}$
ELSE IF Iter $> \frac{2}{3}$ MaxIter
    $\overrightarrow{\mathrm{stepsize}}_{\mathrm{i}} = \vec{\mathrm{R}}_{\mathrm{L}} \otimes \left(\vec{\mathrm{R}}_{\mathrm{L}} \otimes \overrightarrow{\mathrm{Elite}}_{\imath\mathrm{i}} - \overrightarrow{\mathrm{Prey}}_{\mathrm{i}}\right)(\mathrm{i} = 1,2, \dots, \mathrm{n})$
        $\overrightarrow{\mathrm{Prey}}_{\mathrm{i}} = \overrightarrow{\mathrm{Elite}}_{\imath\mathrm{i}} + \mathrm{P} \cdot \mathrm{CF} \otimes \rightleftarrows \overrightarrow{\mathrm{stepsize}}_{\mathrm{i}}$
END IF
Perform memory saving and elite updates based on
    $\overrightarrow{\mathrm{Prey}}_{\mathrm{i}} = \begin{cases} \mathrm{Prey}_{\mathrm{i}} + \mathrm{CF}[\mathrm{X}_{\mathrm{min}} + \mathrm{R} \otimes (\mathrm{X}_{\mathrm{max}} - \mathrm{X}_{\mathrm{min}})] \otimes \mathrm{U} \\ \mathrm{Prey}_{\mathrm{i}} + [\text{ if } \mathrm{r} \leq \mathrm{FAs}(1 - \mathrm{r}) + \mathrm{r}]\left(\overrightarrow{\mathrm{Prey}}_{\mathrm{r}_1} - \overrightarrow{\mathrm{Prey}}_{\mathrm{r}_2}\right) \text{ if } \mathrm{r} > \mathrm{FADs} \end{cases}$
(where FADs=0.2)
current best solution
END
```

4. Results and Discussion

4.1. Experimental Setup

The Python computer language was used to build the proposed approach, which has no direct connection to convolutional networks. The Windows 10 platform was used for the experimental phase, and the machine had 32 GB of RAM and an 11 GB NVIDIA GPU.

4.2. Performance Metrics

A popular statistic for assessing the effectiveness of segmentation models is ACC. According to equation (49), it calculates the percentage of properly recognised samples out of all samples:

$$\mathrm{Accuracy} = \frac{\mathrm{TP+TN}}{\mathrm{TP+TN+FP+FN}} \qquad (49)$$

The PR rate, introduced in equation (50), measures a model's ability to forecast positive samples among those it classifies as positive correctly:

$$\mathrm{Precision} = \frac{\mathrm{TP}}{\mathrm{TP+FP}} \qquad (50)$$

Equation (51), also called the ST and true positive rate, assesses how effectively a prediction model performs at recognising true positives. The ratio of true positives to the sum of actual positives and negatives is used to compute it:

$$\mathrm{Recall} = \frac{\mathrm{TP}}{\mathrm{TP+FN}} \qquad (51)$$

The F1, a measurement of ACC obtained from a weighted average of PR and RC, is defined by equation (52). It assesses the test's capacity to differentiate between favourable and unfavourable results:

$$\mathrm{F1} = \frac{2 \times \mathrm{precision} \times \mathrm{recall}}{\mathrm{precision+recall}} \qquad (52)$$

The overlap between two sets is calculated using the DSC in equation (53), which divides the sum of the sizes of the two sets by the size of their intersection:

$$\mathrm{DSC} = (2 * \mathrm{TP})/(2.\mathrm{TP} + \mathrm{FP} + \mathrm{FN}) \qquad (53)$$

By dividing the intersection's size by the union, the JSC in equation (54) determines how similar two sets are:

$$\mathrm{JSC=TP/(TP+FP+FN)} \qquad (54)$$

4.3. Segmentation Analysis

Table 2 compares the segmentation performance of the two datasets using the following metrics: DSC, JSC, ACC, PR and RC.

Table 2: Segmentation analysis with datasets

Datasets	ACC (%)	PR (%)	RC (%)	F1 (%)	DSC (%)	JSC (%)
ISIC-2020	96.5	96.6	96.5	96.5	96.6	96.6
HAM10000	96.7	96.7	96.8	96.8	96.8	96.7

From Table 2, for the MFETok-MLP-based UNet segmentation process, each image sample was considered. The ISIC-2020 dataset achieved an ACC of 96.5%, PR of 96.6%, RC of 96.5%, F1 of 96.5%, DSC of 96.6% and JSC of 96.6% and the HAM10000 dataset achieved an ACC of 96.5%, PR of 96.6%, RC of 96.5%, F1 of 96.5%, DSC of 96.6% and JSC of 96.6%. The two datasets achieved better ACC results when the proposed model used a robust segmentation process, namely MFETok-MLP-based UNet segmentation. Figure 7 depicts the graphical representation of the segmentation analysis across two datasets.

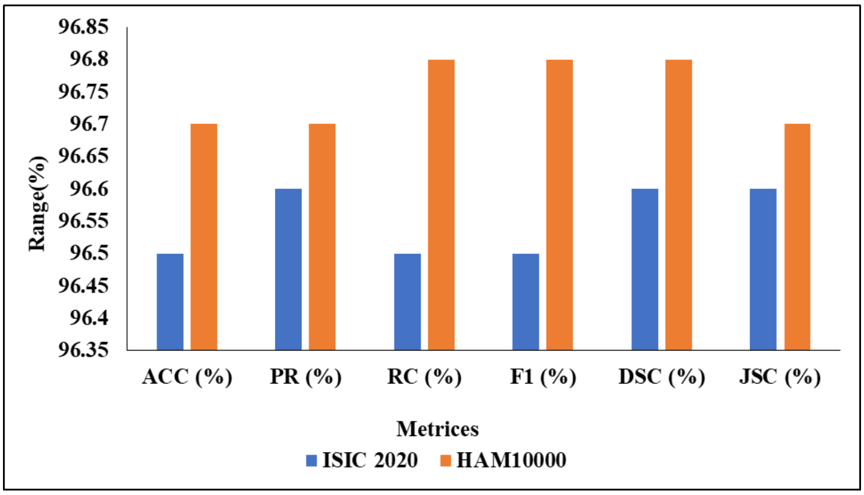


Figure 7: Analysis of segmentation with datasets

4.4. Classification Analysis

The existing models, such as VGG-16, VGG-19, MobileNet, ResNet-152, EfficientNet-B0, RCNN, VGG-13, and DBN, were used to evaluate the proposed model's performance, including ACC, PR, RC, and F1.

Table 3: Analysis of proposed classification technique with existing techniques for the ISIC-2020 dataset

Method	ACC (%)	PR (%)	F1 (%)	RC (%)
VGG-16	91.12	92.09	91.13	90.43
VGG-19	91.68	92.23	91.71	90.57
MobileNet	92.51	92.95	92.17	91.40
ResNet-152	89.32	90.73	89.27	88.21
Proposed Model	99.78	99.72	99.65	99.68

From Table 3, the classification analysis of the proposed technique compared with existing methods shows that VGG-16 achieved ACC of 91.12%, PR of 92.09%, F1 of 91.13%, and RC of 90.43%. VGG-19 had an ACC of 91.68%, PR of 92.23%, F1 of 91.71%, and RC of 90.57%. MobileNet had an ACC of 92.51%, PR of 92.95%, F1 of 92.17%, and RC of 91.40%. ResNet-152 had an ACC of 89.32%, PR of 90.73%, F1 of 89.27%, and RC of 88.21%. The proposed model achieved an ACC of 99.78%, PR of 99.72%, F1 of 99.65%, and RC of 99.68%. The proposed model achieved the best ACC results compared with existing methods, as it employs a unique technique called FRNN Classification with MPOA. As a consequence, the suggested strategy may be useful for detecting skin cancer, and the results of Table 3's analysis demonstrate the proposed Classification's robustness. Figure 8 compares the proposed classification technique with the existing technique on the ISIC-2020 dataset.

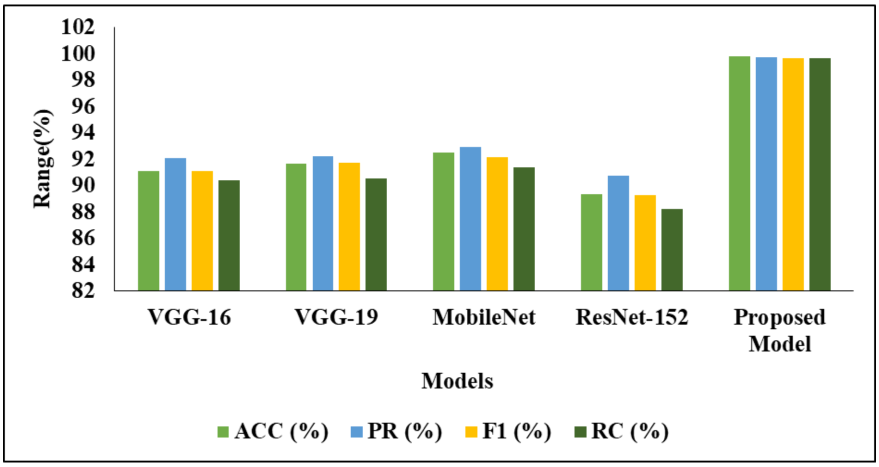


Figure 8: Graphical representation of the proposed classification technique with the existing technique for the ISIC-2020 dataset

In the classification analysis of the proposed technique against existing methods on the HAM10000 dataset, EfficientNet-B0 achieved ACC of 89.46%, PR of 90.21%, F1 of 89.31%, and RC of 88.21%. RCNN had an ACC of 85.50%, a PR of 84.50%, an F1 of 85.30%, and an RC of 83.40%. VGG-13 had an ACC of 89.57%, PR of 89.66%, F1 of 89.65%, and RC of 90.70%. DBN had an ACC of 93.00%, PR of 92.45%, F1 of 92.65%, and RC of 92.91%.

Table 4: Analysis of proposed classification technique with existing techniques for HAM10000 dataset

Method	ACC (%)	PR (%)	F1 (%)	RC (%)
EfficientNet-B0	89.46	90.21	89.31	88.21
RCNN	85.50	84.50	85.30	83.40
VGG-13	89.57	89.66	89.65	90.70
DBN	93.00	92.45	92.65	92.91
Proposed Model	99.81	99.77	99.80	99.75

The proposed model achieved an ACC of 99.81%, PR of 99.77%, F1 of 99.80%, and RC of 99.75%. Notably, across ACC, PR, F1 score, and RC, the proposed approach outperformed all other models.

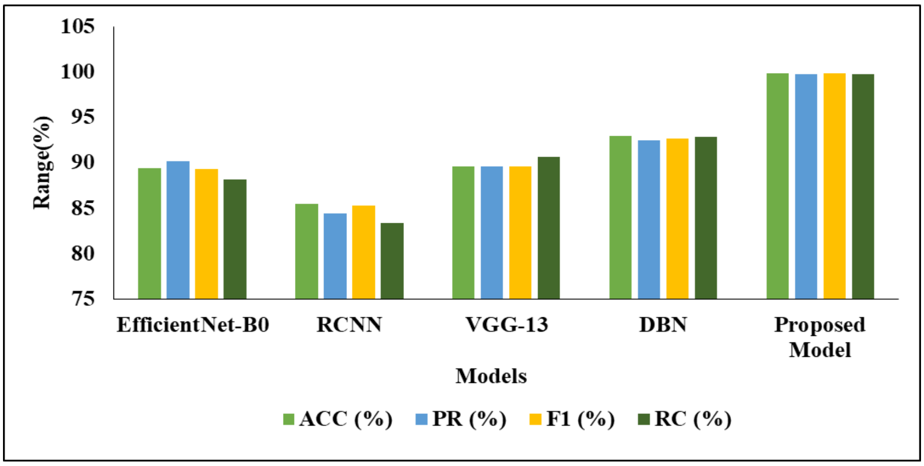


Figure 9: Graphical representation of the proposed classification technique with the existing technique for the HAM10000 dataset

This improved performance indicates that the proposed methodology is highly effective at classifying skin cancer. The remarkable findings presented in Table 4 illustrate how well-suited the proposed method is for skin cancer diagnosis, thanks to its unique approach. Additionally, compared with state-of-the-art approaches, Figure 9 illustrates how well the proposed classification method performs on the HAM10000 dataset.

5. Conclusion

This study presents a novel method for identifying skin cancer using CNNs, a DL technique. This paper discusses the complex problems of identifying and classifying skin lesions. These problems include hair interference, noise, uneven lesion features, and a variety of textures. It does this by implementing a skillfully designed DL architecture and using publicly accessible benchmark datasets, such as HAM10000 and ISIC2020. The suggested model includes processes such as feature segmentation, feature classification, image preprocessing, and hyperparameter tuning. Most notably, it presents a novel technique for preprocessing contrast enhancement using picture luminance information. Additionally, the model uses a classification method based on FRNN and blends the UNet network architecture with an MFETok-MLP-based UNet for segmentation. The model achieves an exceptional ACC rate in skin cancer diagnosis by utilising MPAO to fine-tune the hyperparameters of the FRNN classifiers. With 99.78% ACC on the ISIC-2020 dataset and 99.81% ACC on the HAM10000 dataset, the suggested model's results are quite impressive. These outcomes outperform current methods, indicating that the model has the potential to help dermatologists and other doctors accurately diagnose skin tumours. In the future, this work may expand to handle high-dimensional data and assess its performance across additional categorisation tasks. In the end, this study emphasises the importance of developing flexible algorithms that can handle diverse data-processing challenges.

Acknowledgement: The authors express their sincere gratitude to New Horizon College of Engineering, Karunya Institute of Technology and Science, Saranathan College of Engineering, St. Joseph's College of Engineering, and the Placental Histotherapy Center for their valuable support and collaboration.

Data Availability Statement: The data supporting the findings of this study are available from the corresponding author upon reasonable request.

Funding Statement: This study and the preparation of this manuscript were carried out without any external financial assistance or funding support.

Conflicts of Interest Statement: The authors declare that they have no known conflicts of interest. All information, citations, and references included in this work are original contributions from the authors.

Ethics and Consent Statement: This research was conducted in accordance with established ethical standards, and informed consent was obtained from all participants.

References

1. World Health Organization, “Radiation: Ultraviolet (UV) Radiation and Skin Cancer | How Common Is Skin Cancer,” *WHO*, 2023. Available: https://www.who.int/news-room/q-a-detail/radiation-ultraviolet-(uv)-radiation-and-skin-cancer# [Accessed by 02/10/2024].
2. F. Piccialli, V. Di Somma, F. Giampaolo, S. Cuomo, and G. Fortino, “A survey on deep learning in medicine: Why, how and when?” *Information Fusion*, vol. 66, no. 1, pp. 111–137, 2021.
3. R. Navid, M. Ashourian, M. Karimifard, V. V. Estrela, H. J. Loschi, D. D. Nascimento, R. P. França, and M. Vishnevski, “Computer-aided diagnosis of skin cancer: A review,” *Current Medical Imaging*, vol. 16, no. 7, pp. 781–793, 2020.
4. R. Ashraf, S. Afzal, A. U. Rehman, S. Gul, J. Baber, M. Bakhtyar, I. Mehmood, O.-Y. Song, and M. Maqsood, “Region-of-interest based transfer learning assisted framework for skin cancer detection,” *IEEE Access*, vol. 8, no. 8, pp. 147858–147871, 2020.
5. Y. Zheng, H. Liang, Z. Li, M. Tang, and L. Song, “Skin microbiome in sensitive skin: The decrease of Staphylococcus epidermidis seems to be related to female lactic acid sting test sensitive skin,” *Journal of Dermatological Science*, vol. 97, no. 3, pp. 225–228, 2020.
6. S. S. Chaturvedi, K. Gupta, and P. S. Prasad, “Skin lesion analyser: An efficient seven-way multi-class skin cancer classification using MobileNet,” *in Proc. Int. Conf. Advanced Machine Learning Technologies and Applications*, Jaipur, India, 2020.
7. M. Q. Khan, A. Hussain, S. U. Rehman, U. Khan, M. Maqsood, K. Mehmood, and M. A. Khan, “Classification of melanoma and nevus in digital images for diagnosis of skin cancer,” *IEEE Access*, vol. 7, no. 7, pp. 90132–90144, 2019.
8. J. V. Tembhurne, N. Hebbar, H. Y. Patil, and T. Diwan, “Skin cancer detection using ensemble of machine learning and deep learning techniques,” *Multimedia Tools and Applications*, vol. 82, no. 18, pp. 27501–27524, 2023.
9. D. Keerthana, V. Venugopal, M. K. Nath, and M. Mishra, “Hybrid convolutional neural networks with SVM classifier for classification of skin cancer,” *Biomedical Engineering Advances*, vol. 5, no. 6, p. 100069, 2023.

10. S. Q. Gilani, T. Syed, M. Umair, and O. Marques, “Skin cancer classification using deep spiking neural network,” *Journal of Digital Imaging*, vol. 36, no. 3, pp. 1137–1147, 2023.
11. H. Gururaj, N. Manju, A. Nagarjun, V. N. M. Aradhya, and F. Flammini, “DeepSkin: A deep learning approach for skin cancer classification,” *IEEE Access*, vol. 11, no. 5, pp. 50205–50214, 2023.
12. S. M. Thomas, J. G. Lefevre, G. Baxter, and N. A. Hamilton, “Interpretable deep learning systems for multi-class segmentation and classification of non-melanoma skin cancer,” *Medical Image Analysis*, vol. 68, no. 2, p. 101915, 2021.
13. M. A. Al-Masni, D.-H. Kim, and T.-S. Kim, “Multiple skin lesions diagnostics via integrated deep convolutional networks for segmentation and classification,” *Computer Methods and Programs in Biomedicine*, vol. 190, no. 7, p. 105351, 2020.
14. A. G. Pacheco, A.-R. Ali, and T. Trappenberg, “Skin cancer detection based on deep learning and entropy to detect outlier samples,” *arXiv preprint*, 2019. Available: https://arxiv.org/pdf/1909.04525 [Accessed by 05/10/2024].
15. M. Tahir, A. Naeem, H. Malik, J. Tanveer, R. A. Naqvi, and S. W. Lee, “DSCC_Net: Multi-classification deep learning models for diagnosing skin cancer using dermoscopic images,” *Cancers*, vol. 15, no. 7, p. 2179, 2023.
16. L. Ichim, R. I. Mitrica, M. O. Serghei, and D. Popescu, “Detection of malignant skin lesions based on decision fusion of ensembles of neural networks,” *Cancers*, vol. 15, no. 20, p. 4946, 2023.
17. A. R. Baig, Q. Abbas, R. Almakki, M. E. Ibrahim, L. AlSuwaidan, and A. E. Ahmed, “Light-Dermo: A lightweight pretrained convolution neural network for the diagnosis of multiclass skin lesions,” *Diagnostics*, vol. 13, no. 3, p. 385, 2023.
18. Q. Abbas, Y. Daadaa, U. Rashid, and M. E. Ibrahim, “Assist-Dermo: A lightweight separable vision transformer model for multiclass skin lesion classification,” *Diagnostics*, vol. 13, no. 15, p. 2531, 2023.
19. I. A. Ahmed, E. M. Senan, H. S. A. Shatnawi, Z. M. Alkhraisha, and M. M. A. Al-Azzam, “Multi-models of analyzing dermoscopy images for early detection of multi-class skin lesions based on fused features,” *Processes*, vol. 11, no. 3, p. 910, 2023.
20. H. El-Khatib, A. M. Ştefan, and D. Popescu, “Performance improvement of melanoma detection using a multi-network system based on decision fusion,” *Applied Sciences*, vol. 13, no. 18, p. 10536, 2023.
21. Q. E. Wu, Y. Yu, and X. Zhang, “A skin cancer classification method based on discrete wavelet down-sampling feature reconstruction,” *Electronics*, vol. 12, no. 9, p. 2103, 2023.
22. J. Liutkus, A. Kriukas, D. Stragyte, E. Mazeika, V. Raudonis, W. Galetzka, A. Stang, and S. Valiukeviciene, “ACC of a smartphone-based artificial intelligence application for classification of melanomas, melanocytic nevi, and seborrheic keratoses,” *Diagnostics*, vol. 13, no. 13, p. 2139, 2023.
23. The ISIC 2020 Challenge Dataset, “ISIC 2020 challenge dataset,” 2020. Available: https://challenge2020.isic-archive.com/ [Accessed by 02/10/ 2024].
24. S. Bibi, M. A. Khan, J. H. Shah, R. Damaševičius, A. Alasiry, M. Marzougui, M. Alhaisoni, and A. Masood, “MSRNet: Multiclass skin lesion recognition using additional residual block-based fine-tuned deep models information fusion and best feature selection,” *Diagnostics*, vol. 13, no. 19, p. 3063, 2023.
25. J. M. J. Valanarasu and V. M. Patel, “Unext: MLP-based rapid medical image segmentation network,” in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2022: Proc. 25th Int. Conf.*, Springer, Singapore, 2022.
26. L. Zhu, X. Wang, Z. Ke, W. Zhang, and R. W. Lau, “BiFormer: Vision transformer with bi-level routing attention,” in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, Vancouver, British Columbia, Canada, 2023.
27. K. Vijayalakshmi, S. Al-Otaibi, L. Arya, M. A. Almaiah, T. P. Anithaashri, S. S. Karthik, and R. Shishakly, “Smart agricultural–industrial crop-monitoring system using unmanned aerial vehicle–Internet of Things classification techniques,” *Sustainability*, vol. 15, no. 14, p. 11242, 2023.
28. D. S. A. Elminaam, A. Nabil, S. A. Ibraheem, and E. H. Houssein, “An efficient marine predators’ algorithm for feature selection,” *IEEE Access*, vol. 9, no. 4, pp. 60136–60153, 2021.
29. M. A. A. Al-Qaness, A. A. Ewees, H. Fan, L. Abualigah, and M. Abd Elaziz, “Marine predators algorithm for forecasting confirmed cases of COVID-19 in Italy, USA, Iran and Korea,” *International Journal of Environmental Research and Public Health*, vol. 17, no. 10, p. 3520, 2020.
30. J. Too, G. Liang, and H. Chen, “Memory-based Harris hawk optimization with learning agents: A feature selection approach,” *Engineering Computations*, vol. 38, no. S5, pp. 4457–4478, 2022.