Computer Modeling in Engineering & Sciences

Tech Science Press

Doi:10.32604/cmes.2025.064079

ARTICLE

BioSkinNet: A Bio-Inspired Feature-Selection Framework for Skin Lesion Classification

Tallha Akram[1,*], **Fahdah Almarshad**[1], **Anas Alsuhaibani**[1] and **Syed Rameez Naqvi**[2,3]

[1]Department of Information Systems, College of Computer Engineering and Sciences, Prince Sattam bin Abdulaziz University, Al-Kharj, 11942, Saudi Arabia
[2]Department of Electrical and Computer Engineering, COMSATS Univeristy Islamabad, Wah Campus, WahCantt, 47010, Pakistan
[3]Department of Computer Science, Tulane University, New Orleans, LA 70118, USA

*Corresponding Author: Tallha Akram. Email: t.akram@psau.edu.sa

Received: 04 February 2025; Accepted: 08 April 2025; Published: 30 May 2025

ABSTRACT: Melanoma is the deadliest form of skin cancer, with an increasing incidence over recent years. Over the past decade, researchers have recognized the potential of computer vision algorithms to aid in the early diagnosis of melanoma. As a result, a number of works have been dedicated to developing efficient machine learning models for its accurate classification; still, there remains a large window for improvement necessitating further research efforts. Limitations of the existing methods include lower accuracy and high computational complexity, which may be addressed by identifying and selecting the most discriminative features to improve classification accuracy. In this work, we apply transfer learning to a Nasnet-Mobile CNN model to extract deep features and augment it with a novel nature-inspired feature selection algorithm called *Mutated Binary Artificial Bee Colony*. The selected features are fed to multiple classifiers for final classification. We use PH^2, ISIC-2016, and HAM10000 datasets for experimentation, supported by Monte Carlo simulations for thoroughly evaluating the proposed feature selection mechanism. We carry out a detailed comparison with various benchmark works in terms of convergence rate, accuracy histogram, and reduction percentage histogram, where our method reports 99.15% (2-class) and 97.5% (3-class) accuracy on the PH^2 dataset, while 96.12% and 94.1% accuracy for the other two datasets, respectively, against minimal features.

KEYWORDS: Skin lesion classification; CNN; transfer learning; artificial bee colony; entropy-controlled; bio-inspired; computer-aided diangosis (CAD)

1 Introduction

Cancer is a pathological condition characterized by aberrant proliferation of cells in many anatomical locations within the human body [1]. Skin is a vital organ in the human body that includes all the muscular, skeletal, and visceral components. The fundamental role of the skin is to protect the body from external factors, including chemicals, pathogens, and temperature variations [2]. Naturally, a skin disease, such as skin cancer, will have a consequential impact on all other bodily organs. Several prominent risk factors for skin cancer include genetic susceptibility, fair skin, repeated sunburns, and other medical conditions [3]. Although skin cancer continues to increase worldwide as well, studies suggest that it is the most prevalent kind of cancer in the US [4]. Recent studies [5] suggest that around 9500 people are diagnosed with general skin cancers every day, where more than 2 die every hour only in the US. Similarly, another study by the American Cancer Society [6] reveals that 104,960 people are projected to be diagnosed only with melanoma in 2025, leading to an alarming number of 8430 deaths. The situation, unfortunately, isn't encouraging in the

rest of the world either, especially in Australia and Scandinavia [7]. Due to its rapid growth and metastasis, melanoma is the most severe form of cancer [8]; however, it is curable with early detection and treatment [9]. Malignant and benign examples from various melanoma datasets are shown in Fig. 1.

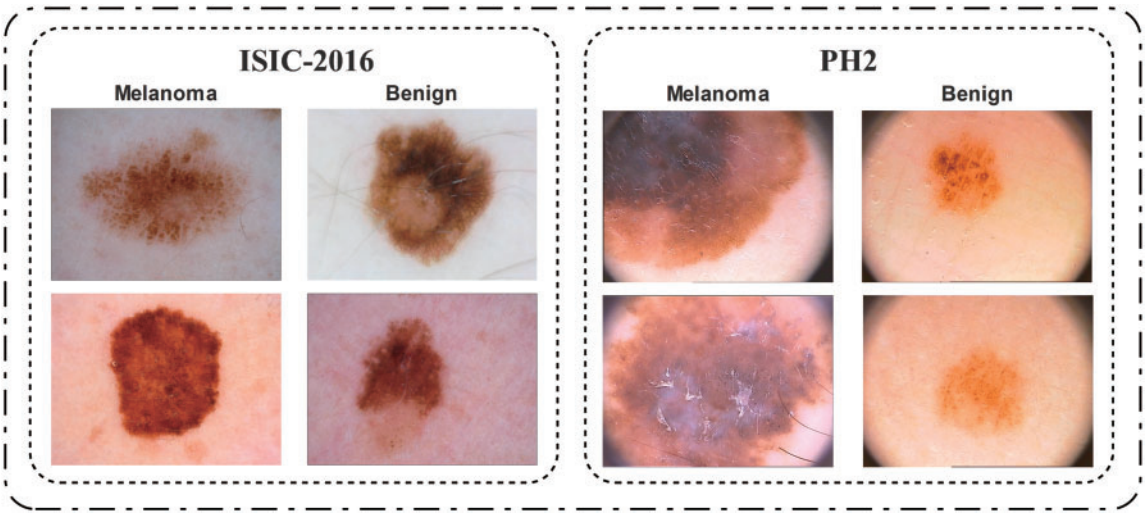

Figure 1: A few image samples from the selected datasets of ISIC-2016 and PH^2-illustrating melanoma and benign classes

Considerable progress has been made by numerous researchers in the field of computer vision for skin cancer detection [4,10–13].

One significant challenge associated with this approach is the substantial data requirement for model training. The problem has been successfully resolved by utilizing transfer learning (TL) [14]. Several researchers utilized multiple pre-trained models to categorize computer vision and medical imaging tasks. These models include VGG [15], GoogleNet [16], ResNet [17], and Densnet201 [18].

Despite the great strides taken thus far, a lot more work is needed to fix the problem with the detection and classification phases. Given the significance of feature selection for improving the model's overall accuracy and decreasing computation time, this research focuses primarily on the feature selection process as a strategy to alleviate the challenge referred to as the *curse of dimensionality*. Given the increasing recognition of bio-inspired techniques in the field of feature selection, we present a unique approach known as the mutated binary artificial bee colony (MBABC) algorithm. The primary advantage of employing MBABC compared to traditional methods, such as particle swarm optimization (PSO), genetic algorithm (GA), grey wolf optimization (GWO), and BMNABC, resides in its improved exploration and exploitation balance, broader search space, and its ability to adjust the selection strategy. Moreover, the implementation of the entropy function for fitness evaluation has necessitated adaptations in the employed bee phase. This extension not only enhances diversity but also facilitates fast convergence and reduces the likelihood of being stuck in local minima.

The following is the logical progression of this article: The current approaches are discussed in Section 2, and then the problem statement and contributions are presented in Section 3. Section 4 will focus on the datasets and models used in the study, followed by the proposed section. Section 5 discusses the results, followed by Section 6, which concludes our work.

2 Literature Review

Skin cancer classification is a well-researched field, with numerous contributions, directly towards deep learning models and feature selection methods, proposed in the past few years. Bassel et al. [4] introduced an automated technique for classifying benign and malignant skin cancer by utilizing a hybrid deep learning approach. They utilized several pre-trained deep models, such as Xception, ResNet50, and VGG16, to perform feature extraction on the ISIC-2019 dataset. Afterwards, the deep features wee passed to the classifiers, that included support vector machines (SVM), random forest (RF), neural networks (NN), and K-nearest neighbors (KNN). To evaluate the effectiveness of the proposed framework, various performance metrics such as F1 score, accuracy, AUC, and sensitivity were employed. The proposed method achieved an accuracy of around 90.9% when applied to the extracted features using the Xception model.

Ali et al. presented a multi-class classification of skin cancer by applying pre-processing and transfer learning techniques [11]. During the preprocessing stage, the images from HAM10000 dataset were scaled and augmented, and the hair were removed using existing techniques. They evaluated the effectiveness of different variants of EfficientNet. To accomplish this, they applied transfer learning to Efficientnet B0-B7 nets. The evaluation criteria used in this article were recall, accuracy, and F1 score. Among all, EfficientNet models, B4 performed well with an accuracy of 87.91% and an F1 score of 87%.

In their study, Bechelli et al. [12] employed a combination of conventional and deep learning approaches to classify skin cancer from dermoscopic images. They used multiple machine learning algorithms, including logistic regression (LR), linear discriminant analysis (LDA), decision tree classifiers (DTC), k-nearest neighbors classifiers (KNN), and Gaussian Naive bayes (GNB). Additionally, they utilized their own 11-layered custom convolutional neural network (CNN) model, as well as pre- models such as ResNet50, Xception, and VGG16. Their model achieved an accuracy rate of 84%, while among the conventional machine learning models, LR exhibited the highest performance with an accuracy rate of 72%. On the other hand, among the deep learning models, VGG16 outperformed the rest with an accuracy of 88%.

Chen et al. [19] proposed a multimodal data fusion diagnosis network (MDFNet) for the purpose of classifying skin cancer. The proposed approach extensively integrated feature fusion in conjunction with the attention mechanism. The proposed method consisted of three main steps: 1. Construction of a feature extraction mechanism based on two modes; 2. Implementation of attention mechanisms to effectively handle multimodal features; and 3. Integration of multimodal characteristics through fusion. Various models, including ResNet50, VGGNet19, Inception-V3, and DenseNet121, were employed for feature extraction. The developed approach achieved an accuracy of 80.42%, exhibiting a 9% improvement in comparison to models solely utilizing medical images.

Hosny et al. [20] proposed a refined residual deep convolutional neural network for skin cancer classification. This research work utilized six skin cancer datasets: ISIC (2016, 2017, 2018), MED-NODE, DermIS, Quest, and PH^2 and considered a set of experiments for evaluating the proposed model. Initially, the proposed residual deep convolutional neural network (RDCNN) was trained and tested on the dataset without undergoing any preprocessing. Subsequently, during the second experimentation, the model underwent testing on the segmented images. Finally, the model obtained from the second experimentation was employed as a pre-trained model for final classification. The results demonstrated that their proposed RDCNN model surpassed other models and achieved the best accuracy, around 96.29%, on the ISIC 2017 dataset.

Tembhurne et al. [21] proposed a method for detecting skin cancer using a combination of traditional and modern machine learning approaches. The dataset used in this study was sourced from Kaggle, a publicly accessible platform, and consisted of processed images from the ISIC dataset. The developed

technique relied on four key steps: 1. Feature extraction; 2. Concatenation; 3. Dimensionality reduction; and 4. Classification. The features were initially retrieved using the contourlet transform (CT) and local binary pattern histogram (LBP). Subsequently, the extracted features from both algorithms were combined before applying the principal component analysis (PCA) for dimensionality reduction. The concatenated features were later classified using two classifiers: logistic regression (LR) and linear support vector machines (SVM). The proposed technique achieved a precision of 93%. Similarly, other researchers have also explored several different pre-trained CNN models, either solely or in conjunction with the other pre-trained models [22–25]. In Table 1, the existing techniques are summarized briefly on the basis of the methodologies proposed, the precision, and the datasets utilized.

Table 1: A few existing techniques

Year	Authors	Method	Highest accuracy (%)	Datasets
2025	Padhy et al. [22]	Deep R-LSTIM50 model by combining ResNet50 for feature extraction, and Long Short-Term Memory (LSTM) for temporal modeling	95.72%, 94.23%	ISIC 2020, HAM10000
2024	Bibi et al. [26]	Contrast enhancement, Feature extraction (DenseNet 53, 201), Fusion (Serial-harmonic mean), Selection (Marine predator optimization)	85.4, 98.80	ISIC 2018, 2019
2024	Chen et al. [19]	Multimode data fusion diagnosis network (MDFNet)	80.42	PAD-UFES-20
2023	Tahir et al. [27]	DSCC_Net	94.17	ISIC 2020, DermIS, and HAM10000
2023	Tembhurne et al. [21]	Feature extraction, Fusion, dimensionality reduction, and classification	93	Kaggle dataset
2023	Gilani et al. [28]	Spiking VGG-13	89.57	ISIC 2019
2023	Kibriya et al. [29]	SegNet	89	ISIC 2016
2022	Khan et al. [30]	Optimized Color Feature, Feature Extraction DCNN-9, Normal Parallel Fusion, Normal Distribution base High-ranking Feature Selection	92.1, 96.5, and 85.1	ISBI 2016,2017,2018
2022	Kaur et al. [31]	Lesion classification network (LCNet)	81.41, 88.23, and 90.42	ISIC 2016, ISIC 2017, and ISIC 2020
2022	Bassel et al. [4]	Transfer learning (Xception, ResNet50, and VGG16), Feature Extraction, and classification	Xception = 90.9	ISIC 2019
2022	Ali et al. [11]	Transfer learning (Efficientnet B0-B7), and classification	B4 = 87.91	HAM10000

(Continued)

Table 1 (continued)

Year	Authors	Method	Highest accuracy (%)	Datasets
2022	Bechelli et al. [12]	ML(LDA, LR, DTC, KNN, and GNB), DL (Custom NN, ResNet50, Xception and VGG16)	LR = 72, CCNN = 84, VGG16 = 88	HAM10000
2022	Saarela et al. [32]	Customized deep model	80	HAM10000
2022	Atta et al. [10]	Custom 15 layer deep model	89.4	Claudio Fanconi
2021	Srinivasu et al. [23]	Comparison of MobileNet-V2 and LSTM	85%	HAM10000
2020	Khan et al. [33]	Contrast Stretching, Boundary Localization, Transfer Learning DenseNet, Feature Extraction, Feature Selection Newton Raphson	94.5, 93.4	ISBI 2016, 2017

Despite extensive work on skin lesion classification, several gaps remain. First, many methods strive to boost accuracy but pay limited attention to dimensionality reduction—leading to high feature redundancy and potential overfitting. Second, computational complexity is seldom discussed in detail, making it difficult to assess a model's practical feasibility. Third, bio-inspired optimization techniques, such as swarm and evolutionary strategies, are underutilized despite their potential to mitigate the "curse of dimensionality". Especially, the advanced metaheuristics remain underexplored for class-imbalanced datasets like HAM10000. Fourth, many approaches evaluate performance on a single or closely related dataset, leaving questions about cross-dataset generalization. Finally, although some studies handle multi-class problems, many focus primarily on binary classification, inhibiting broader clinical adoption.

3 Problem Statement & Contributions

Existing methods in automated skin lesion classification are limited in their practical efficacy due to numerous significant challenges that have not been responded: despite substantial research in the field. Mostly, several presented methods emphasize accuracy improvement without sufficient fixing of dimensionality reduction, leading to duplicated features and elevated overfitting risk. Additionally, computational complexity and efficiency are often overlooked, posing difficulties in real-world implementation, particularly in resource-constrained clinical environments. Although bio-inspired optimization methods, especially advanced metaheuristics, have shown the potential to resolve these challenges, their use remains constrained, mostly in class-imbalanced datasets. Ultimately, the majority of research primarily focuses on binary classification, overlooking the complex challenges present in multi-class clinical scenarios. Therefore, it is evident that a strong and computationally efficient feature selection method is necessary to overcome these limitations, thereby improving the reliability, scalability, and practical applicability of skin lesion classification systems.

In this study, we consider a labeled dataset $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^{N}$ of images $\mathbf{x}_i$ with class labels y_i. By passing each $\mathbf{x}_i$ through MobileNet, we obtain high-dimensional feature vectors $\mathbf{f}_i \in \mathbb{R}^D$, which together form the feature matrix $\mathbf{F} \in \mathbb{R}^{N \times D}$. We then define a binary selection vector $\mathbf{s} = (s_1, \ldots, s_D) \in \{0,1\}^D$, where

$s_j = 1$ indicates the j-th feature is chosen; thus each sample i is represented by the reduced vector $\tilde{\mathbf{f}}_i(\mathbf{s}) = (f_{i1}s_1, \ldots, f_{iD}s_D)$. To measure classification performance, we let $\mathcal{C}$ be a classifier trained on the selected features with average accuracy.

$$A(\mathbf{s}) = \frac{1}{N}\sum_{i=1}^{N}\mathbb{I}\big(y_i = \mathcal{C}\big(\tilde{\mathbf{f}}_i(\mathbf{s})\big)\big) \tag{1}$$

To balance accuracy with feature reduction, we solve

$$\max_{\mathbf{s}\in\{0,1\}^D}\Big[A(\mathbf{s}) - \lambda\frac{\|\mathbf{s}\|_1}{D}\Big] \tag{2}$$

where $\|\mathbf{s}\|_1$ counts the selected features, and $\lambda \geq 0$ is a regularization parameter. Our proposed algorithm efficiently tackles this combinatorial optimization, identifying an optimal (or near-optimal) feature subset $\mathbf{s}^*$, which is then used to train the final classifier.

The significant contributions of this work are as follows:

1. A novel bio-inspired metaheuristic algorithm, mutated binary artificial bee colony (MBABC), for feature selection is proposed to address the challenges posed by over-fitting, "*curse of dimensionality*", and computational cost. The core objective is to identify and ultimately choose the most discriminatory feature information for the final classification.
2. A thorough evaluation framework is developed that not only assesses robustness and generalization of the proposed feature selection method using multiple datasets, but also compares it with the baseline models using various metrics The results are compiled by conducting Monte-Carlo simulations for stability analysis.

4 Materials and Method

4.1 Dataset

Our proposed methodology in this research is assessed using three publicly accessible benchmark datasets: ISIC-2016 [34], PH^2 [35], and HAM10000 [36]. Both ISIC-2016 and PH^2 consist of two distinct categories, specifically malignant and benign. Whereas the HAM10000 dataset consists of 7 classes. The ISIC-2016 dataset comprises of 1145 images, PH^2 has 200 images, whereas HAM10000 has 10,015 images. The dataset's summary is presented in Table 2.

Table 2: Datasets summary

Benchmark dataset	Classes	Total images
ISIC-2016	Melanoma	232
	Benign	913
PH^2	Melanoma	40
	Nevus	80
	Atypical	80

(Continued)

Table 2 (continued)

Benchmark dataset	Classes	Total images
HAM10000	Actinic Keratosis (AKIEC)	327
	Basal Cell Carcinoma (BCC)	514
	Benign Keratosis (BKL)	1099
	Dermatofibroma (DF)	115
	Melanoma (MEL)	1113
	Melanocytic Nevi (NV)	6705
	Vascular Lesions (VASC)	142

4.2 Evaluation Matrics

To evaluate the performance of the proposed method comprehensively, several standard evaluation metrics were utilized. Initially, a confusion matrix was constructed to visualize the performance through the classification of predictions into four distinct categories: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) [37]. Based on these categories, the following performance measures are calculated:

1. **Accuracy** calculates the ratio of accurate predictions relative to the total number of predictions:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

2. **Precision** quantifies how precise the positive predictions are, defined as the ratio of correctly identified positive cases to all predicted positive cases:

$$Precision = \frac{TP}{TP + FP}$$

3. **Recall** (Sensitivity) evaluates the method's ability to identify positive instances correctly. High recall is especially critical in medical diagnosis to minimize the risk of missing potentially malignant lesions:

$$Sensitivity = \frac{TP}{TP + FN}$$

4. **Specificity** measures the proportion of correctly identified negative instances:

$$Specificity = \frac{TN}{TN + FP}$$

5. **F1-score** is the harmonic mean of precision and recall, providing a balanced measure especially useful when dealing with imbalanced datasets:

$$F1\text{–}score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

6. **Matthews Correlation Coefficient (MCC)** evaluates the quality of binary classifications, considering all four confusion matrix categories (TP, TN, FP, FN):

$$MCC = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

4.3 CNN

Convolutional Neural Network (CNN) is a specialized architecture that has been specifically designed for computer vision applications. These architectures have significantly transformed the domain of computer vision by facilitating the learning of complex visual patterns and automatic recognition, thus enabling the implementation of image segmentation, object detection, and classification, among other applications.

Nasnet-mobile

Construction of the Nasnet-Mobile convolutional neural network model utilizes the neural architecture search (NAS) technique. Neural architecture search is a procedure that autonomously looks for the optimal network architecture for a specified task. Designed specifically for mobile devices, this version of the NasNet architecture is more compact and efficient.

Compared to deep models such as ResNet and InceptionV3, NasNet-mobile is an exceptionally quick, shallow, and compact convolutional neural network model with 70 layers and 4.2 million parameters. The Nasnet-mobile model utilizes the Adam Optimizer in combination with the Rectified Linear Unit (ReLU) activation function. The comprehensive Nasnet-mobile architecture is illustrated in Table 3.

Table 3: Nasnet-mobile architecture

Layers	Input size	Output size
Conv, 3 × 3	224 × 224 × 3	112 × 112 × 32
MBConv1, 3 × 3	112 × 112 × 32	112 × 112 × 16
MBConv6, 3 × 3	112 × 112 × 16	56 × 56 × 24
MBConv6, 3 × 3	56 × 56 × 24	56 × 56 × 24
MBConv6, 5 × 5	56 × 56 × 24	28 × 28 × 40
MBConv6, 5 × 5	28 × 28 × 40	28 × 28 × 40
MBConv6, 3 × 3	28 × 28 × 40	28 × 28 × 80
MBConv6, 3 × 3	28 × 28 × 80	28 × 28 × 80
MBConv6, 3 × 3	28 × 28 × 80	28 × 28 × 80
MBConv6, 5 × 5	28 × 28 × 80	14 × 14 × 112
MBConv6, 5 × 5	14 × 14 × 112	14 × 14 × 112
MBConv6, 5 × 5	14 × 14 × 112	14 × 14 × 112
MBConv6, 5 × 5	14 × 14 × 112	7 × 7 × 192
MBConv6, 5 × 5	7 × 7 × 192	7 × 7 × 192
MBConv6, 5 × 5	7 × 7 × 192	7 × 7 × 192
MBConv6, 5 × 5	7 × 7 × 192	7 × 7 × 192
MBConv6, 3 × 3	7 × 7 × 192	7 × 7 × 320
Conv 1 × 1 and Pooling and FC	7 × 7 × 320	7 × 7 × 1280

4.4 Transfer Learning

A critical concern arises when confronted with the substantial volume of data needed for the CNN framework. Deep learning models demand a significantly greater volume of data for training purposes when compared to conventional machine learning models. Nevertheless, this issue has been resolved since the inception of transfer learning [38]. By utilizing transfer learning, it is feasible to train a deep learning model despite the limited availability of data. Fig. 2 demonstrates a visual illustration of transfer learning. In this

process, the final layers of the model are replaced with the new layers, and the weights in the previous layers remain frozen.

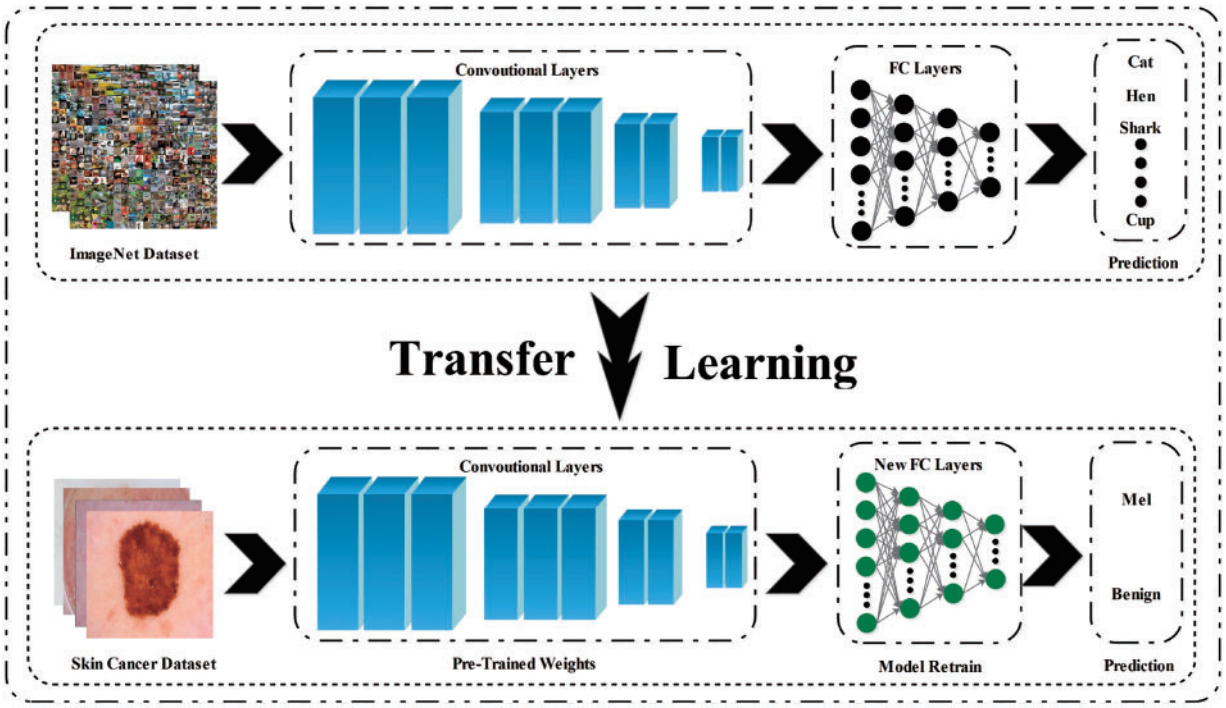

Figure 2: Transfer Learning: A CNN model trained on ImageNet is adapted for skin lesion classification by substituting and fine-tuning its final layers

The steps involved in transfer learning are listed below:

1. Selection of a pre-trained model P_θ, that is trained to perform tasks T_s, with a source domain S_D.
2. Change the fully-connected (FC) layer of the P_θ, and fine-tune it on your dataset.
3. The output model is fine-tuned on new data N_d, to perform new tasks T_t on target domain T_D.

4.5 Proposed Framework

The proposed framework in Fig. 3 encompasses a comprehensive pipeline that spans from the initial acquisition of images to the last stage of classification. The images obtained from the chosen databases are initially fed into the selected pre-trained model, followed by transfer learning application. This process involves substituting the last three layers with the newly initialized layers. By maintaining the frozen state of the existing weights in the previous layers, the model undergoes retraining on the chosen datasets, which then extracts the resulting features. It is crucial to acknowledge that the last layers possess the potential to provide redundant information. Consequently, it becomes necessary to employ a feature selection strategy to improve efficiency.

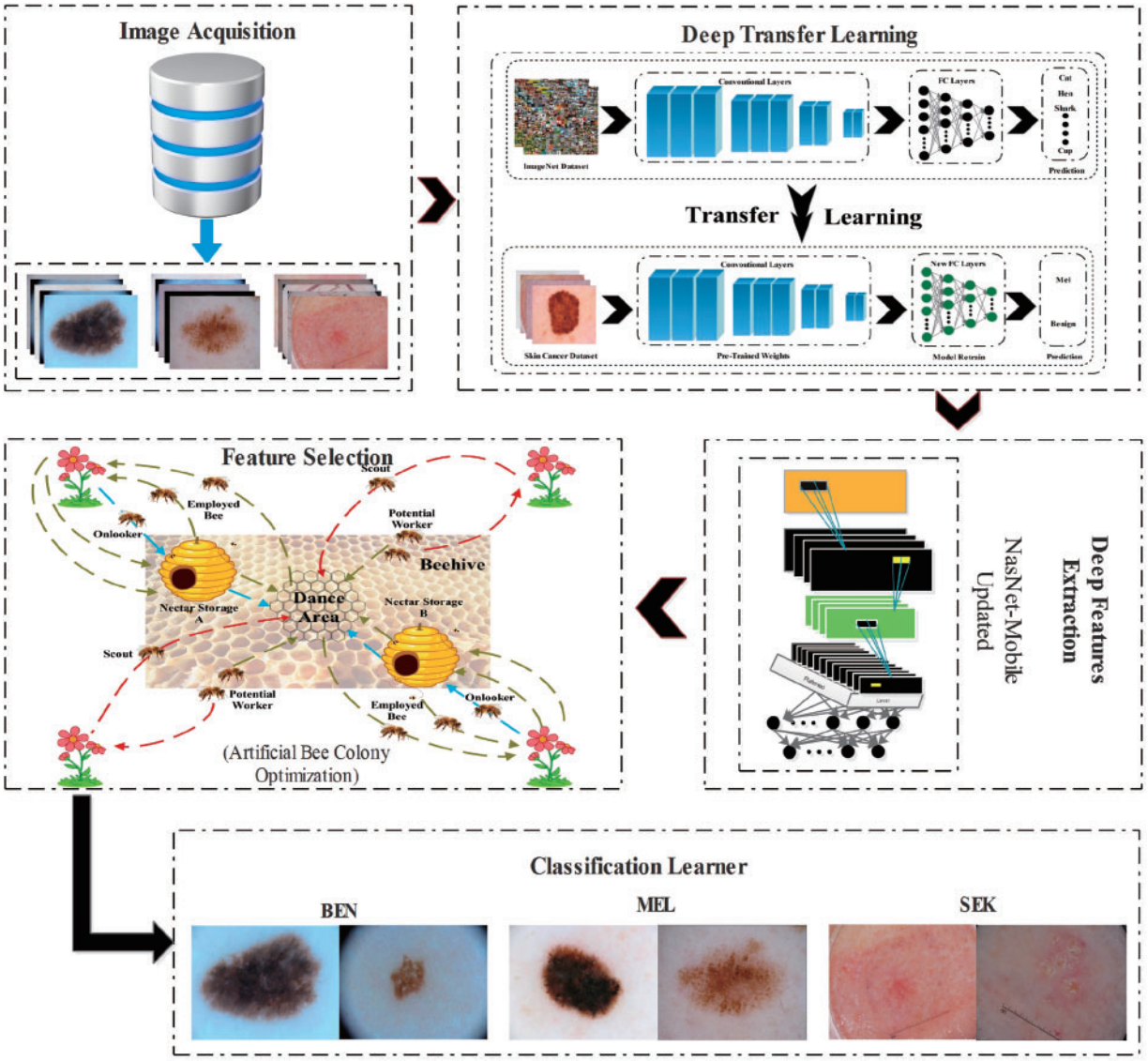

Figure 3: Schematic of the proposed BioSkinNet framework: From the image acquisition block to the final classification block

In the field of machine learning, and feature selection in particular, bio-inspired algorithms have recently emerged as key players. These algorithms draw inspiration from the principles of evolution and biological systems, and encompass a category of optimization methodologies that have been effectively employed in diverse sectors for problem-solving purposes [18]. In this research, we focus on the artificial bee colony (ABC) algorithm, which emulates the foraging behavior of honeybees to explore and improve the solutions. The ABC algorithm employs two distinct categories of bees: employed bees and observer bees. The foraging behavior involves employed bees actively exploring the solution space surrounding their present food source. On the other hand, the observer bees make decisions regarding food sources by relying on the information acquired from the employed bees. Bees engage in a dance-like technique to transmit the quality of their food sources to one another. As the algorithm advances, an increasing number of spectator bees opt for the most optimal food sources, thereby facilitating the exploration of novel and superior problem-solving strategies. The proposed method integrates an entropy-based fitness function with

a balanced exploration-exploitation mechanism. Moreover, with mutation-driven diversity, it often surpasses other feature selection methods. While many traditional algorithms (e.g., PSO, GA, GWO) risk early convergence or poor local minima in high-dimensional spaces, MBABC's added mutation phase that flips feature bits stochastically - allowing swarm to escape suboptimal regions. Simultaneously, its entropy-based fitness penalizes redundancy more precisely than just accuracy-based approaches by rewarding the most discriminative feature subsets. This combination consistently yields higher classification performance, more robust convergence, and better feature reduction than competing approaches, making MBABC well-suited to real-world tasks that demand both accuracy and efficiency.

4.5.1 Mutated Binary Artificial Bee Colony Algorithm (MBABC)

In order to optimize the mathematical operations, Karaboga [39] presented the ABC algorithm as a swarm intelligence method. It is based on how honey bee colonies hunt for food. The bee colony consists of three distinct types of bees, namely employed bees, onlooker bees, and scout bees.

In the initial phase, the employed bees engage in foraging activities to locate food sources. They communicate their findings to onlookers, who then evaluate the information provided by the employed bees to determine which food sources to exploit. The decision-making process is predicated on the employed bees' perception of whether a specific food source constitutes a geographically optimal solution. When a food source does not experience any advancements within a designated period of time, referred to as the abandonment counter (AC), the bee linked to that food source transitions into a scout bee. During the scout bees' phase, the bees look for more food sources using an algorithm called *AC* that operates with a threshold value set by the user.

The variables χ_m and l represent the random food source and dimension, respectively. Additionally, φ denotes a randomly produced number within the range of $[-1, +1]$. The ζ_{kl} is evaluated by following a greedy criterion. If the new solution ζ_k is improved compared to the existing solution χ_k, the employed bees update the new solution by resetting the counter *AC* to either zero or one. The fitness value is calculated based on an improved entropy-based criterion.

$$\mathbb{O}_k = -\sum_{p=1}^{n} \zeta_p \, log_2 \, \zeta_p \tag{3}$$

for the given objective function $\mathbb{O}$, the fitness is calculated based on predefined threshold value.

$$fit_k = \begin{cases} \mathbb{O}_k & if \, (\mathbb{O}_k > \tau) \\ \mathbb{O}_{k-1} & otherwise \end{cases} \tag{4}$$

4.5.2 Initialization Phase

During this phase, the value of *AC* is set to zero and a predetermined limit value is established. Moreover, within the predetermined boundaries, the food sources are designated at random.

$$\gamma_{kl} = \chi_l^{min} + rand(0,1)(\chi_l^{max} - \chi_l^{min}) \qquad k = 1 \ldots N_{food}, \quad l = 1 \ldots U_{dim} \tag{5}$$

The variable N_{food} signifies the overall number of food sources, whereas U_{dim} represents the potential dimensions, while U_{dim} denotes the potential dimensions. Similarly, χ_l^{min} and χ_l^{max} are the respective lower and upper bounds of the lth dimension.

$$\chi_{kl} = round(|\gamma_{kl} \mod 2|) \mod 2) \tag{6}$$

where χ_{kl} is the respective binary value.

4.5.3 Employed Bees Phase

The new solution is produced as follows with the improved solution at this phase:

$$\zeta_{kl} = \chi_{kl} + \omega_l \cdot \varphi(\chi_{kl} - \chi_{ml})$$
$$k,\ m \in \{1, 2, \ldots, N_{food}\},\ k \neq m$$
$$l \in \{1, 2, \ldots, U_{dim}\} \tag{7}$$

where τ is the user-defined threshold value, ω_l is the weighting criterion that may vary across dimensions*I*. to prioritize certain features during the update.

4.5.4 Onlooker Bees Phase

As the employed bees return to the hive, they impart knowledge to the observer bees concerning the food sources they have utilized. The observer bee selects its food source through the utilization of probabilistic information at its disposal, which leads to the discovery of the new solution denoted as ζ_k. The selection probability ρ_k is determined based on the improved fitness value of the solution.

$$\rho_k = \frac{\omega_k \cdot fit_k}{\sum_{l=1}^{N_{food}} \omega_l \cdot fit_l} \tag{8}$$

ω_k is introduced as a weight factor that prioritizes solutions based on criteria, not just fitness alone, and introduces more strategic selection by improving the diversity. If the observer bee's new solution is better, then an updated solution is considered. The value of *AC* is reset to zero if it is not already zero; else, the value of *AC* is incremented by 1.

4.5.5 Scout Bees Phase

In this phase, each food source's *AC* value is explored. If the count goes above the maximum allowed, the AC is set back to zero. Eq. (5) is used to generate a new supply of food, and the current solution is the best solution, that is, (χ_{kl} or g_{best}). Regarding computational complexity, for a D-dimensional search space, using a population of size *N*, over *T* iterations, the computational complexity of the phases (employer + onlooker + scout) altogether is dominated by $O(N \times D)$. Over *T* total iterations, the time complexity is $O(T \times N \times D)$.

Algorithm 1 explains the detailed flow of our proposed MBABC technique. The process initiates with a random initialization of a population of candidate solutions, where each solution is represented as a binary vector, indicating features that are either selected or discarded. Each iteration cycle employs three distinct bee phases: Employed bees perturb existing solutions to explore neighboring subsets, guided by a mutation operator to improve solution diversity. Subsequently, based on an entropy-controlled fitness criterion, onlooker bees probabilistically select the best possible solutions, ensuring a balanced exploration and exploitation of the search space. Scouts replace solutions stalling beyond a threshold level, therefore fostering diversity and lowering the chance of early convergence. The method balances exploration and exploitation until a stopping criterion—such as a convergence threshold or a maximum iteration limit—is fulfilled, so iteratively updating a global best solution.

Algorithm 1: Proposed feature selection algorithm (MBABC)

Initialization:

$\mathcal{N}, \mathcal{C}$, stopping criteria (max iterations or convergence threshold).

$\mathcal{AC}_k \leftarrow 0, \forall k \in \{1, \ldots, \mathcal{N}\}$.

$\chi_{kl} \sim \mathcal{U}(\chi_l^{\min}, \chi_l^{\max}), \quad \forall k \in \{1, \ldots, \mathcal{N}\}, \; l \in \{1, \ldots, U_{dim}\}$,

where $\mathcal{U}$ is the uniform distribution.

Convert χ_{kl} to binary using $f_{\text{binary}}(\chi_{kl})$.

$F(\chi_k) = \frac{\omega_k \cdot fit_k}{\sum_{j=1}^{\mathcal{N}} \omega_j \cdot fit_j}$, where $\omega_k = g(\chi_k)$.

$g_{\text{best}} = \arg\max_k F(\chi_k)$.

Main Iterative Process:

While TRUE:

Step 1: Employed Bee Phase

$\forall k \in \{1, \ldots, \mathcal{N}\}, \quad m \in \{1, \ldots, \mathcal{N}\} \backslash \{k\}$,

$\zeta_{kl} = \chi_{kl} + \omega_l \cdot \varphi(\chi_{kl} - \chi_{ml}), \quad \forall l \in \{1, \ldots, U_{dim}\}$,

where $\varphi \sim \mathcal{U}(-1, 1), \quad \omega_l = f(\chi_{kl}, \text{diversity})$.

$$F(\zeta_k) = \frac{\omega_k \cdot fit_k}{\sum_{j=1}^{\mathcal{N}} \omega_j \cdot fit_j}.$$

$$\chi_k = \begin{cases} \zeta_k, & \text{if } F(\zeta_k) > F(\chi_k), \\ \chi_k, & \text{otherwise.} \end{cases}$$

$$\mathcal{AC}_k = \begin{cases} \mathcal{AC}_k + 1, & \text{if } F(\zeta_k) \leq F(\chi_k), \\ 0, & \text{otherwise.} \end{cases}$$

Step 2: Onlooker Bee Phase (Updated)

$$\mathcal{P}_k = \frac{\omega_k \cdot F(\chi_k)}{\sum_{j=1}^{\mathcal{N}} \omega_j \cdot F(\chi_j)}, \quad \forall k \in \{1, \ldots, \mathcal{N}\}.$$

$\forall$ onlooker bees, sample χ_k based on $\mathcal{P}_k$.

$\mathcal{V}_{kl} = \chi_{kl} + \omega_l \cdot \varphi(\chi_{kl} - \chi_{ml}), \quad m \neq k, \; \forall l$.

$$F(\mathcal{V}_k) = \frac{\omega_k \cdot fit_k}{\sum_{j=1}^{\mathcal{N}} \omega_j \cdot fit_j}.$$

$$\chi_k = \begin{cases} \mathcal{V}_k, & \text{if } F(\mathcal{V}_k) > F(\chi_k), \\ \chi_k, & \text{otherwise.} \end{cases}$$

$$\mathcal{AC}_k = \begin{cases} \mathcal{AC}_k + 1, & \text{if } F(\mathcal{V}_k) \leq F(\chi_k), \\ 0, & \text{otherwise.} \end{cases}$$

Step 3: Scout Bee Phase (Updated Fitness)

If $\mathcal{AC}_k > \mathcal{C}, \quad \forall k$:

$\chi_{kl} \sim \mathcal{U}(\chi_l^{\min}, \chi_l^{\max}), \quad \forall l$.

$\mathcal{AC}_k \leftarrow 0$.

Update Global Best Solution:

$g_{\text{best}} = \arg\max_k F(\chi_k)$.

Output:

Best solution found: g_{best}.

5 Results and Discussion

The simulations are performed using three publicly available benchmark datasets, including PH^2, ISIC-2016, and HAM10000. Various families of classifiers, including k-nearest neighbor (KNN), ensemble (ES), support vector machines (SVM), neural network (NN), and linear discriminant (LD), are employed for classification due to their improved performance. This study primarily focuses on the proposed feature selection approach. Therefore, a comprehensive analysis has been provided for the readers. The ratio of training to testing is set at a fixed value of 80/20 for the simulations. Additionally, the k-fold cross-validation approach is selected for training validation. The default Matlab values are chosen for the selected deep models. Further, to validate the classification performance, a set of parameters is used, including accuracy, sensitivity, specificity, precision, negative prediction value (NPV), false discovery rate (FDR), false negative rate (FNR), F1-score, and Mathews correlation coefficient (MCC), respectively.

For the selected pre-trained deep model of Nasnet-mobile, the resultant feature vector is of dimension $(N \times 1056) \in \mathbb{R}^{(r,\ c)}$, where $N = [900,\ 160,\ 10000]$ for ISIC-2016, PH^2, and HAM10000, respectively. This research aims to down-sample the given dimension of each image. For a fair comparison, we selected the original feature selection algorithm [40]. Fig. 4 demonstrates the convergence plot of the proposed and the existing in terms of average error and number of iterations after applying Monte-Carlo simulations.

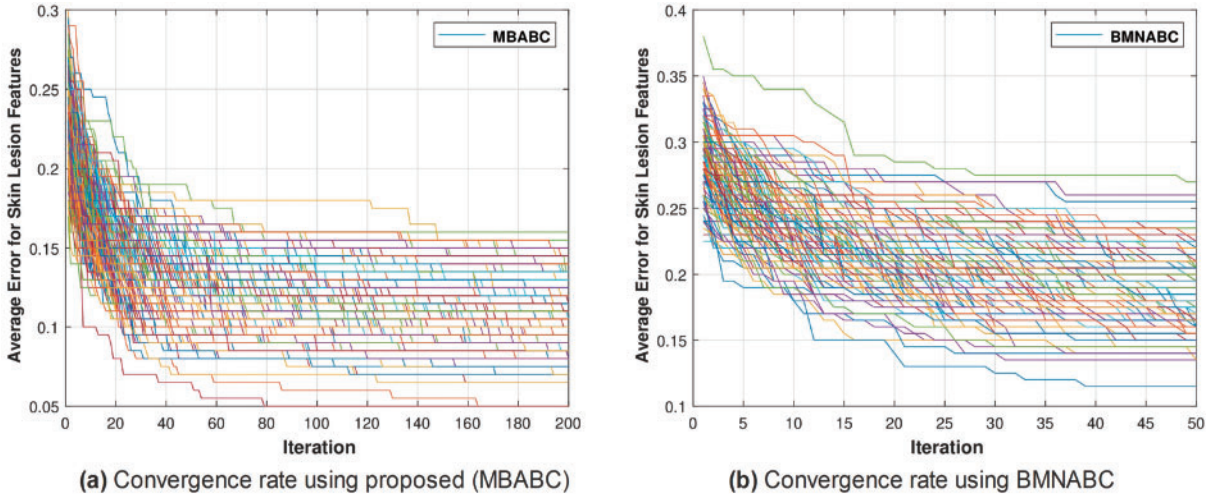

Figure 4: Comparison of MBABC vs. BMNABC convergence on ISIC-2016. Plot (a) illustrates faster convergence rate to a lower average error using proposed method, while plot (b) depicts the slower convergence rate of BMNABC

One could observe two main events: 1) a sudden drop in average error; and 2) a minimum average error. It is clear from the comparison that the proposed method has shown a sudden drop in average error compared to the existing method, as well as a greater number of times the algorithm touches the lowest value of less than 0.1. On the contrary, the existing method has not shown a sharp trend or a minimum average error. Similar trends have been observed in Fig. 5, where with the proposed feature selection algorithm, the classification framework performs better compared to the existing one. Fig. 5a has an accuracy range of (84%–96%) compared to (72–90) for the existing Fig. 5b.

Utilizing a combination of parameters to obtain a more comprehensive understanding of the model's performance is often recommended. As a result, we opted for the collection of performance parameters mentioned earlier. The majority of these parameters exhibit interdependence and collectively exert an impact on the average accuracy of the model. Accuracy is a measure that tests the overall correctness of a system or

model. Sensitivity, on the other hand, quantifies the ability of the system or model to correctly identify real positives. Specificity, in contrast, evaluates the system's or model's ability to avoid false positives. Precision is a metric that specifically reviews the accuracy of positive predictions made by a system or model. Lastly, NPV is a measure that assesses the accuracy of negative predictions made by the system or model. FDR is used to assess the fraction of inaccurate positive predictions, while FNR quantifies the rate of false negatives. The F1-Score is a metric that strikes a compromise between precision and sensitivity, giving preference to a model that exhibits both high precision and recall. In conclusion, MCC provides a comprehensive evaluation of the model's performance by taking into account both true and erroneous positives and negatives. A thorough examination of these metrics is crucial to improve a classification model for particular applications since they jointly provide insights into its accuracy and efficacy.

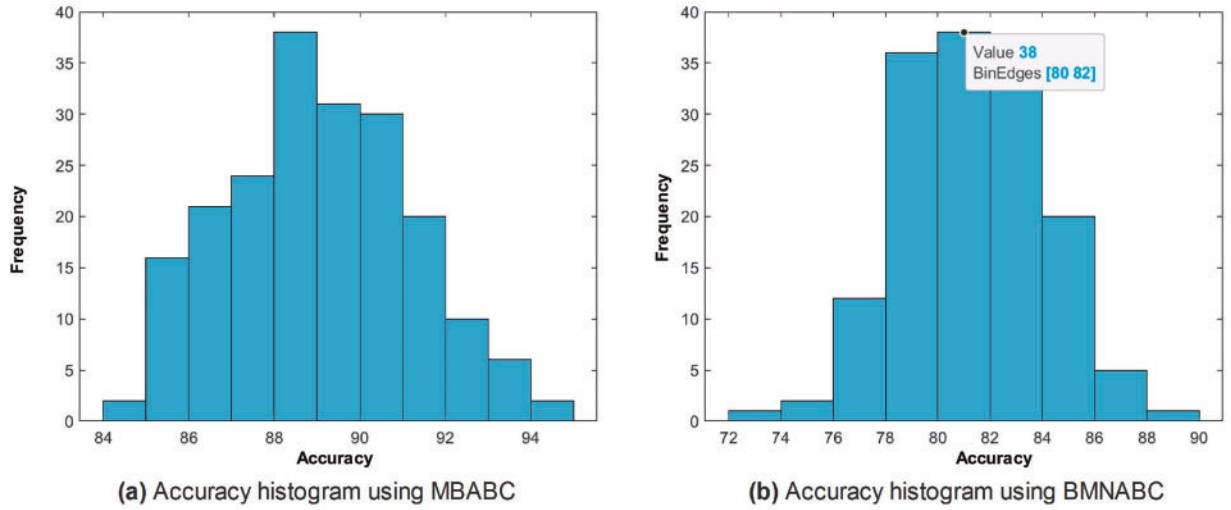

Figure 5: Comparison of accuracy histogram (%) on ISIC-2016 after Monte-Carlo simulations

In the case of the proposed, the frequency is maximum for an accuracy of (88%–90%), whereas in the case of the existing, the frequency is maximum within the range of (78%–84%). This is clear evidence of the performance of the proposed feature selection method.

Similar trends have been observed in the reduction plot in Fig. 6, where, after applying Monte-Carlo simulations, the histogram is plotted. It could be observed that in Fig. 6a, with the proposed reduction method, maximum features are reduced within the range of (70%–75%) percent, whereas on the other hand with the existing method, mostly the features are reduced within the range of (60%–75%). For a fair comparison, a confidence interval is plotted in terms of a box plot, where it could be seen that the proposed framework outperforms the existing method with a greater margin, whereas in terms of a reduction percentage, the difference is somewhat comparable, see Fig. 7a,b. For better understanding and to make well-informed decisions, the confusion matrix and receiver-operating characteristic curve (ROC) plots have been included in the analysis. These visualizations, depicted in Fig. 8, help to distinguish between positive and negative examples. The ISIC-2016 dataset presents TPR and FNR values that are inversely related: when TPR increases, FNR drops, and vice versa. The elevated true positive rate (TPR) values demonstrate the effectiveness of the proposed approach in accurately identifying positive events. Similarly, the low FNR value plainly implies that the model is successfully minimizing the amount of false negatives, which reveals that the model is good at finding positive cases. Similarly, the Area under the curve-Receiver-operating characteristic curve (AUC-ROC) score demonstrates superior discriminatory capability, as observed in our ISIC-2016 case.

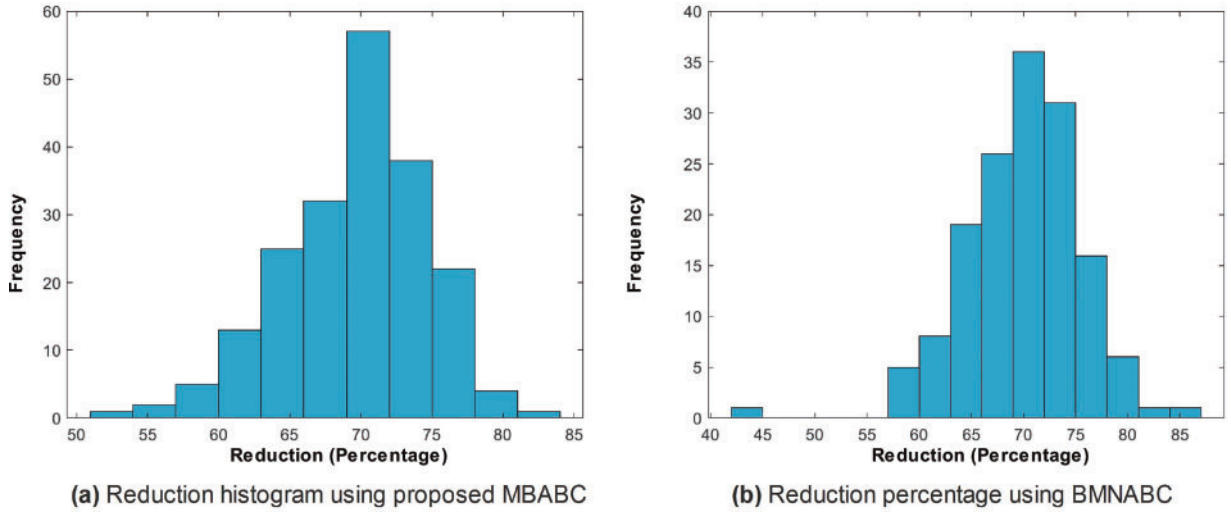

Figure 6: Comparison of feature reduction percentage histograms using proposed MBABC (a) vs. BMNABC (b) on ISIC-2016 dataset

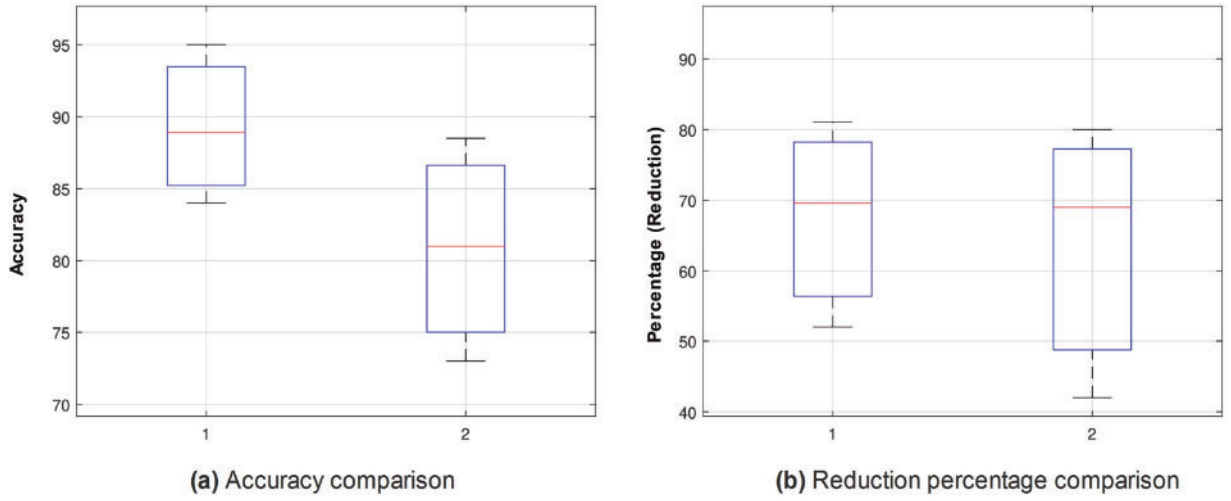

Figure 7: Box plots for accuracy (%) (a) and feature reduction percentage (b): Comparing MBABC (1) vs. BMNABC (2) on ISIC-2016 dataset

In the case of PH^2, we considered two possible scenarios, employing two classes and three classes, respectively. Initially, we conducted Monte-Carlo simulations for two classes and subsequently generated charts depicting the convergence, accuracy, and reduction rate. The behavior of the proposed feature selection method on PH^2 is quite similar to that of ISIC-2016.

Fig. 9 demonstrates the convergence of the proposed framework to that of the existing one. The convergence rate of the proposed framework clearly shows a sharp drop as well as a low average error in most of the cases. On the contrary, the existing method shows a few cases of zero average error. Considering the accuracy histogram, Fig. 10a,b, for the PH^2 dataset, both cases have achieved maximum accuracy. The difference is in the frequency and range, as with the proposed framework, the accuracy histogram reveals a range of (88%–100%) compared to the existing method, which shows a range of (86%–98.2%). Further,

most of the cases (bin value) with the proposed framework are in the range of (96%–100%), whereas for the existing cases, the range is (94%–98%).

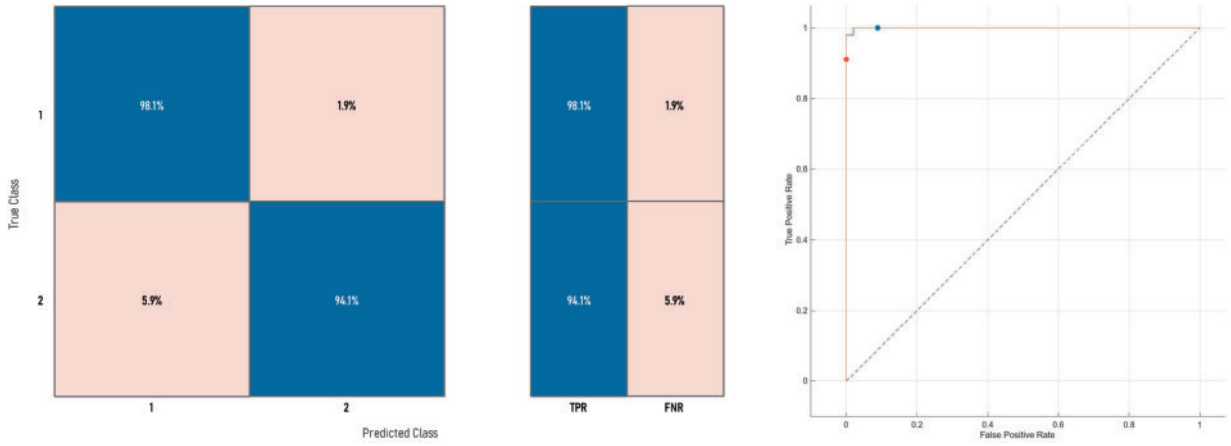

Figure 8: Confusion matrix and ROC curve of Medium-NN on ISIC-2016 dataset

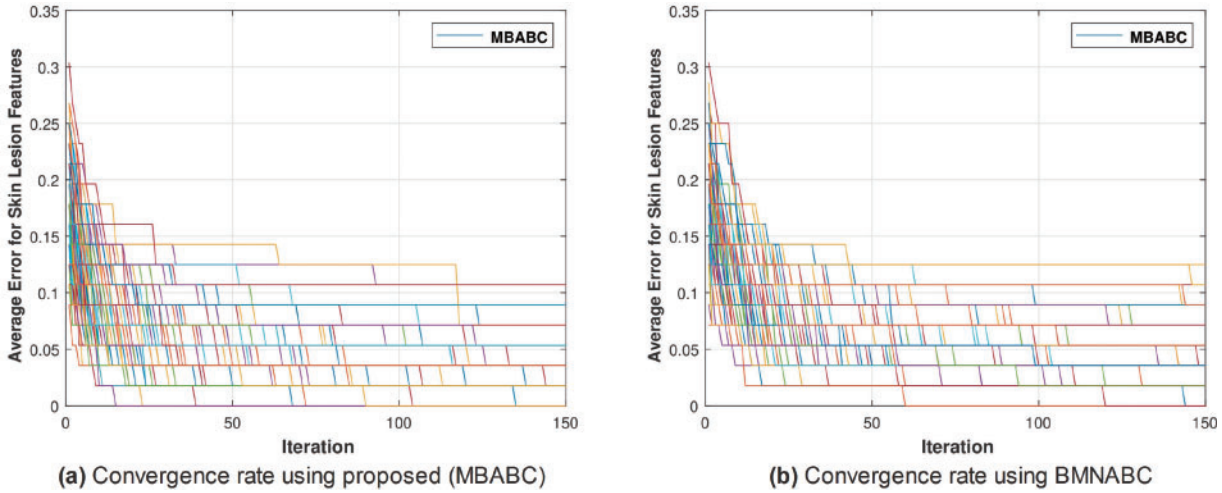

Figure 9: Comparison of MBABC vs. BMNABC convergence on PH^2 dataset. Plot (a) illustrates faster convergence rate to a lower average error using the proposed method, while plot (b) depicts the slower convergence rate of BMNABC

Similarly, Fig. 10c demonstrates a maximum reduction trend, as it reaches the limit of 98%. This shows, that even with only 2% of the original features, the framework is able to achieve maximum accuracy. Fig. 10c,d indicates the same pattern as that of ISIC-2016. With the proposed framework, the reduction percentage lies within the range of (85%–96%) compared to the existing method, which is in the range of (82%–96%).

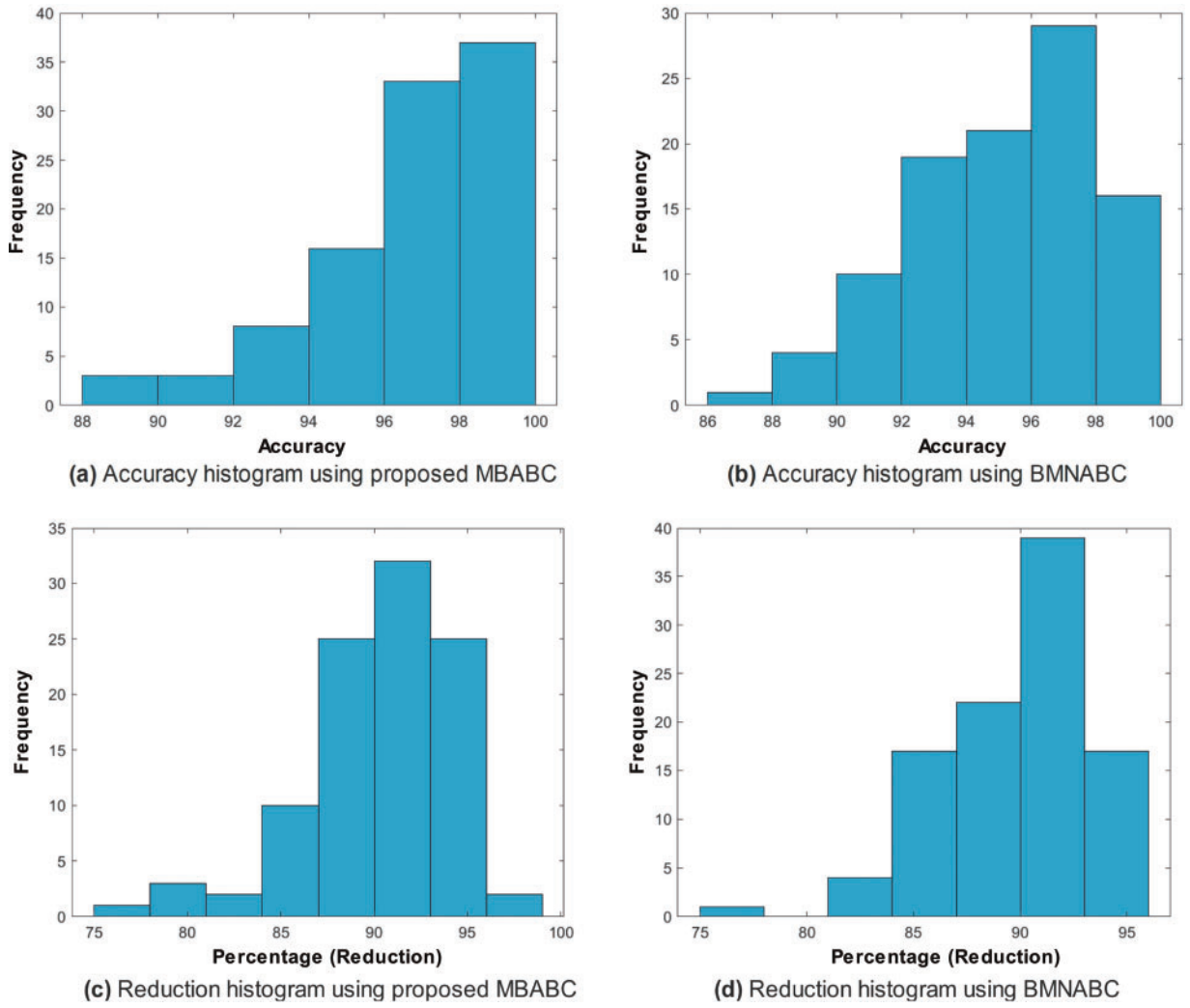

Figure 10: Top: comparison of accuracy (%) histograms on PH^2 dataset using the proposed BMABC (a) and BMNABC algorithms (b); Bottom: comparison of feature reduction percentage histograms using the proposed MBABC (c) vs. BMNABC algorithms (d) on PH^2 dataset

The box plot in Fig. 11 shows the confidence interval for both accuracy and reduction percentages. From both plots, one could easily observe the difference in accuracy as well as the reduction percentage.

Similarly, as in ISIC-2016, the confusion matrix and AUC-ROC for the PH^2 dataset for both configurations (2 & 3 classes) are also provided in Fig. 12. The elevated true positive rates (TPR) of 98.3% and 100% clearly demonstrate the effectiveness of the proposed approach in accurately identifying positive events for the PH^2 dataset. Similarly, the FNR value of 1.7% indicates that the model effectively reduces the occurrence of false negatives, demonstrating its ability to identify positive cases. Similarly, the AUR-ROC score for the PH^2 dataset clearly exhibits exceptional discriminatory capacity.

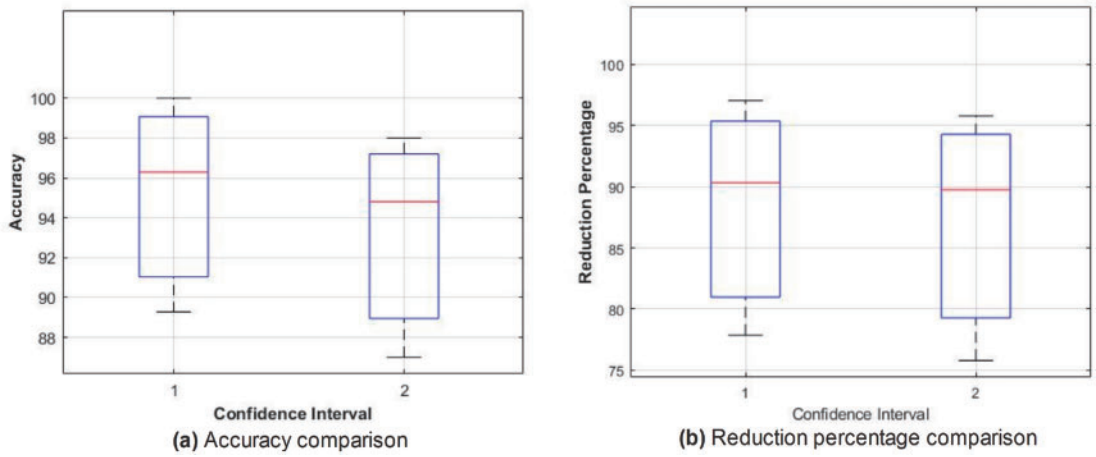

Figure 11: Box plots for accuracy (%) (a) and feature reduction percentage (b): Comparing MBABC (1) vs. BMNABC (2) on PH^2 dataset

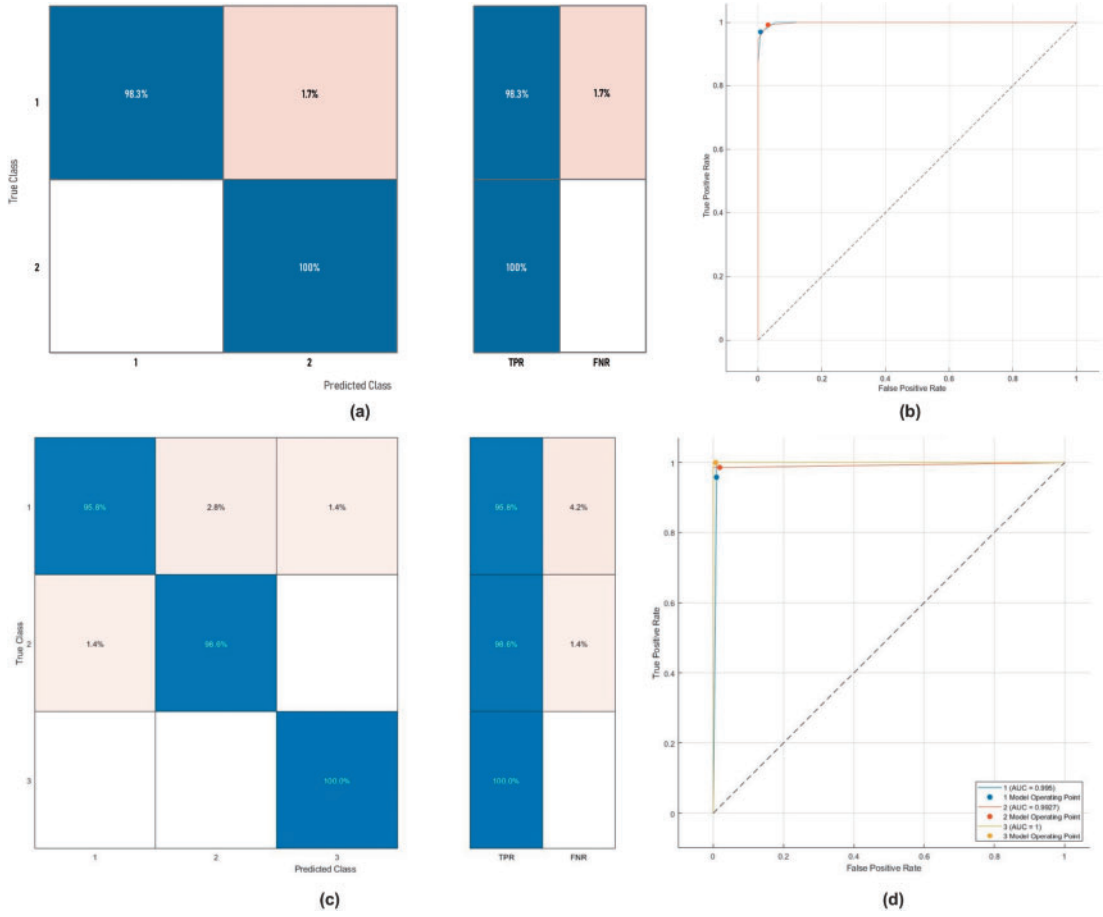

Figure 12: The ROC curve and confusion matrix of PH^2 dataset: two-class CM, two-class ROC; three-class CM, three-class ROC

In the case of HAM10000, after applying Monte-Carlo simulations, we generated convergence and accuracy plots. The proposed feature selection strategy exhibits distinct behavior when applied to the HAM10000 dataset compared to the ISIC-2016 and PH^2 datasets. Fig. 13 demonstrates the convergence of the proposed framework to that of the existing one on the HAM10000 dataset. Similarly, Fig. 14a–d indicates the difference in pattern as that of the other two datasets. In case of accuracy, Fig. 14a, the bounds are lower compared to the other datasets, as in this case, the range is (64%–76%) with the proposed model, whereas the range is between (63%–75%) for the baseline model (Fig. 14b). The pattern is also the same with the reduction percentage compared to the other datasets (Fig. 14c). Though, with the proposed method, the reduction percentage is up to 81%, with the peak between (72%–78%). On the contrary, with the baseline model, the reduction percentage is minimum (Fig. 14d). This clearly indicates that the proposed feature selection method works exceptionally well for a fewer number of classes, but for a larger number of classes, the reduction percentage is compromised.

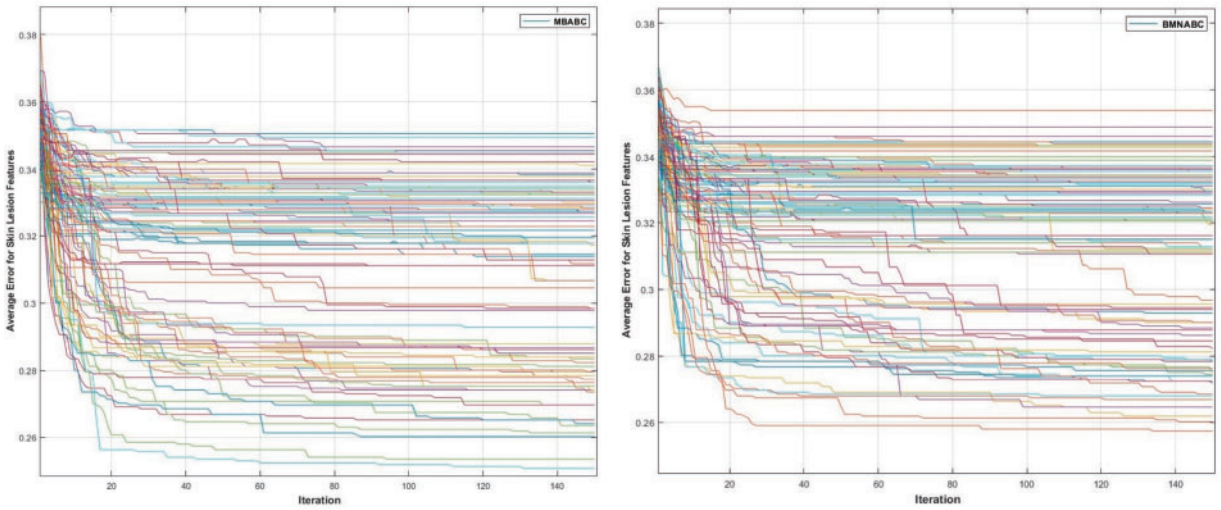

Figure 13: Comparison of MBABC vs. BMNABC convergence on HAM10000 dataset. Plot (a) illustrates faster convergence rate to a lower average error using the proposed method, while plot (b) depicts the slower convergence rate of BMNABC

Furthermore, Fig. 15a presents the confusion matrix for HAM10000. The high TPR values of 97.8% and 90.7% are considered acceptable for classes 6 (NV) and 3 (BKL). However, for classes 4 (Df), 7 (VASC), and 1 (AKIEC), the misclassification rates are very high, leading to a high FNR and lower TPR values. The TPR and FNR for the remaining classes are within acceptable ranges. An identical trend has been noted in the AUC-ROC plot, as depicted in Fig. 15b. The main factor contributing to the lower TPR and larger FNR for the given classes is the limited number of image samples. Specifically, there are 327 samples for AKIEC, 115 samples for Df, and 142 samples for VASC.

Table 4 shows the list of classifiers in a given order from highest accuracy to the lowest. Similarly, other parameters also show the variations as per the generated confusion matrix. It is quite clear that for both datasets, different classifiers have achieved maximum and minimum accuracies. In the case of the PH^2 dataset, quadratic SVM has shown maximum accuracy with the proposed framework, which is 99.15%, compared to other sets of classifiers. For the same classifier, other performance measures are also significant,

including sensitivity 0.983, specificity & precision (1.0), and FNR is 0.017. It could be observed that in the case of PH^2, the KNN family has shown the best performance compared to NN, SVM, and ensemble classifiers.

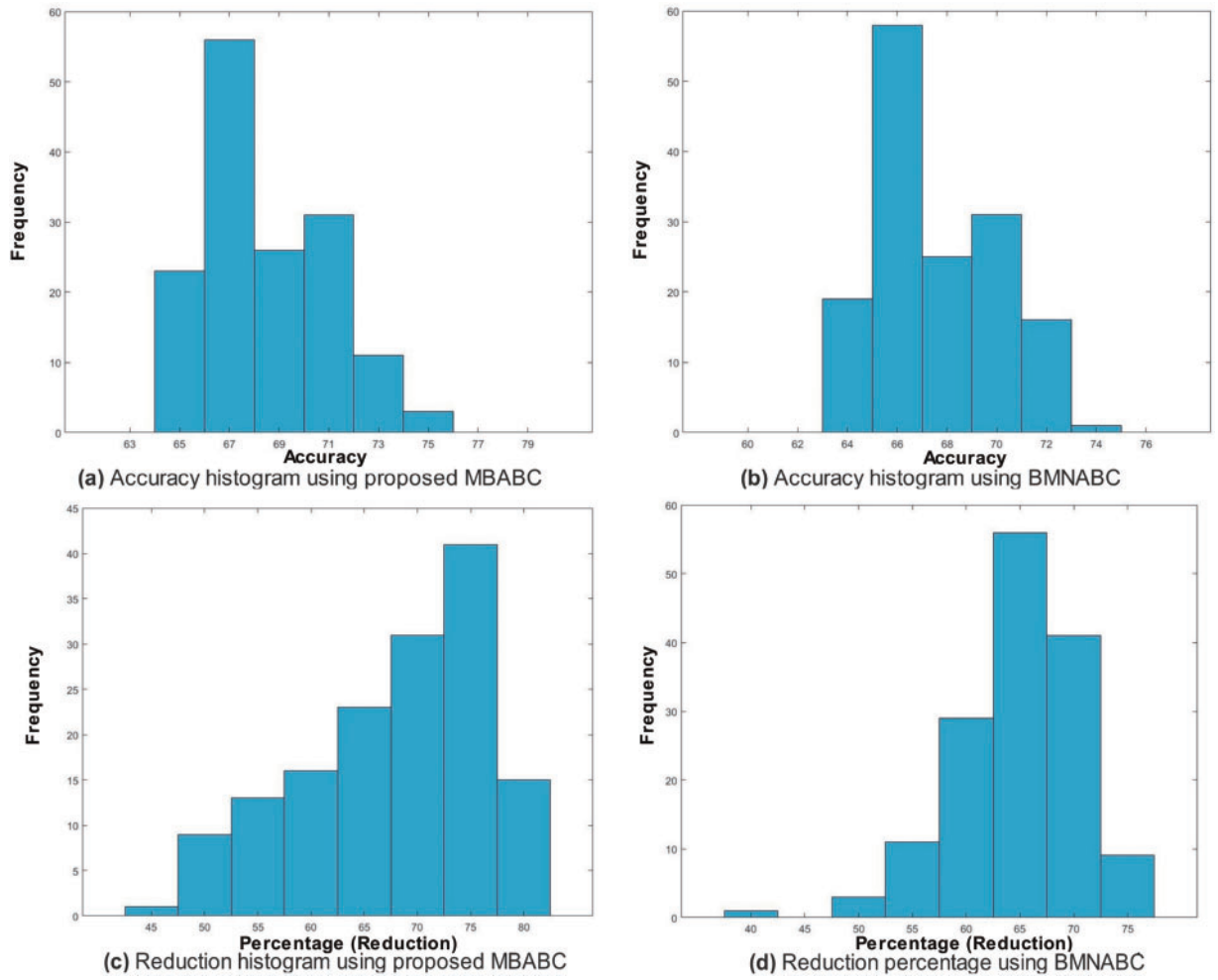

Figure 14: Top: comparison of accuracy (%) histograms on HAM10000 dataset using the proposed BMABC (a) and BMNABC algorithms (b); Bottom: comparison of feature reduction percentage histograms using the proposed MBABC (c) vs. BMNABC algorithms (d) on HAM10000 dataset

In this table, seven selected classifiers with improved classification accuracy are provided. The minimum accuracy achieved is 96.96% which is with medium KNN. On the contrary, in the case of ISIC-2016, the maximum accuracy achieved is 96.12% with medium NN. Other performance measures, including sensitivity is 0.971, specificity at 0.951, NPV at 0.97, and FNR at 0.029.

Regarding ISIC-2016, both the SVM and KNN families have exhibited excellent performance in comparison to other families, such as linear discriminant and ensemble. In the concluding instance of HAM10000, the ensemble of neural networks has exhibited excellent performance in comparison to alternative groups of classifiers. The results clearly demonstrate that the neural network family possesses the capability to achieve outstanding performance across several classes.

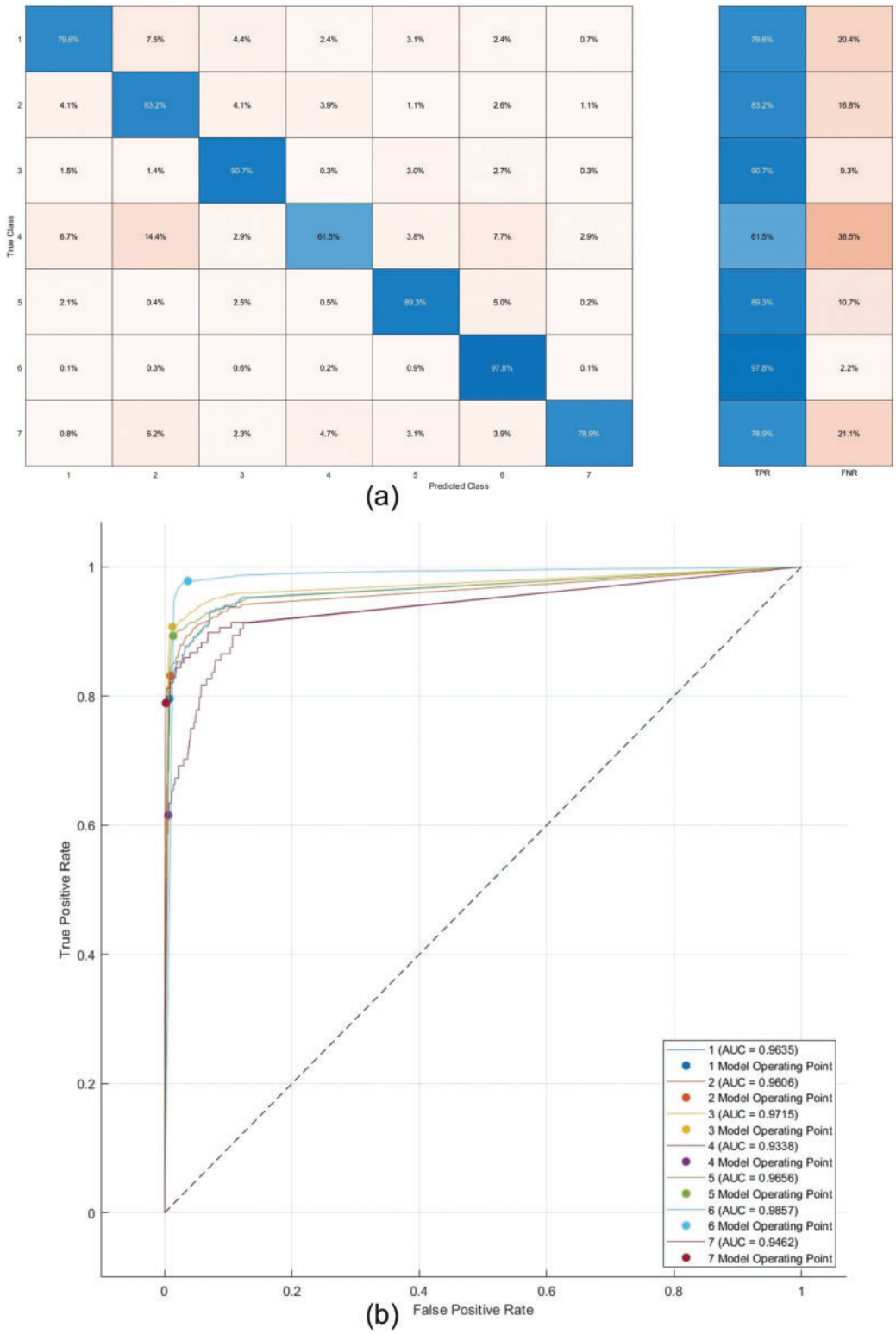

Figure 15: Confusion matrix and ROC of Trilayered-NN on HAM10000 dataset

Table 4: Performance measure of the several classifiers' families over benchmark datasets

	HAM10000								
	Acc (%)	**MCC**	**Spe**	**Sen**	**NPV**	**Pre**	**FNR**	**FDR**	**F1-Score**
Trilayered-NN	94.10	0.835	0.959	0.881	0.964	0.866	0.119	0.134	0.873
Logistic Regression Kernel	92.09	0.773	0.970	0.763	0.928	0.889	0.237	0.112	0.821
Coarse Gaussian SVM	90.70	0.736	0.961	0.737	0.921	0.856	0.263	0.144	0.790
Subspace-KNN	89.50	0.701	0.953	0.711	0.913	0.828	0.289	0.172	0.765
Fine-KNN	86.00	0.599	0.930	0.638	0.891	0.742	0.362	0.259	0.686
Weighted-KNN	85.10	0.573	0.924	0.619	0.885	0.720	0.381	0.280	0.666
Cubic-KNN	84.50	0.554	0.919	0.608	0.883	0.700	0.393	0.300	0.651
	ISIC-2016								
	Acc (%)	**MCC**	**Spe**	**Sen**	**NPV**	**Pre**	**FNR**	**FDR**	**F1-Score**
Medium-NN	96.12	0.921	0.971	0.951	0.951	0.970	0.049	0.029	0.961
Ensemble Subspace Discriminant (ESD)	95.40	0.908	0.968	0.940	0.941	0.967	0.058	0.032	0.954
Cubic-SVM	94.58	0.890	0.961	0.930	0.932	0.958	0.068	0.040	0.945
Quadratic-SVM	94.23	0.880	0.954	0.930	0.931	0.949	0.069	0.051	0.940
Cosine-KNN	93.01	0.861	0.947	0.920	0.921	0.938	0.078	0.060	0.930
Weighted-KNN	91.75	0.835	0.924	0.915	0.915	0.919	0.084	0.080	0.917
Linear Discriminant	90.50	0.810	0.901	0.910	0.909	0.090	0.091	0.904	
	PH^2 Dataset								
	Acc (%)	**MCC**	**Spe**	**Sen**	**NPV**	**Pre**	**FNR**	**FDR**	**F1-Score**
Quadratic-SVM	99.15	0.983	0.983	1.000	1.000	0.983	0.000	0.017	0.991
ESD	99.01	0.981	0.981	1.000	1.000	0.981	0.000	0.019	0.990
Cosine-KNN	98.62	0.972	0.981	0.990	0.989	0.981	0.010	0.019	0.985
Medium-NN	98.51	0.970	0.982	0.991	0.989	0.980	0.010	0.020	0.984
Weighted-KNN	98.40	0.969	0.979	0.990	0.989	0.979	0.010	0.021	0.984
Linear Discriminant	97.57	0.959	0.971	0.980	0.979	0.971	0.020	0.029	0.975
Medium-KNN	96.96	0.939	0.969	0.970	0.971	0.969	0.030	0.031	0.969

To ensure a fair evaluation of feature selection strategies in optimizing classification accuracy for this application, a comparison table is also included (Table 5). In contrast to established feature selection methods such as genetic algorithm (GA), particle swarm optimization (PSO), grey wolf optimization (GW), and BMNABC, the proposed method demonstrated remarkably high performance, as evidenced by its classification accuracy. As compared to alternative methodologies, GA achieves a higher average classification accuracy across all datasets. Undoubtedly, the genetic algorithm stands out as the optimal solution for feature selection in this specific application, surpassing other alternatives. Additionally, F1-score has been incorporated alongside accuracy metrics for skin lesion classification. The rationale for adding the F1-score is twofold: 1) HAM10000 skin lesion dataset exhibits significant class imbalance, where accuracy alone can mask poor performance on minority classes. The F1-score, which balances precision and recall, provides a more robust evaluation of model effectiveness in such scenarios; and 2) by reporting both metrics, we demonstrate that MBABC not only achieves higher accuracy but also maintains superior F1-scores compared to PSO and GWO. This emphasizes its ability to optimize feature selection for both majority and minority classes, a critical requirement in medical diagnostics. Table 6 presents a comparison of classification accuracy between a baseline method (without feature selection), and a proposed method for all selected datasets. The results clearly demonstrate a consistent improvement in accuracy when the proposed feature selection method is applied. The baseline model attained an accuracy of 96.12% on the Ph2 dataset with quadratic SVM classifiers, whereas the proposed method achieved an accuracy of 99.15%.

Moreover, employing ISIC-2016, the achieved accuracy is 92.68% with a medium-NN classifier, in contrast to 96.64% with the proposed method. Ultimately, the baseline model attained an accuracy of 87.41% on the HAM10000 dataset with a quadratic SVM, whereas the proposed technique achieved an accuracy of 94.10%. These findings underscore the effectiveness of the proposed feature selection technique in improving model's performance across diverse datasets and classifiers, highlighting its potential to optimize classification tasks in medical imaging and related domains.

Table 5: Comparison of accuracy and approximate F1-scores after feature selection methods

Dataset	Classifier	Method	Accuracy (%)	F1-Score
HAM10000	Trilayered-NN	GA	93.09	0.855
		PSO	91.63	0.820
		GWO	93.20	0.860
		BMNABC	93.17	0.860
		MBABC	94.10	0.873
HAM10000	Logistic Reg. Kernel	GA	90.97	0.820
		PSO	91.30	0.830
		GWO	92.00	0.840
		BMNABC	91.43	0.830
		MBABC	92.09	0.821
ISIC-2016	Medium-NN	GA	96.64	0.947
		PSO	93.24	0.932
		GWO	94.16	0.942
		BMNABC	95.78	0.958
		MBABC	96.12	0.961
ISIC-2016	ESD	GA	94.96	0.949
		PSO	93.63	0.936
		GWO	95.17	0.952
		BMNABC	95.10	0.951
		MBABC	95.40	0.954
PH^2	Quadratic-SVM	GA	96.93	0.969
		PSO	95.80	0.958
		GWO	94.86	0.949
		BMNABC	94.47	0.945
		MBABC	99.15	0.991
PH^2	ESD	GA	96.65	0.967
		PSO	98.19	0.982
		GWO	96.17	0.961
		BMNABC	97.12	0.971
		MBABC	99.01	0.990

Table 6: Classification accuracy comparison: Proposed method vs. baseline (without feature selection)

Dataset	Without feature selection		Proposed	
	Accuracy (%)	Classifier	Accuracy (%)	Classifier
PH^2	96.12	Quadratic-SVM	99.15	Quadratic-SVM
ISIC-2016	92.68	Medium-NN	96.64	Medium-NN
HAM10000	87.41	Quadratic-SVM	94.10	Trilayered-NN

6 Conclusion

Melanoma is the form of skin cancer that has the highest mortality rate, and its incidence has been rapidly rising over the past several years. The main objective of this study is to prioritize the identification and selection of feature information that exhibits the highest level of discrimination, ultimately resulting in improved accuracy in classification. Three benchmark datasets, namely PH^2, ISIC-2016, and HAM10000 are employed for the purpose of re-training the pre-trained Nasnet-Mobile CNN model. The features that have been retrieved, after applying transfer learning, are subsequently fed into the proposed feature selection module, which is defined as the mutated binary artificial bee colony (MBABC) algorithm. The selected features are subsequently fed into several families of classifiers to perform the ultimate classification. This study successfully addresses the issue of the "curse of dimensionality" by converting the high-dimensional data into lower dimensions. However, it does not perform exceptionally well in presence of a greater number of classes. One explanation is the increased correlation rate among the features of various classes. Furthermore, the presence of a limited number of image samples in certain classes significantly diminishes the accuracy of categorization.

In subsequent research, the number of image samples in specific categories will be augmented through the utilization of either data augmentation techniques or generative adversarial networks. Additionally, exploring the use of additional pre-trained models and developing a customized CNN model could be a promising avenue to pursue. When using a feature selection method, the effectiveness of the proposed algorithm relies heavily on the fitness function. Therefore, it is important to build a suitable fitness function that can efficiently handle a large feature set with more number of classes.

Acknowledgement: The authors extend their appreciation to Prince Sattam bin Abdulaziz University for funding this research work.

Funding Statement: The authors extend their appreciation to Prince Sattam bin Abdulaziz University for funding this research work through the project number (PSAU/2024/03/31540).

Author Contributions: Tallha Akram: Conceptualization, methodology, original draft, simulation, funding; Fahdah Almarshad: Methodology, review & editing; Anas Alsuhaibani: Methodology, original draft, review & editing; Syed Rameez Naqvi: Conceptualization, methodology, original draft, review & editing. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: The data used in this study were obtained from Kaggle.com. The dataset is publicly available and open to research community.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present research.

References

1. Malik S, Akram T, Awais M, Khan MA, Hadjouni M, Elmannai H, et al. An improved skin lesion boundary estimation for enhanced-intensity images using hybrid metaheuristics. Diagnostics. 2023;13(7):1285. doi:10.3390/diagnostics13071285.
2. Peate I. The skin: largest organ of the body. British J Healthcare Assist. 2021;15(9):446–51. doi:10.12968/bjha.2021.15.9.446.
3. Calabrò F, Sternberg C. Cancer and its management. BJU Int. 2006;97(3):651–1. doi:10.1111/j.1464-410X.2006.06115_2.x.
4. Bassel A, Abdulkareem AB, Alyasseri ZAA, Sani NS, Mohammed HJ. Automatic malignant and benign skin cancer classification using a hybrid deep learning approach. Diagnostics. 2022;12(10):2472. doi:10.3390/diagnostics12102472.
5. Foundation SC. Nonmelanoma Skin Cancer Statistics. Skin Cancer Facts, Estimated cases of basal cell carcinoma and squamous cell carcinoma in the United States; 2025. [cited 2025 Apr 7]. Available from: https://www.skincancer.org/skin-cancer-information/skin-cancer-facts/.
6. Society AC. Melanoma Statistics 2025. Cancer Facts & Figures. 2025. Estimated new cases and deaths for melanoma in the United States. [cited 2025 Apr 7]. Available from: https://www.cancer.org/cancer/types/melanoma-skin-cancer/about/key-statistics.html.
7. World Cancer Research Fund. Skin Cancer Statistics; 2025. [cited 2025 Mar 6]. Available from: https://www.wcrf.org/preventing-cancer/cancer-statistics/skin-cancer-statistics/.
8. Khayyati Kohnehshahri M, Sarkesh A, Mohamed Khosroshahi L, HajiEsmailPoor Z, Aghebati-Maleki A, Yousefi M, et al. Current status of skin cancers with a focus on immunology and immunotherapy. Cancer Cell Int. 2023;23(1):174. doi:10.1186/s12935-023-03012-7.
9. Bhatt H, Shah V, Shah K, Shah R, Shah M. State-of-the-art machine learning techniques for melanoma skin cancer detection and classification: a comprehensive review. Intell Medi. 2023;03(3):180–90. doi:10.1016/j.imed.2022.08.004.
10. Atta A, Khan MA, Asif M, Issa GF, Said RA, Faiz T. Classification of Skin Cancer empowered with convolutional neural network. In: 2022 International Conference on Cyber Resilience (ICCR). Dubai, United Arab Emirates; 2022. p. 1–6.
11. Ali K, Shaikh ZA, Khan AA, Laghari AA. Multiclass skin cancer classification using EfficientNets—a first step towards preventing skin cancer. Neurosci Inform. 2022;2(4):100034. doi:10.1016/j.neuri.2021.100034.
12. Bechelli S, Delhommelle J. Machine learning and deep learning algorithms for skin cancer classification from dermoscopic images. Bioengineering. 2022;9(3):97. doi:10.3390/bioengineering9030097.
13. Malik S, Akram T, Ashraf I, Rafiullah M, Ullah M, Tanveer J. A hybrid preprocessor DE-ABC for efficient skin-lesion segmentation with improved contrast. Diagnostics. 2022;12(11):2625. doi:10.3390/diagnostics12112625.
14. Ribani R, Marengoni M. A survey of transfer learning for convolutional neural networks. In: 2019 32nd SIBGRAPI Conference on Graphics, Patterns and Images Tutorials (SIBGRAPI-T). Rio de Janeiro, Brazil: IEEE; 2019. p. 47–57.
15. Devi GM, Neelambary V. Computer-aided diagnosis of white blood cell leukemia using VGG16 convolution neural network. In: 2022 4th International Conference on Inventive Research in Computing Applications (ICIRCA). Tamil Nadu, India: IEEE; 2022. p. 1064–8.
16. Bhatt B, Iyer A, Writer D, Schau G. A comparative study of transfer learning networks and siamese networks for acute lymphoblastic leukemia (ALL) diagnosis. J Student Res. 2022;11(4):1. doi:10.47611/jsrhs.v11i4.3489.
17. Ain QU, Akbar S, Hassan SA, Naaqvi Z. Diagnosis of leukemia disease through deep learning using microscopic images. In: 2022 2nd International Conference on Digital Futures and Transformative Technologies (ICoDT2). Rawalpindi, Pakistan: IEEE; 2022. p. 1–6.
18. Ahmad R, Awais M, Kausar N, Akram T. White blood cells classification using entropy-controlled deep features optimization. Diagnostics. 2023;13(3):352. doi:10.3390/diagnostics13030352.
19. Chen Q, Li M, Chen C, Zhou P, Lv X, Chen C. MDFNet: application of multimodal fusion method based on skin image and clinical data to skin cancer classification. J Cancer Res Clinic Oncol. 2023;149(7):3287–99. doi:10.1007/s00432-022-04180-1.

20. Hosny KM, Kassem MA. Refined residual deep convolutional network for skin lesion classification. J Digit Imaging. 2022;35(2):258–80. doi:10.1007/s10278-021-00552-0.
21. Tembhurne JV, Hebbar N, Patil HY, Diwan T. Skin cancer detection using ensemble of machine learning and deep learning techniques. Multimed Tools Appl. 2023;82(18):27501–24. doi:10.1007/s11042-023-14697-3.
22. Padhy S, Dash S, Kumar N, Singh SP, Kumar G, Moral P. Temporal integration of ResNet features with LSTM for enhanced skin lesion classification. Res Eng. 2025;25(1):104201. doi:10.1016/j.rineng.2025.104201.
23. Srinivasu PN, SivaSai JG, Ijaz MF, Bhoi AK, Kim W, Kang JJ. Classification of skin disease using deep learning neural networks with MobileNet V2 and LSTM. Sensors. 2021;21(8):2852. doi:10.3390/s21082852.
24. Zhong L, Li T, Cui M, Cui S, Wang H, Yu L. DSU-Net: dual-stage U-Net based on CNN and transformer for skin lesion segmentation. Biomed Signal Process Cont. 2025;100(1):107090. doi:10.1016/j.bspc.2024.107090.
25. Tan L, Wu H, Zhu J, Liang Y, Xia J. Clinical-inspired skin lesions recognition based on deep hair removal with multi-level feature fusion. Pattern Recognit. 2025;161(1):111325. doi:10.1016/j.patcog.2024.111325.
26. Bibi S, Khan MA, Shah JH, Damaševičius R, Alasiry A, Marzougui M, et al. MSRNet: multiclass skin lesion recognition using additional residual block based fine-tuned deep models information fusion and best feature selection. Diagnostics. 2023;13(19):3063. doi:10.3390/diagnostics13193063.
27. Tahir M, Naeem A, Malik H, Tanveer J, Naqvi RA, Lee SW. DSCC_Net: multi-classification deep learning models for diagnosing of skin cancer using dermoscopic images. Cancers. 2023;15(7):2179. doi:10.3390/cancers15072179.
28. Gilani SQ, Syed T, Umair M, Marques O. Skin cancer classification using deep spiking neural network. J Digit Imaging. 2023;36(3):1137–47. doi:10.1007/s10278-023-00776-2.
29. Kibriya H, Abdullah I, Kousar F. Melanoma lesion segmentation and classification using SegNet; 2023:1–6.
30. Khan MA, Sharif MI, Raza M, Anjum A, Saba T, Shad SA. Skin lesion segmentation and classification: a unified framework of deep neural network features fusion and selection. Expert Syst. 2022;39(7):e12497. doi:10.1111/exsy.12497.
31. Kaur R, GholamHosseini H, Sinha R, Lindén M. Melanoma classification using a novel deep convolutional neural network with dermoscopic images. Sensors. 2022;22(3):1134. doi:10.3390/s22031134.
32. Saarela M, Geogieva L. Robustness, stability, and fidelity of explanations for a deep skin cancer classification model. Appl Sci. 2022;12(19):9545. doi:10.3390/app12199545.
33. Khan MA, Sharif M, Akram T, Bukhari SAC, Nayak RS. Developed Newton-Raphson based deep features selection framework for skin lesion recognition. Pattern Recognit Letters. 2020;129(4/5):293–303. doi:10.1016/j.patrec.2019.11.034.
34. Mahbod A, Schaefer G, Wang C, Ecker R, Ellinge I. Skin lesion classification using hybrid deep neural networks. In: ICASSP, 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); Brighton, UK; 2019. p. 1229–33.
35. Akram T, Junejo R, Alsuhaibani A, Rafiullah M, Akram A, Almujally NA. Precision in dermatology: developing an optimal feature selection framework for skin lesion classification. Diagnostics. 2023;13(17):2848. doi:10.3390/diagnostics13172848.
36. Akram T, Alsuhaibani A, Khan MA, Khan SU, Naqvi SR, Bilal M. Dermo-optimizer: skin lesion classification using information-theoretic deep feature fusion and entropy-controlled binary bat optimization. Int J Imaging Syst Technol. 2024;34(5):e23172. doi:10.1002/ima.23172.
37. Vanacore A, Pellegrino MS, Ciardiello A. Fair evaluation of classifier predictive performance based on binary confusion matrix. Computat Statist. 2024;39(1):363–83. doi:10.1007/s00180-022-01301-9.
38. Torrey L, Shavlik J. Transfer learning. Vol. 2, In: Handbook of research on machine learning applications and trends: algorithms, methods, and techniques. IGI Global. 2010. p. 242–64.
39. Karaboga D. Artificial bee colony algorithm. Scholarpedia. 2010;5(3):6915.
40. Beheshti Z. BMNABC: binary multi-neighborhood artificial bee colony for high-dimensional discrete optimization problems. Cybernet Syst. 2018;49(7–8):452–74. doi:10.1080/01969722.2018.1541597.