RESEARCH ARTICLE

WILEY

A two-stream deep neural network-based intelligent system for complex skin cancer types classification

Muhammad Attique Khan[1] | Muhammad Sharif[1] |
Tallha Akram[2] | Seifedine Kadry[3] | Ching-Hsien Hsu[4,5,6]

[1]Department of Computer Science, COMSATS University Islamabad, Wah Campus, Wah Cantt, Pakistan

[2]Department of Electrical and Computer Engineering, COMSATS University Islamabad, Wah Campus, Wah Cantt, Pakistan

[3]Faculty of Applied Computing and Technology, Noroff University College, Kristiansand, Norway

[4]Guangdong-Hong Kong-Macao Joint Laboratory for Intelligent Micro-Nano Optoelectronic Technology, School of Mathematics and Big Data, Foshan University, Foshan, China

[5]Department of Computer Science and Information Engineering, Asia University, Taichung, Taiwan

[6]Department of Medical Research, China Medical University Hospital, China Medical University, Taichung, Taiwan

Correspondence
Muhammad Attique Khan, Department of Computer Science, COMSATS University Islamabad, Wah Campus, Wah Cantt 47040, Pakistan.
Email: attique@ciitwah.edu.pk

Abstract
Medical imaging systems installed in different hospitals and labs generate images in bulk, which could support medics to analyze infections or injuries. Manual inspection becomes difficult when there exist more images, therefore, intelligent systems are usually required for real-time diagnosis. Melanoma is one of the most common and severe forms of skin cancer that begins from the cells beneath the skin. Through dermoscopic images, it is possible to diagnose the infection at the early stages. In this regard, different approaches have been exploited for improved results. In this study, we propose a two-stream deep neural network information fusion framework for multiclass skin cancer classification. The proposed technique follows two streams: initially, a fusion-based contrast enhancement technique is proposed, which feeds enhanced images to the pretrained DenseNet201 architecture. The extracted features are later optimized using a skewness-controlled moth–flame optimization algorithm. In the second stream, deep features from the fine-tuned MobileNetV2 pretrained network are extracted and down-sampled using the proposed feature selection framework. Finally, most discriminant features from both networks are fused using a new parallel multimax coefficient correlation method.

A multiclass extreme learning machine classifier is used to classify lesion images. The testing process is initiated on three imbalanced skin data sets—HAM10000, ISBI2018, and ISIC2019. The simulations are performed without performing any data augmentation step in achieving an accuracy of 96.5%, 98%, and 89%, respectively. A fair comparison with the existing techniques reveals the improved performance of our proposed algorithm.

1 | INTRODUCTION

Melanoma is the deadliest type of skin cancer, which is showing a high occurrence frequency for the last three decades.[1] The malignant lesion, also known as melanoma, begins at the melanocyte cells level, and later spreads to the upper skin. The malignant melanoma has a higher mortality rate if not diagnosed at the earlier stages.[2] According to the World Health Organization (WHO), in the United States alone during 2019, a total number of skin cancer cases registered were 104,350 and deaths 11,650.[3] In 2020, the approximated skin cancer diagnosed cases are 196,060 which include 95,710 noninvasives and 100,350 invasives. From the latter, 40,160 and 60,190 are approximated to be men and women, respectively.[4] From these facts, it is evident that the diagnosed cases in the year 2020 will be increased by almost 2%, whereas the death rate will decrease by almost 5.3%. In the United States during the year 2021, the expected new diagnosed melanoma cases will be 106,110, whereas the number of deaths will be 7180. This shows that the number of melanoma cases will be increased with time. The early identification of skin cancer can help in reducing the mortality rate and it is highly recommended.[5]

Manual inspection of skin lesions with a naked eye is quite a time-consuming process and is also fallible.[6,7] Dermoscopy is one of the latest noninvasive technologies that diagnose the lesion with high accuracy.[8,9] The main idea of dermoscopy is to produce high-resolution images of infected skin regions. In clinics, dermatologists use few clinical methods, such as ABCDE rule,[10] 3-point checklist,[11] and 7-point checklist.[12] But due to time consumption, none of them is preferable.

Therefore, a computer-aided diagnosis (CAD) system is always required for a second opinion in the area of medical imaging[13,14] but especially for the dermatologists to support their decision.[15] The CAD system not only bypasses manual inspection, but also improves the diagnosis accuracy.[16,17] The CAD system for this specific purpose involves two building blocks: lesion segmentation and lesion classification.[18] In the segmentation task, the lesion regions are localized and boundaries of the infected part are drawn, whereas in the classification task, the localized lesions are classified into the relevant category (i.e., melanoma, benign keratosis [bkl], and nevi [nv]).[19] These tasks are challenging due to the following reasons: (i) the low contrast among infected skin lesion pixels and surrounding normal skin pixels, (ii) inter- and intratype

lesions similarity, (iii) the presence of lesion region on the boundary, (iv) some lesions are similar in the appearance but in actual they are from different classes, and (v) some lesions are corrupted due to few artifacts, such as hair, vessels, and bubbles. Many techniques are introduced in the literature for lesion segmentation and classification, but due to these issues, they did not achieve significant accuracy. However, a few techniques follow a sequential process, such as lesion segmentation at the first stage for region of interest (ROI) detection, which later provides discriminative features. These features are later fed into a classifier for the classification of lesions. But the classification accuracy depends on the strength of extracted features, and most of the researchers focus on two steps features selection and classification. However, the extracted features are difficult to analyze without any experimental process, and there exists a chance that some of the relevant features are discarded. Hence, the localized region is not confirmed to classify the correct category, and the classification accuracy is constrained.

Nowadays, machine learning (ML) showing significant success in the area of medical imaging for the prediction of diseases. ML utilized in several industrial applications for the prediction purposes: prediction of lung cancer,[20] prediction of COVID-19 from computed tomography (CT) images,[21] tracking of COVID-19,[22] spacecraft autonomy models,[23] and prognosis of aerospace systems using ML,[24] to name a few. Broad categorization includes medical imaging, visual surveillance, object classification, signal processing, and so forth. A convolutional neural network (CNN) is a class of deep neural networks that are used to analyze images. A simple CNN model includes a series of layers, such as convolutional layer, pooling layer, activation layer (rectified linear unit [ReLu]), normalization, fully connected layer, and output layer (Softmax). The CNN model works well for both simple and complex data sets. However, in case of complex data sets, an architecture modification improves the results. These CNN models perform better compared with the conventional techniques—for both segmentation and classification. To improve the accuracy, feature fusion and selection steps have been introduced by several researchers working in this area. Feature fusion step has the disadvantage of computational time, which could be compensated using an efficient feature selection technique. Many feature selection techniques are implemented in the literature, such as heuristic and metaheuristic. These areas are fully embedded in our proposed framework where the lesion images are classified after applying contrast stretching, deep feature selection, and feature fusion steps. The main motivation is to utilize only the most discriminant features to achieve an improved classification accuracy.

Many techniques are introduced in the literature for binary class skin lesion classification; however, the multiclass lesion classification is a complex task compared with the binary classification problem.[25-27] Most of the aforementioned techniques follow both segmentation and classification steps. According to these studies, it can be observed that the segmentation step has an impact on classification accuracy. High however, for multiclass skin pigment classification, the preprocessing and segmentation are not performed. They only focused on feature extraction through deep learning or conventional techniques and performed classification.[28] But due to the complex lesion shape, texture, and similarity in color, there still exists a space for improvement. Moreover, they faced the problem of imbalanced data sets. The imbalanced data set has a high impact on the classification accuracy due to the reason that the classes with more a number of images are having high occurring probability compared with the classes with a less number of images.

In this study, we proposed two-stream deep neural network architecture for multiclass skin cancer type classification using imbalanced data sets. Our major contributions are as follows:

i. A contrast stretching scheme is proposed which considers both local and global information. The local information is processed using a new activation function—calculated on the basis of variance and mean values. The global information is calculated by applying the top-hat filtering operation. Later, both operations are fused to generate enhanced images. A feature selection mechanism, skewness-controlled moth–flame optimization algorithm (ScMFO), is proposed that selects the most discriminant feature information from both pretrained models.

ii. A feature fusion scheme, multimax coefficient correlation (MCC), is also proposed to fuse the resultant feature vectors generated after applying the feature selection.

The rest of the manuscript is organized as follows: Related work of this manuscript which includes the discussion of some recent studies, given in Section 2. The proposed methodology, which includes two-stream network for lesion types classification, is presented in Section 4. Results are presented in Section 6. Finally, the conclusion of this study is given in Section 7.

2 | RELATED WORK

More recently, the deep learning is showing a huge success in the area of medical imaging.[29] For skin cancer patients, the process of multiclass skin lesion classification is not easy due to high similarity between different pigmented lesions.[30] As compared with deep learning techniques, particularly CNNs show significant performance in medical imaging for both disease segmentation and classification. Xie et al.[25] presented a mutual bootstrap CNN approach for automated skin lesion classification. In the initial phase, they performed lesion segmentation using an enhanced network of segmentation and coarse segmentation network. Simultaneously, the mask-guided CNN was applied through the bootstrapping method and performed the classification. Experiments were performed on two data sets named ISIC2017 and PH2. On these data sets, they achieved an accuracy of 93.8% and 97.7%, respectively. This method proved useful for simultaneous segmentation and classification of skin lesions when the data set was imbalanced. Al-Masni et al.[18] presented an integrated deep learning method for multiple skin lesion segmentation and classification. A full resolution deep learning model was proposed for lesion boundaries extraction. Later, four pretrained deep learning models were reused for the classification of skin lesions in the relevant category. Three data sets were used for the experimental process, such as ISBI2016 (includes two classes), ISIC2017 (includes three classes), ISBI2018 (includes seven classes), and achieved notable accuracy. Sikkandar et al.[26] introduced an automated system for skin lesion classification by the combination of adaptive Neuro Fuzzy classifier (NeFy) and GrabCut algorithm. In this method, the contrast of original images was improved using the top-hat filtering approach, and then segmentation was performed using the GrabCut algorithm. Next, the deep learning features were extracted and classified using the NeFy classifier. This method was evaluated on ISIC data set and achieved a sensitivity rate of 93.47% and an accuracy of above 97%. Mahbod et al.[27] presented a performance-based automated system for skin lesion classification. In this method, they performed segmentation through the cropping approach and extracted features. A baseline classifier was proposed and trained the extracted features for classification. In the output, the results showed that the segmentation step was useful for relevant feature extraction for only the segmented lesions. Al Nazi et al.[31] presented a deep learning method for skin lesion detection. They used U-net method for lesion boundaries segmentation. However, the main problem which they faced was data augmentation, which was resolved

through transfer learning. A few pretrained models were used for feature extraction and classified by the support vector machines (SVM) classifier. This method was tested on PH2 data set and achieved a mean accuracy of 92%. Garg and Jindal[32] introduced a segmentation technique for skin lesion using dermoscopic images. The preprocessing was performed in the initial step to remove hair and improve the contrast. In the next step, K-means segmentation was performed for lesion segmentation. The segmented lesions were further optimized through the firefly algorithm. The PH2 and ISIC data sets were employed for the experimental process and achieved an average accuracy of 98%. Miglani and Bhatia[33] presented an Efficient-Bo model for multiclass skin lesion classification. In this model, they decreased the number of parameters than the ResNet model. The HAM10000 data set was employed for the experimental process and achieved an improved performance. Iqbal et al.[34] presented a CNN-based framework for multiclass lesions classification. The presented CNN model includes several layers and filter sizes; however a few filters improved the efficiency of the system. Three data sets, such as ISIC2017, 2018, and 2019, were used for the experimental process and showed a state-of-the-art performance. Cano et al.[35] presented an augmentation-based framework for skin lesion classification. They trained NasNet deep network using augmented data set, whereas none of preprocessing steps was performed. The ISIC data set was used for the experimental process and obtained improved accuracy for eight classes' classification. López-Leyva et al.[36] considered Fourier Spectral Transform (FST) and Adaptive Color Transformation (ACT) along with artificial neural network (ANN) for the diagnosis of multiclass skin lesion.

A few other techniques were also presented in the literature, such as DermaDL-based approach,[37] ML-based techniques[38] densely connected based method for multiclass lesion classification,[39] a Newton–Raphson-based selection of the most optimal features for lesion classification,[40] fusion of multiple deep learning features,[41] and saliency method along with deep neural network,[28] to name a few. Most of the techniques rely on the preprocessing steps, such as contrast enhancement and lesion segmentation, which later feed to deep learning models for feature extraction.[42]

The addition of preprocessing and segmentation steps improves the classification accuracy, which is not fully covered in the existing techniques, detailed above. Similarly, the existence of irrelevant and redundant feature information leads to the problem of false labeling. The mentioned problems are not the primary focus of the aforementioned techniques. Therefore, it is somewhat essential to propose a new method that successfully tackles these kinds of issues.
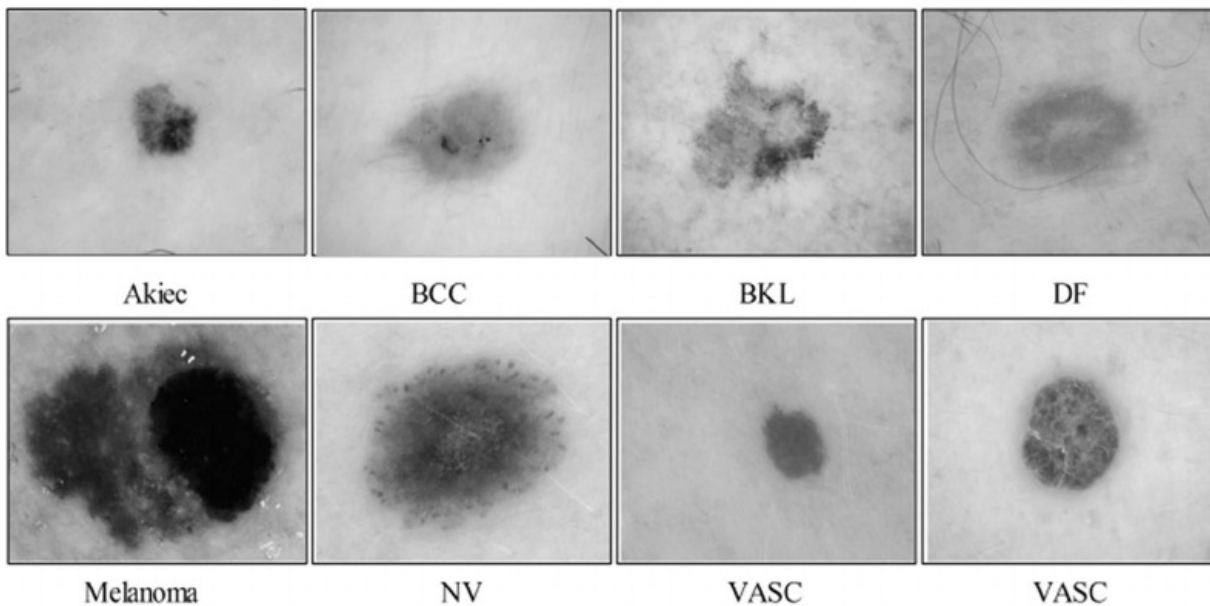
3 | DATA SETS

Three publically available benchmark data sets are used in this study for the evaluation of proposed framework. The selected data sets are HAM10000, ISBI2018, and ISBI2019. These data sets belong to the category of imbalanced data sets with different image samples of each class. The HAM10000 data set consists of 10,015 dermoscopic images with seven classes: actinic keratoses (akiec), basal cell carcinoma (bcc), bkl, dermatofibroma (df), melanoma (mel), melanocytic nv, and vascular (vasc). The number of images of akiec is 327, bcc 514, bkl 1099, df 115, mel 1113, nv 6705, and vasc 142. This number shows that this data set is highly imbalanced and suitable for a simple classification method. Sample images are shown in Figure 1.

ISBI2018[44] data set consists of 10,015 dermoscopic images for the classification task. Similar to HAM10000 data sets, ISBI2018 includes seven classes, including akiec, bcc, bkl, df, mel, nv, and vasc. For validation and testing, the number of images is 193 and 1512, respectively. A few
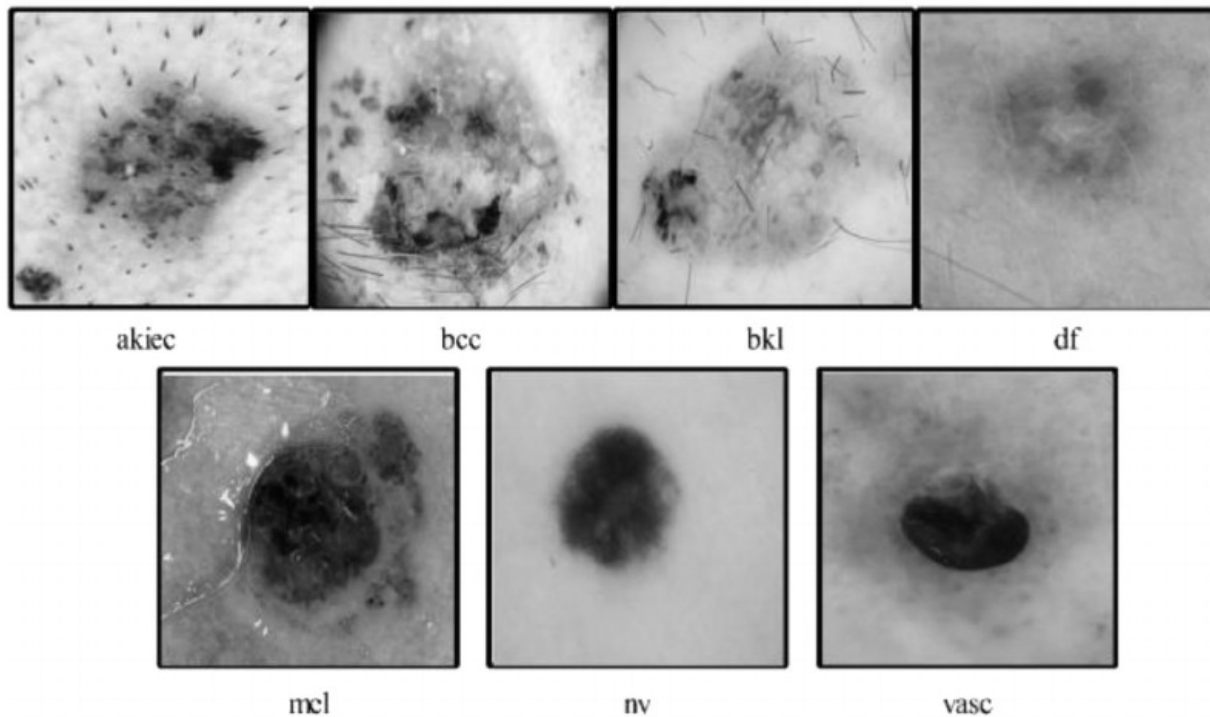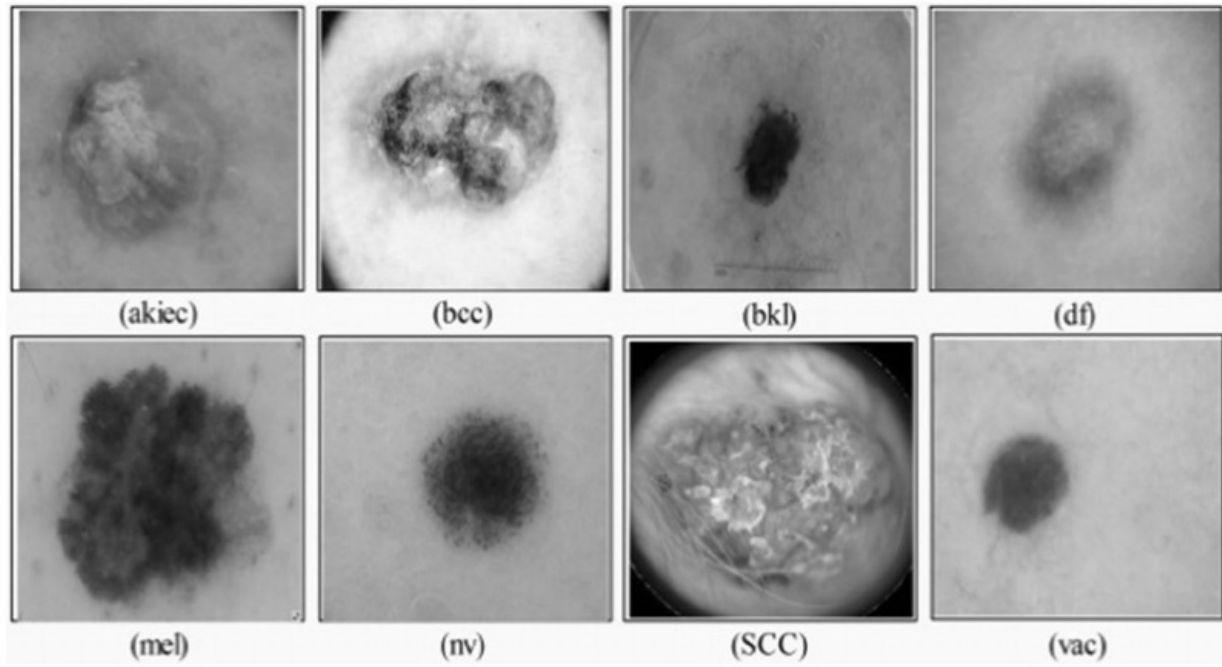
FIGURE 1 Sample skin cancer image samples from HAM10000 data set.[43] akiec, actinic keratoses; bcc, basal cell carcinoma; bkl, benign keratosis; df, dermatofibroma; mel, melanoma; nv, nevi; vasc, vascular [Color figure can be viewed at wileyonlinelibrary.com]

FIGURE 2 Sample images of ISBI2018 data set.[44] akiec, actinic keratoses; bcc, basal cell carcinoma; bkl, benign keratosis; df, dermatofibroma; mel, melanoma; nv, nevi; vasc, vascular [Color figure can be viewed at wileyonlinelibrary.com]

sample images are shown in Figure 2. Similarly, ISIC2019 data set[45] consists of 25,331 dermoscopic images with eight classes. The eight classes are akiec (867 images), bcc (3323 images), bkl (2624 images), df (239 images), mel (4522 images), nv (12,875 images), squamos cell carcinoma (628 images), and vasc (253 images). A few sample images are shown in Figure 3.

FIGURE 3 Sample skin cancer images collected from the ISBI2019 data set. akiec, actinic keratoses; bcc, basal cell carcinoma; bkl, benign keratosis; df, dermatofibroma; mel, melanoma; nv, nevi; vasc, vascular [Color figure can be viewed at wileyonlinelibrary.com]

4 | PROPOSED METHODOLOGY

In this article, we proposed a new two-stream deep neural network architecture for multiclass skin lesion classification. Initially, a new contrast enhancement technique has been proposed and used for both streams. In the first stream, the DenseNet201 pretrained model is fine-tuned and extract features which are later optimized using modified algorithm name ScMFO. In the second stream, MobileNetV2 pretrained deep neural network is first fine-tuned to extract features which are later down-sampled using a new approach, ScMFO. The output of streams at this point is fused using a new parallel fusion approach, namely, MCC. The final features are classified using a multiclass extreme learning machine (MCELM) classifier. The results are obtained in the form of numerical and visual (labeled). The complete architecture of the proposed two-stream method is illustrated in Figure 4.

4.1 | Local–global information fusion-based contrast enhancement

Consider an input image $Im(x, y)$ having dimensions $512 \times 512 \times 3$ (seen in Figure 5). Local contrast of input image $Im(x, y)$ is represented by $Im_{lc}(x, y)$ and global contrast is denoted by $Im_{gl}(x, y)$, where the dimension of resultant image should be the same as the input image. Then, the resultant fusion image is denoted by $Im_{rs}(x, y)$ with dimensions $512 \times 512 \times 3$. Initially, the local contrast of an input image is enhanced using a new hybrid approach named bottom-hat-controlled local mean (BHcLM). In this approach, initially, the bottom-hat transform is applied to the original image with a dynamic structural element. Then, the local mean-based brightness is computed and increased based on multiplication operations. Given, $Im(x, y)$ is an original image, then bottom-hat transform is applied as follows:
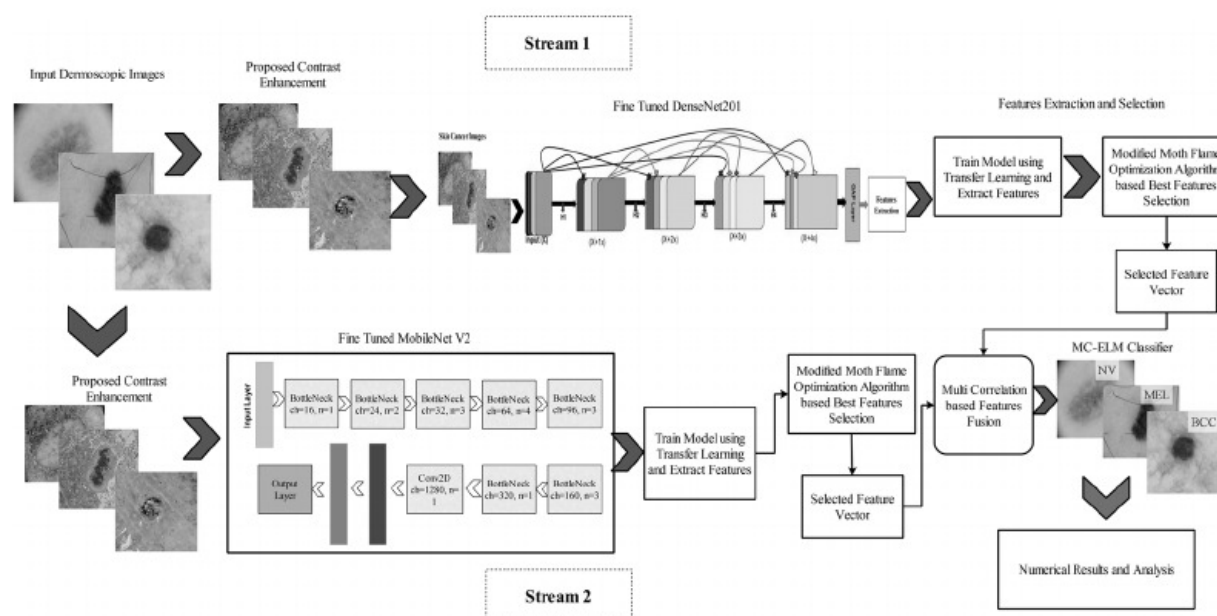
FIGURE 4 Proposed architecture diagram of multiclass skin lesion classification using Two-Stream Network [Color figure can be viewed at wileyonlinelibrary.com]
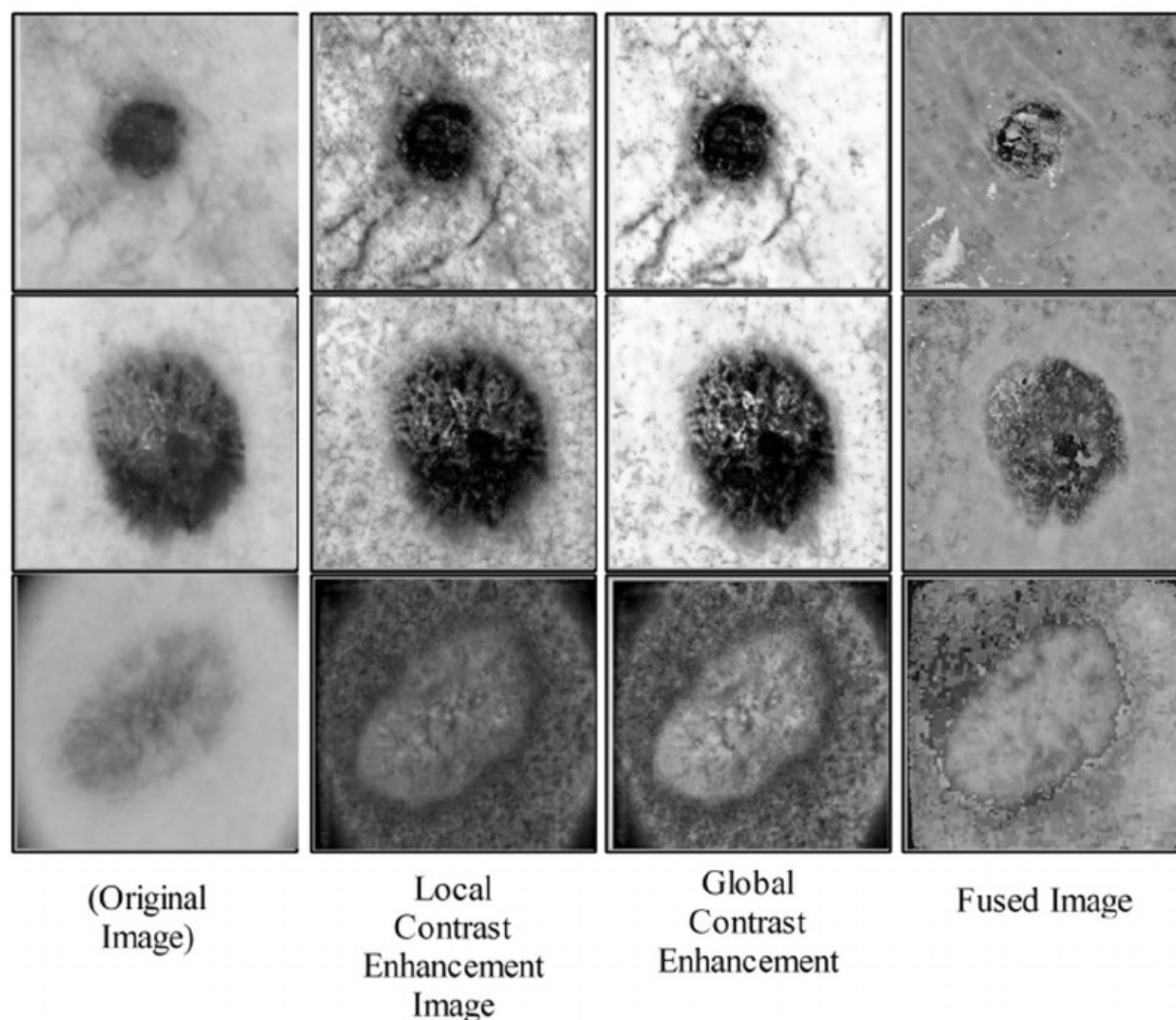
FIGURE 5 Proposed local–global information fusion-based contrast enhancement [Color figure can be viewed at wileyonlinelibrary.com]

$$Im_{\text{bot}}(x, y) = Im(x, y) \cdot se - Im(x, y). \tag{1}$$

This transformation is further refined using log transform operation $\log(Im(x, y))$.

$$Im_{\text{bot}}(x, y) = \log(Im(x, y)) \times Im_{\text{bot}}(x, y). \tag{2}$$

Here, we utilize the dynamic structuring element (se) instead of static value. The dynamic value of se is computed as follows:

$$se = \frac{\sum_{i=1}^{N} \sum_{j=1}^{M} (Im_{ij}(x, y))}{MN}. \tag{3}$$

Finally, local mean is computed from the bottom-hat transformed image. Through this operation, the brightness of an image is computed to get the textural information:

$$Bright = \underset{\mu}{argmin} \sum_{i=1}^{N} \left\{ (Im_{\text{bot}}(i) - \mu)^2 + 0.2 \sum_{i=(x,y)} |\partial_i Im_{\text{bot}}(i)| \right\}. \tag{4}$$

Finally, the final local contrast-enhanced image is computed using Equation (5) and visual examples are provided in Figure 5.

$$Im_{\text{lc}}(x, y) = Im_{\text{bot}}(x, y) \times Bright. \tag{5}$$

For global contrast enhancement, we first performed modified top-hat transformation and then multiplied the output with an existing techniques named brightness preserving Bi HE (BPiHE).[46] This hybrid method increases the global image contrast. Mathematically, this process is formulated as follows:

$$Im_{\text{top}}(x, y) = Im(x, y) - Im(x, y) \circ se, \tag{6}$$

$$Im_{\text{top}}(x, y) = \frac{Im_{\text{top}}(x, y) + \mu}{\sigma + 0.1} \times 2, \tag{7}$$

$$Im_{\text{BPiHE}}(x, y) = g_{\text{L}}(Im_{\text{L}}) \cup g_{\text{U}}(Im_{\text{U}}), \tag{8}$$

$$Im_{\text{gl}}(x, y) = Im_{\text{BPiHE}}(x, y) \times Im_{\text{top}}(x, y). \tag{9}$$

Here, $Im_{\text{top}}(x, y)$ represents the modified top-hat transformation, $Im_{\text{BPiHE}}(x, y)$ the BPiHE-based enhanced image, and $Im_{\text{gl}}(x, y)$ the final global contrast transformed image. Visual results can be observed in Figure 5 (global contrast enhancement). Finally, the fusion of both local and global images is performed using the following objective function.

$$Im_{\text{rs}}(x, y) = \frac{Im_{\text{gl}}(x, y) \times \sigma(Im_{\text{gl}}(x, y)) + Im_{\text{lc}}(x, y) \times \sigma(Im_{\text{lc}}(x, y))}{\sum(Im_{\text{lc}}, Im_{\text{gl}}) + 0.1}, \tag{10}$$

where $Im_{\text{rs}}(x, y)$ is the final fused image and visual results are illustrated in Figure 5 (fused image).

4.2 | Convolutional neural network

A CNN consists of input layer, several hidden layers, and one output layer. In the hidden layers there exists the convolutional layers, pooling layers, batch normalization layers, and ReLu layers. The input layer considers an input image with dimensions $h \times w \times c$, where c denotes the number of channels. In the hidden layers, the first layer is a convolutional layer. This is one of the most important layers and requires three inputs w, h, and c for the convolutional operation. Feature maps are constructed in this layer which later utilizes for the visualization and in the activation layers. To get a better insight, a number of research articles are available.

4.3 | Transfer learning

Transfer learning is a process of reusing the existing models, trained on one set of data, on a different data. This process is exceedingly useful when the number of image samples is not enough for a comprehensive training. Consider, ψ_S represents source domain define as $\psi_S = \{\{\phi_{s1}, y_{s1}\}, ..., \{\phi_{sn}, y_{sn}\}\}$, where $\phi_{si} \in \chi_S$. Here, ϕ_{si} represents the input source features and $y_{si} \in Y_S$ are source labels. As shown in Figures 6 and 7, the ImageNet data set is used as a source data, DenseNet201 and MobileV2 pretrained models are source models, and total 1000 class labels. Similarly, ψ_T represents the target domain defined by $\psi_T = \{\{\phi_{T1}, y_{T1}\}, ..., \{\phi_{Tn}, y_{Tn}\}\}$, where $\phi_{Ti} \in \chi_T$. Here, ϕ_{Ti} represents the target features and y_{Ti} the target labels.

4.4 | Fine-tuned DenseNet201

DenseNet is an extension of ResNet with direct connectivity among the layers. Therefore, there is no chance of information loss because the upcoming layer takes input from the previous and passes it to the next layer. The concatenation of attributes is done and forwarded to the upcoming layers. Hence there are p number of inputs at the pth layer. All previous attributes are considered as input
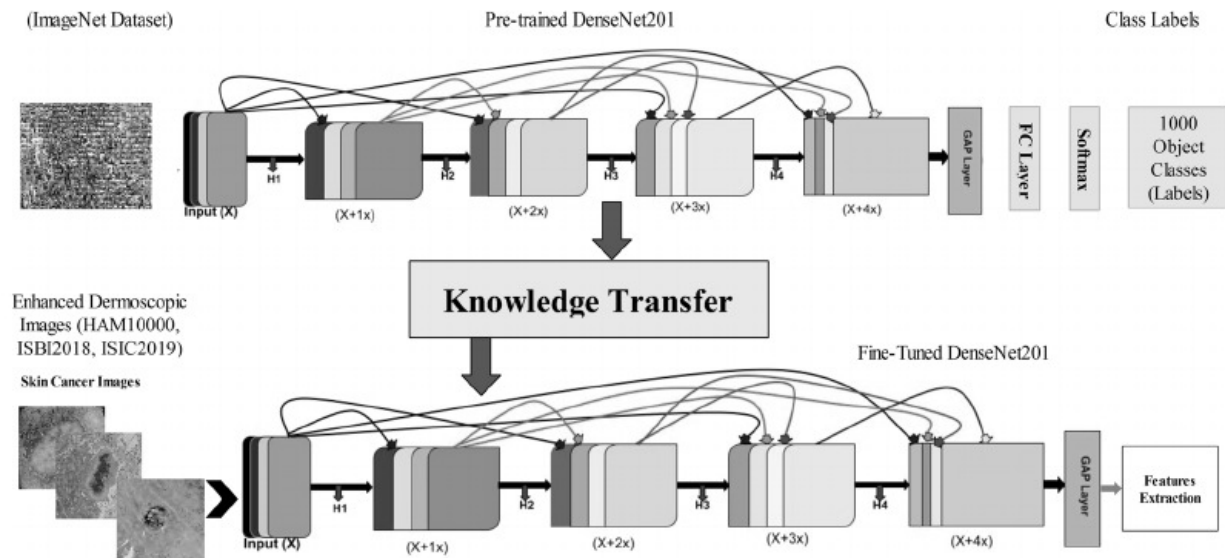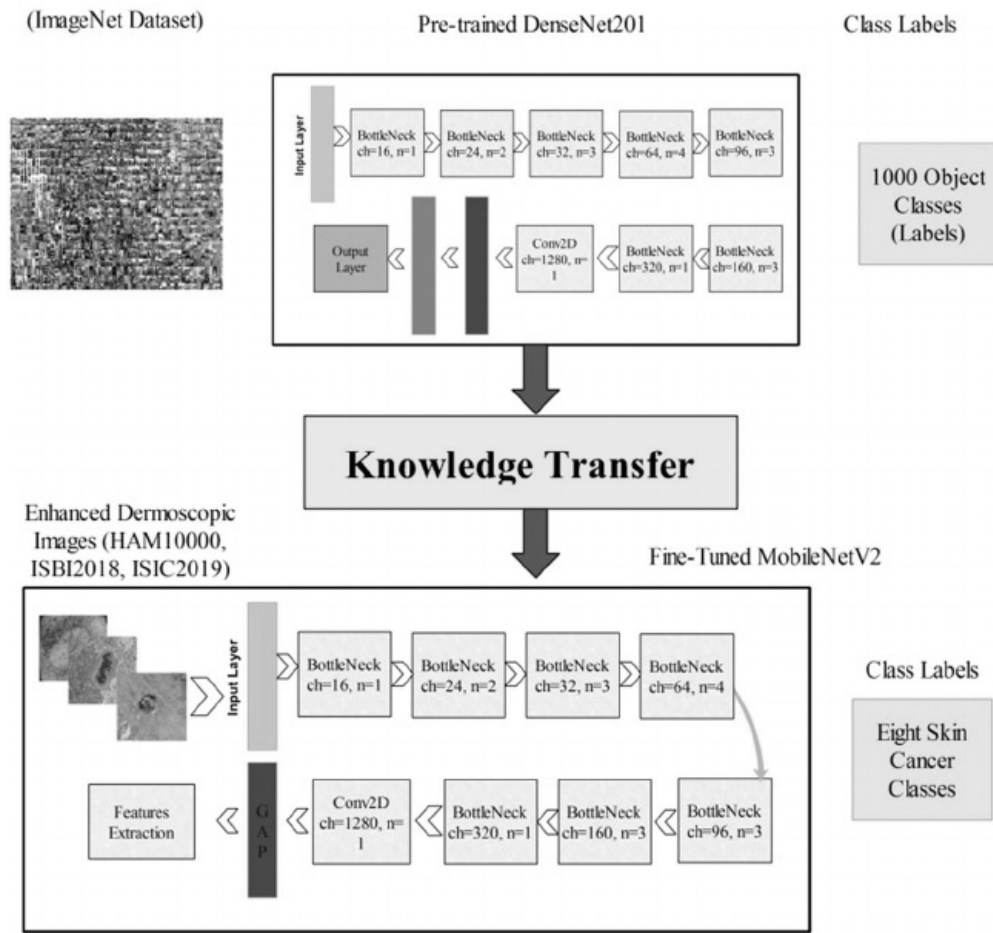

FIGURE 6 Features extraction using target label (TL) of Stream 1 [Color figure can be viewed at wileyonlinelibrary.com]

FIGURE 7   Features extraction using target label (TL) of Stream 2 [Color figure can be viewed at wileyonlinelibrary.com]

for the next layer. These attributes of the present layers are fed to upcoming layers $-p$. The total number of cumulative connections is $\frac{P(P+1)}{P}$ in the network of P-layers. While in case of the traditional networks there are P number of connections. This network minimizes the number of parameters compared with the old networks. The key advantage of this network is that layers directly access the information which results in an implied deep regulation.[47]

Let F_0 is an image frame and is fed as input to the network having P layers. The transformation is nonlinear and denoted as $V_p(\cdot)$, where the layers are guided by p. $V_p(\cdot)$ has few operations, such as ReLu,[48] Pooling,[49] Convolution, and Batch Normalization,[50] and this function is reflected as composite. The demonstration of output is done as F_p at pth layer. The connection is mathematically defined as follows:

$$F_p = V_p\left([F_0, F_1, ..., F_{p-1}]\right),\tag{11}$$

where input $F_0, ..., F_{p-1}$ is obtained at layer from pth from prior layers and $([F_0, F_1, ..., F_{p-1}])$ demonstrates the final features after concatenation in layers $0, ..., p-1$. A composite function is also applied in this network which followed the pooling layer. For the growth rate, hyperparameter α is used. This parameter shows a total channel number of a layer and its value is $\alpha = 32$. The main advantage of this network is in terms of minimized parameters. The features x are directly accessed by the layer, hence, the value of α is based on newly added features after
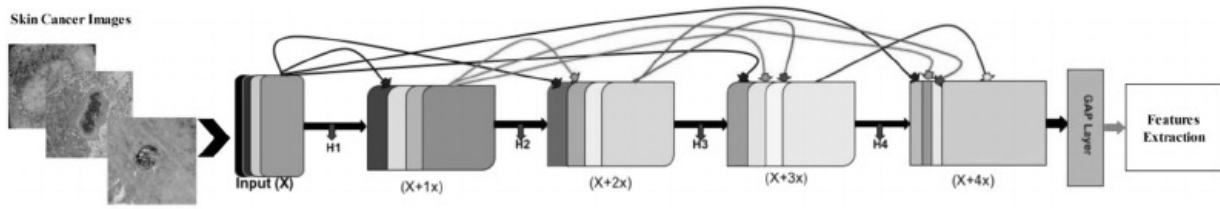
FIGURE 8 Fine-tuned DenseNet201 architecture for skin lesion feature extraction [Color figure can be viewed at wileyonlinelibrary.com]

a transition is made. For the minimization of features, a 1×1 convolution layer is added—acting as a bottleneck, which overall improves the system's performance. This is the main reason for a bottleneck layer in the DenseNet that computes $4k$ feature map. In the end, all features are compressed through a transition layer.

The fine-tuned model is shown in Figure 8. In this figure, it can be seen that the proposed contrast-enhanced images are given as an input to the network and also Softmax and fully connected layers are removed. Then, a new fully connected layer is added that has eight classes (i.e., skin cancer categories). TL is employed to train this fine-tuned model. Then, the features are extracted from the global average pool layer (GAP). However, we analyze the extracted features and observed that the current performance of this layer feature is not sufficient for the multiclass skin cancer classification. Therefore, we save these features into a new vector represented by ψ_{v1}^{N}, where $v1$ represents the dimension of extracted vector of this layer which is 2048 and N the total number of training images. The features of this layer are optimized in the late stage for better results.

4.5 | Fine-tuned MobileNetV2

This is one of the most robust deep learning models that are built under the resource-constrained environment. Instead of residual blocks, in this network, the inverted residual connections are adopted. This is used for the connection between thin blocks. The process of feature extraction is started from the spatial convolution layer with a kernel size 3×3 for the enhanced input image. The size of the input image for the convolutional layer is $224 \times 224 \times 3$. The result of this convolutional layer is entered into 19 bottleneck layers. Each bottleneck layer consists of three convolutional operations as 1×1 convolution, 3×3 depth convolutions, and 1×1 pointwise convolutions. The convolutional operation 1×1 increases the number of features, whereas 3×3 depth convolution operation separates the channels. The input $g \times w \times r$, the convolutional operation is used followed by an ReLu6. Here, g represents the height, w the width, and r the channels. In the output, it returns an output of $g \times w \times (tr)$. In the next time, this output is selected as an input and 3×3 depthwise convolutional operation is used. This operation returns an output of $\frac{g}{s} \times \frac{w}{s} \times tr$. This output is further transformed through a pointwise convolutional operation and obtained output of $\frac{g}{s} \times \frac{w}{s} \times r'$. Here, r' represents the number of channels. A detailed architecture of this model is presented in Table 1.

In this paper, we fine-tuned this model for multiclass skin cancer classification. For this purpose, the last Softmax and convolutional layer are removed and added a new convolutional layer that includes eight skin cancer categories. Originally, this layer includes 1000 object

TABLE 1 Architecture of MobileNetV2

Input	Operator	_ft_	_ch_	_n_	Stride (s)
$224 \times 224 \times 3$	Conv2D	–	32	1	2
$112 \times 112 \times 32$	BtN	1	16	1	1
$112 \times 112 \times 16$	BtN	6	24	2	2
$56 \times 56 \times 24$	BtN	6	32	3	2
$28 \times 28 \times 32$	BtN	6	64	4	2
$14 \times 14 \times 64$	BtN	6	96	3	1
$14 \times 14 \times 96$	BtN	6	160	3	2
$7 \times 7 \times 160$	BtN	6	320	1	1
$7 \times 7 \times 320$	Conv 1×1	–	1280	1	1
$7 \times 7 \times 1280$	GAP 7×7	–	–	1	–
$1 \times 1 \times 1280$	Conv 1×1	–	K	–	

Note: BtN represents a bottleneck.

Abbreviations: GAP, global average pool layer.
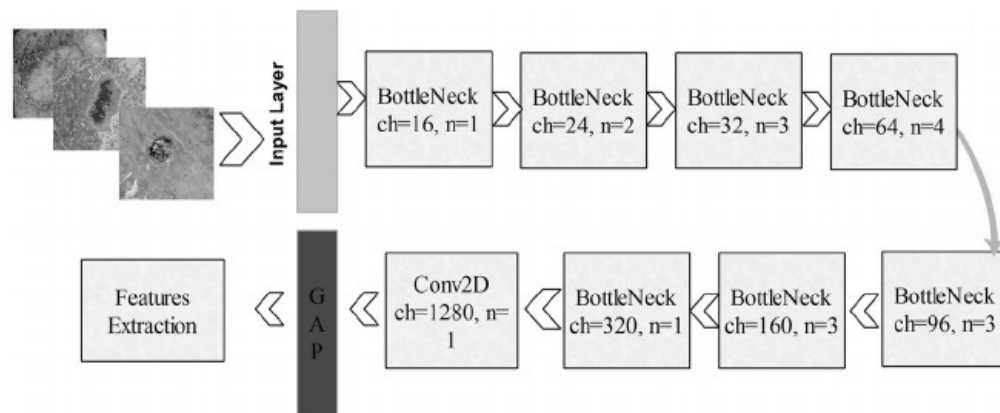
FIGURE 9 Fine-tuned MobileNetV2 for skin lesion features extraction [Color figure can be viewed at wileyonlinelibrary.com]

classes. Visually, this fine-tuned model is illustrated in Figure 9. The TL is employed for the training of this fine-tuned model and features are extracted from the GAP layer for the classification purpose. However, based on the Softmax accuracy, it is compulsory to optimize this feature vector. Hence, we store the features of this layer in a new vector represented by ψ_{vec2}^{N}, where *vec2* represents the features length and N are the total training images.

4.6 | Features selection

In this stream, we applied a modified moth–flame optimization (MMFO) algorithm for the best features selection. The moth–flame optimization (MFO) algorithm[51] is a population-based search algorithm. This algorithm starts with the initial population parameters initialization and

generated initial moths. These moths are randomly generated in this algorithm. Each moth in this algorithm is considered as a feasible solution with decision variables (moth position). Then, the fitness is computed to find the moth position and tagging the best position by flame. As in this algorithm, the moths are moved in a spiral way, so their position is updated in the next phase to achieve the best position.[52] This process is continuing until the termination criterion is not satisfied. When the termination criterion is satisfied then the best features (positions) are obtained for each moth.

Consider ψ_{v1}^N is deep feature vector of $v1$ features dimension and $v1 = 2048$, extracted through the fine-tuned DenseNet201 pretrained model. Also, ψ_{v2}^N is a deep feature vector of $v2$ and $v2 = 1056$ extracted through a fine-tuned MobileNetV2 model. Consider, we have three tuples in MFO and defined as

$$MFO = (IN_p, UP_p, Te_p),$$ (12)

where IN_p is the initialization phase in which random positions of moths are generated and fitness is computed to find the best position for each flame. The second phase is updating phase represented by UP_p. In this phase, moths position is updated using the following formulation:

$$\psi_i = \vec{D}_i e^{bl} \cos(2\pi l) + Fact_u, \quad i, u = 1, 2, 3, ..., N,$$ (13)

$$\vec{D}_i = |Fact_u - \psi_i|,$$ (14)

where $l = [-1,1]$ and b is the uth flame. These parameters are used to define the shape of logarithmic spiral function. To balance between the exploration and exploitation, the following formulation is adopted:

$$FNO = round\left(N - C_r \times \frac{N-1}{C_{\max}}\right),$$ (15)

where C_r is the current iteration, $C_{\max}$ the maximum iterations, and N the different moth locations. The third phase is termination (stopping condition). This phase is based on the number of iterations. In this study, we set the number of iterations is 200 and the number of populations is 20. For the fitness function, distance is computed between the moth and flame. In this study, we modify the updation phase by utilizing two-phase selection approach. For this purpose, we proposed a skewness value-based activation function, define as follows:

$$Activation = \begin{cases} F\tilde{N}O & \text{for } FNO_i \geq SK, \\ Not \quad Update \quad Position & Otherise. \end{cases}$$ (16)

The set of features that fulfill this activation function updates the position. The main advantage of this step is to minimize the dimension of the final selected feature vector. After 200 iterations, we obtain two optimal selected feature vectors for ψ_{v1}^N and ψ_{v2}^N of dimension $N \times 816$ (for this study according to the experimental setup) and $N \times 828$. Both selected vectors are denoted by ψ_{sv1}^N and ψ_{sv2}^N. Further, this modified algorithm is also presented in Algorithm 1.

Algorithm 1

Output: Selected feature vector $\leftarrow \psi_{sv}^{N}$ of dimension $N \times v_s$.

Input: Deep Feature vector $\leftarrow \psi_{sv}^{N}$ of dimension $N \times v$.

Start

Step 1: Define parameters, population $= 20$, Iterations $= 200$.

Step 2: Generate random moths.

Step 3: Tagging best moth position based on Fitness function.

Step 4: Update flame number.

Step 5: Compute Distance between moth and flame.

Step 6: Update moth positions based on the corresponding moths.

Step 7: Two-phase selection using Skewness Activation.

Step 8: Repeat Steps 3–7 until condition is not terminated.

Step 9: Best features (position) are selected.

End

4.7 | Deep features fusion

Consider, we have two best feature vectors ψ_{sv1}^{N} and ψ_{sv2}^{N} of dimension $N \times 816$ and $N \times 828$, respectively. Suppose ψ_{fus}^{N} is a fused feature vector with dimension $N \times fus$, where fus represents the length of a fused feature vector. Suppose, $\mathbb{V}^{(1)} \in \psi_{sv1}^{N}$ and $\mathbb{V}^{(2)} \in \psi_{sv2}^{N}$, then

$$\mathbb{V}^{(1)\prime} = \left(v_1^{(1)\prime}, v_2^{(1)\prime}, \dots, v_{sv1}^{(1)\prime} \right), \tag{17}$$

$$\mathbb{V}^{(2)\prime} = \left(v_1^{(2)\prime}, v_2^{(2)\prime}, \dots, v_{sv2}^{(2)\prime} \right), \tag{18}$$

where $sv1$ and $sv2$ represent the dimension of feature vectors $\mathbb{V}^{(1)\prime}$ and $\mathbb{V}^{(2)\prime}$. As we know the dimension of each feature vector $N \times 816$ and $N \times 816$; therefore, first we make the equal length of both vectors. For this purpose, we compute the length of feature vectors and based on the maximum length feature vector, we performed mean padding for another feature vector (low dimensional vector). Mathematically, it is defined as follows:

$$MaxLength = \mathbb{V}^{(max)\prime} = Max(\mathbb{V}^{(1)\prime}, \mathbb{V}^{(2)\prime}), \tag{19}$$

$$MinLength = \mathbb{V}^{(min)\prime} = Min(\mathbb{V}^{(1)\prime}, \mathbb{V}^{(2)\prime}), \tag{20}$$

$$pad_{(\mu)} = \sum_{i=1}^{sv^{(i)}} (MaxLength(\mathbb{V}^{(i)})), \tag{21}$$

$$\mathbb{V}^{(min)\prime} = [\mathbb{V}^{(min)\prime} \leftarrow \mathbb{V}^{(max)\prime}(pad_{(\mu)})]. \tag{22}$$

The $\mathbb{V}^{(min)\prime}$ represents the less dimensional feature vector and $\mathbb{V}^{(max)\prime}$ the maximum dimensional feature vector. The $pad_{(\mu)}$ represents the mean value of the maximum length feature vector which is used for padding. After padding, the linear combination[53] is computed among

both vectors $\mathbb{V}^{(1)'}$ and $\mathbb{V}^{(2)'}$, where in this study $\mathbb{V}^{(1)'} \in \mathbb{V}^{(max)'}$ and $\mathbb{V}^{(2)'}$ is updated $\mathbb{V}^{(min)'}$. The linear combination is computed as follows:

$$\Phi^{(1)} = \alpha^{(1)'} \mathbb{V}^{(1)}, \quad \text{then } \sigma^2(\Phi^{(1)}) = \alpha^{(1)'} S_{11} \alpha^{(1)'}, \tag{23}$$

$$\Phi^{(2)} = \alpha^{(2)'} \mathbb{V}^{(2)}, \quad \text{then } \sigma^2(\Phi^{(2)}) = \alpha^{(2)'} S_{22} \alpha^{(2)'}, \tag{24}$$

where $\alpha^{(1)'}$ and $\alpha^{(2)'}$ represent the first and second projection components of a group of projection directions. The S_{11} and S_{22} described the between-group and within-group covariance matrix and $S_{11} = S_{22} = 1,2,\ldots,n$. Hence, maximize the relation coefficient as follows:

$$Max_{\mathbb{V}}(\alpha) = \sqrt{1 - \frac{det\left(G\left(\mathbb{V}_1^{(1)}, \mathbb{V}_2^{(2)}\right)\right)}{\left\|\mathbb{V}_1^{(1)}\right\| \cdot \left\|\mathbb{V}_2^{(2)}\right\|}}. \tag{25}$$

This formulation maximizing the relation coefficient between projections and does not stop until all features' groups are completed. The features whose relation coefficient is above ≥ 0.4 are selected for final classification. The main reason behind the selection of this value is an analysis of features on different values, such as 0.2, 0.3, 0.5, 0.6, and 0.7. The features of the final fused vector are classified using MCELM.[54] The method of extreme learning machine (ELM) is multiclass. Through MCELM, the features are classified in skin class labels, such as akiec, bkl, MEL, and so forth. Figure 10 shows the scatter plot. This figure shows the correct and incorrect prediction of skin cancer types. The detailed results of the proposed methodology are discussed in Section 5.
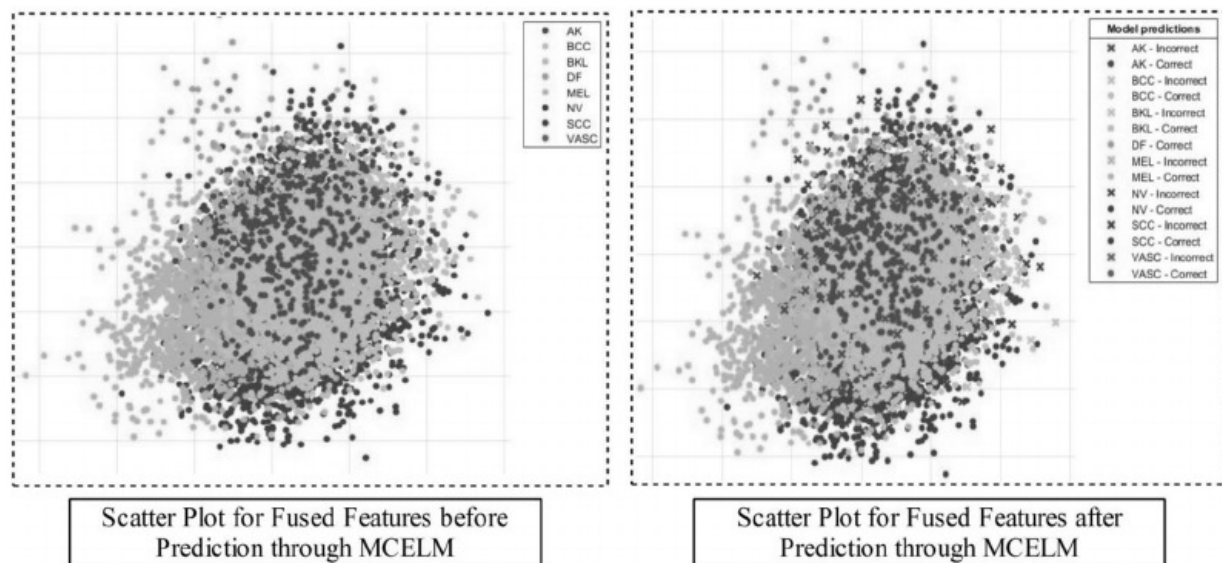
FIGURE 10 Prediction through MCELM using fused feature vector. akiec, actinic keratoses; bcc, basal cell carcinoma; bkl, benign keratosis; df, dermatofibroma; MCELM, multiclass extreme learning machine; mel, melanoma; nv, nevi; vasc, vascular [Color figure can be viewed at wileyonlinelibrary.com]

5 | EXPERIMENTAL RESULTS AND ANALYSIS

The results are computed on three selected data sets—keeping the training/testing ratio to be 50:50. For the training of both streams, the training rate is 0.001, the number of epochs is 500, the mini-batch size is 64, and the momentum is 0.7. For each epoch, five iterations are performed. The MCELM is utilized as the main classifier and compares its performance with other well-known classifiers, such as multiclass SVM (MCSVM), Gaussian naïve Bayes (GN Bayes), weighted k-nearest neighbors (W-KNN), Softmax, and ensemble baggage tree (EBT). The performance of each classifier is validating by the following evaluation protocols, such as sensitivity rate (Sen), precision rate (Prec), F1-Score, accuracy (Acc), and computation time (s).[55] All results are obtained using 10-fold cross-validation. The simulation of this study is conducted on MATLAB2020b using Personal Desktop having 16 GB of RAM and 16 GB of the graphics card.

6 | RESULTS

The results are presented for the proposed framework and compared with individual steps in this framework. Moreover, confusion matrix-based results are also validated. Table 2 presents the proposed classification results against the HAM10000 data set. In this table, it is described that the MCELM generated the best sensitivity rate and accuracy of 89.01% and 96.5%. The best precision rate is 97.47% achieved by MCSVM. Moreover, the highest F1-Score and accuracy rate are 90.83% and 96.5% for MCELM. The accuracy obtained by MCSVM is 94.6%, which is the second-best accuracy for this data set. The lowest sensitivity rate is 65.37% obtained by the GN Bayes classifier. The obtained accuracy for the rest of classifiers is 92.9%, 93%, 93.6%, and 92.4%. Moreover, the computational time is also noted, and the minimum time is 10.761 (s) for MCELM. The highest time is 50.486 (s) for the GN Bayes classifier. The proposed accuracy of MCELM is also compared with individual streams, illustrated in Figure 3. The individual results are given in Table 3. In this table, the results are computed for Stream 1 selected feature (S1SL) and Stream 2 selected feature (S2SL). The achieved accuracy of S1SL is 92.3% and noted time is 9.446 (s). For the S2SL achieved accuracy is 93.8% and time is 11.427 (s). This showed that the accuracy is improved by S2SL. However, the proposed framework accuracy is 96.5%, which is significantly

TABLE 2 Proposed multiclass skin lesion classification for HAM10000 data set

Classifier	Evaluation protocols				
	Sen (%)	Prec (%)	F1-Score (%)	Acc (%)	Time (s)
MCELM	**89.01**	**92.72**	**90.83**	**96.5**	**10.761**
MCSVM	74.68	97.47	84.57	94.6	48.936
GN Bayes	65.37	84.21	73.60	92.9	50.486
W-KNN	78.40	89.57	83.61	93.0	28.349
Softmax	68.87	82.51	75.08	93.6	22.877
EBT	70.60	95.20	81.08	92.4	18.298

*Note*: Bold values represents best values.

Abbreviations: Acc, accuracy; EBT, ensemble baggage tree; GN, Gaussian naïve; MCELM, multiclass extreme learning machine; MCSVM, multiclass support vector machines; Prec, precision; Sen, sensitivity; W-KNN, weighted k-nearest neighbors.

TABLE 3 Stepwise propose method accuracy and time comparison for HAM10000 data set

Method	Features stream			Measure	
	S1SL	S2SL	Final	Acc (%)	Time (s)
MCELM	√			92.3	**9.446**
		√		93.8	11.427
			√	**96.5**	10.761

*Note*: Bold values represents best values.

Abbreviations: Acc, accuracy; MCELM, multiclass extreme learning machine; S1SL, Stream 1 selected feature; S2SL, Stream 2 selected feature.
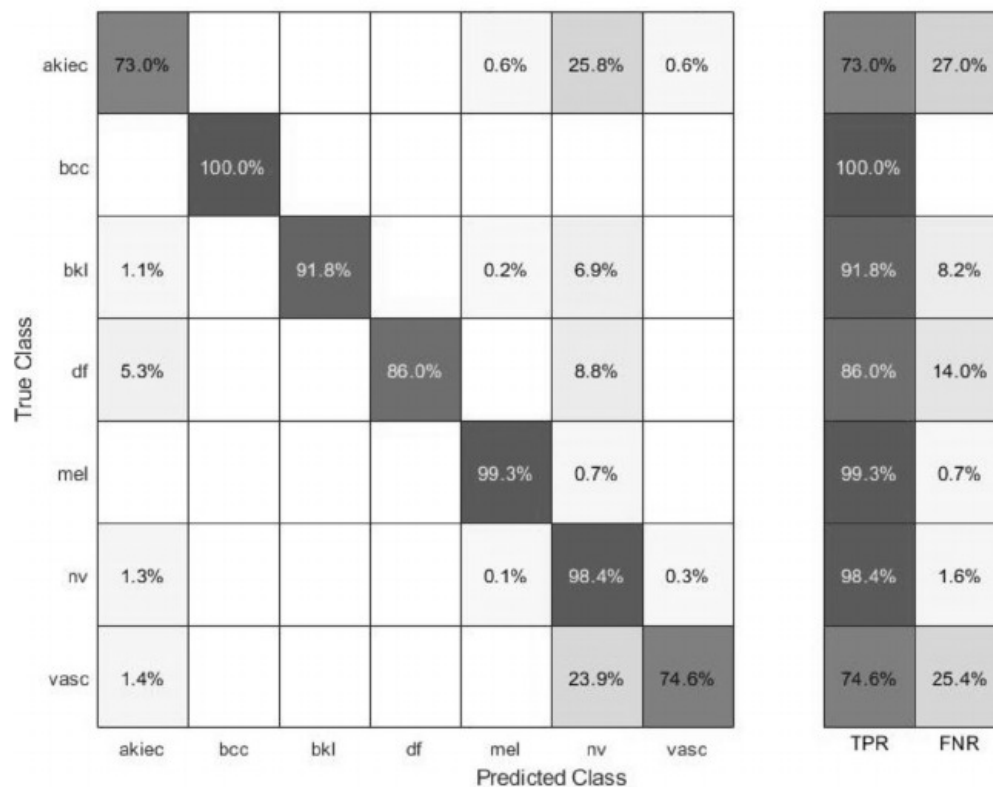
FIGURE 11 Confusion matrix of MCELM for HAM10000 data set using the proposed methodology. akiec, actinic keratoses; bcc, basal cell carcinoma; bkl, benign keratosis; df, dermatofibroma; MCELM, multiclass extreme learning machine; mel, melanoma; nv, nevi; vasc, vascular [Color figure can be viewed at wileyonlinelibrary.com]

improved. In Figure 11, confusion matrix shows that the akiec and vasc skin classes having low accuracy, which impact on the classification performance. Additionally, the proposed framework's accuracy is also compared with a few well-known neural nets as illustrated in Figure 12. On the basis of this figure, one can clearly observe the performance of the proposed framework.

Table 4 presents the classification results of on ISBI2018 data set. In this table, it is presented that MCELM provides better results with a sensitivity rate of 90.21%, a precision rate of 98.60%, F1-score is 94.22%, and an accuracy of 98%. The best precision rate is 97.47% achieved by MCELM and for the rest of classifiers, the precision rate is 97.72%, 92.20%, 96.04%, 95.44%, and 95.17%, respectively (can be seen in Table 4).
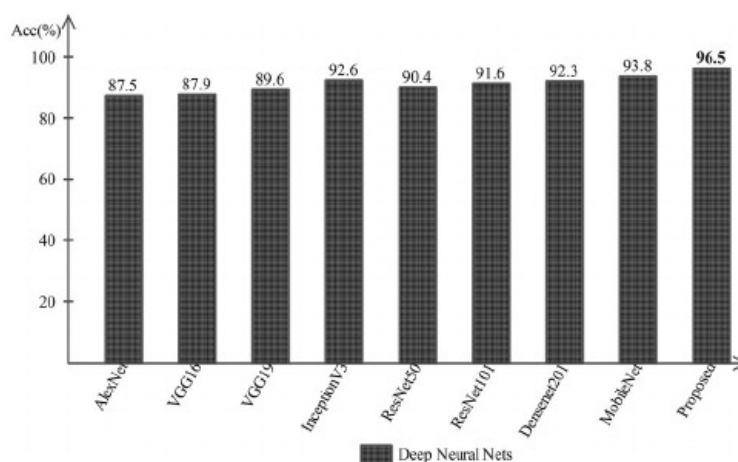
FIGURE 12 Proposed method accuracy compared with individual deep neural nets. All networks features are validated through feature selection step. Acc, accuracy

TABLE 4 Proposed multiclass skin lesion classification for ISBI2018 data set

Classifier	Evaluation protocols				
	Sen (%)	Prec (%)	F1-Score (%)	Acc (%)	Time (s)
MCELM	**90.21**	**98.60**	**94.22**	**98.0**	46.751
MCSVM	81.8	97.72	89.05	96.5	55.715
GN Bayes	84.4	92.20	88.13	95.0	22.629
W-KNN	77.78	96.04	85.95	94.2	22.284
Softmax	81.84	95.44	88.12	95.5	**19.251**
EBT	77.12	95.17	85.20	93.9	52.730

Note: Bold values represents best values.

Abbreviations: Acc, accuracy; EBT, ensemble baggage tree; GN, Gaussian naïve; MCELM, multiclass extreme learning machine; MCSVM, multiclass support vector machines; Prec, precision; Sen, sensitivity; W-KNN, weighted *k*-nearest neighbors.

The highest F1-Score is 94.22% obtained by MCELM. For the rest of classifiers, the F1-Score is 89.05%, 88.13%, 85.95%, 88.12%, and 85.20%, respectively (can be seen in Table 4). The minimum sensitivity rate of this data set is 77.12%, obtained by the EBT classifier. Moreover, the computational time is also noted and minimum time is 19.251 (s) for the Softmax classifier. The highest time is 55.715 (s) for the MCSVM classifier. The proposed accuracy of MCELM is also compared with individual streams as illustrated in Figure 3. The individual results are given in Table 5. In this table, the results are computed for S1SL and S2SL. The achieved accuracy of S1SL is 93.9% and the noted time is 36.116 (s). For the S2SL achieved accuracy is 94.6% and time is 40.204 (s). On the basis of this accuracy, the S2SL showed improved accuracy but a minor time is increased. However, the proposed framework accuracy is 98.0% and significantly better as compared with S1SL and S2SL. The performance of proposed accuracy is also proved in Figure 13. In this figure, a confusion matrix is illustrated which shows that the DF and VASC skin classes having high misclassification. But it showed only a little impact on the classification performance because these classes have fewer images as compared with other classes, like, melanoma, akiec, and NV. In the addition, proposed framework accuracy is also

TABLE 5 Stepwise propose method accuracy and time comparison for ISBI2018 data set

| Method	Features stream			Measure	
	S1SL	S2SL	Final	Acc (%)	Time (s)
MCELM				93.9	**36.116**
				94.6	40.204
				98.0	46.751

Note: Bold values represents best values.

Abbreviations: Acc, accuracy; MCELM, multiclass extreme learning machine; S1SL, Stream 1 selected feature; S2SL, Stream 2 selected feature.
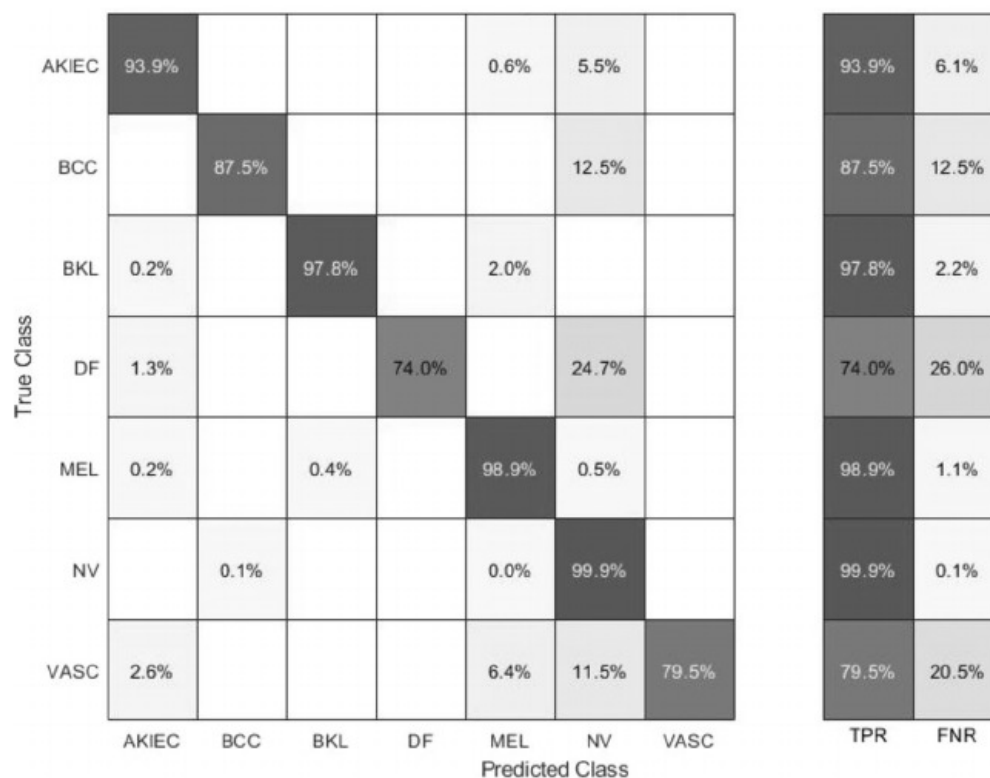
FIGURE 13 Confusion matrix of MCELM for ISBI2018 data set using the proposed methodology. akiec, actinic keratoses; bcc, basal cell carcinoma; bkl, benign keratosis; df, dermatofibroma; MCELM, multiclass extreme learning machine; mel, melanoma; nv, nevi; vasc, vascular [Color figure can be viewed at wileyonlinelibrary.com]

compared with few well-known neural nets as illustrated in Figure 14. On the basis of this figure, we clearly say that our proposed method outperformed.

Similar to HAM10000 and ISBI2018, the results of the ISBI2019 data set are presented in Table 6. In this table, it is presented that the best-achieved accuracy is 89% by the MCELM classifier. The sensitivity and precision rate of the MCELM are 79.62% and 88.25%. This classifier is executed in 122.621 (s) but the minimum noted time is 102.176 (s) for the Softmax classifier. The MCELM accuracy is further compared with few other classifiers as listed in Table 6. In this table, it is noted that the Softmax achieved the second-best accuracy of 88.5%. The rest of classifiers accuracies are 86.5%, 82.15, 85.0%, and 85.1%, respectively. Table 7 represents the comparison of both streams' accuracy and time with the proposed performance.
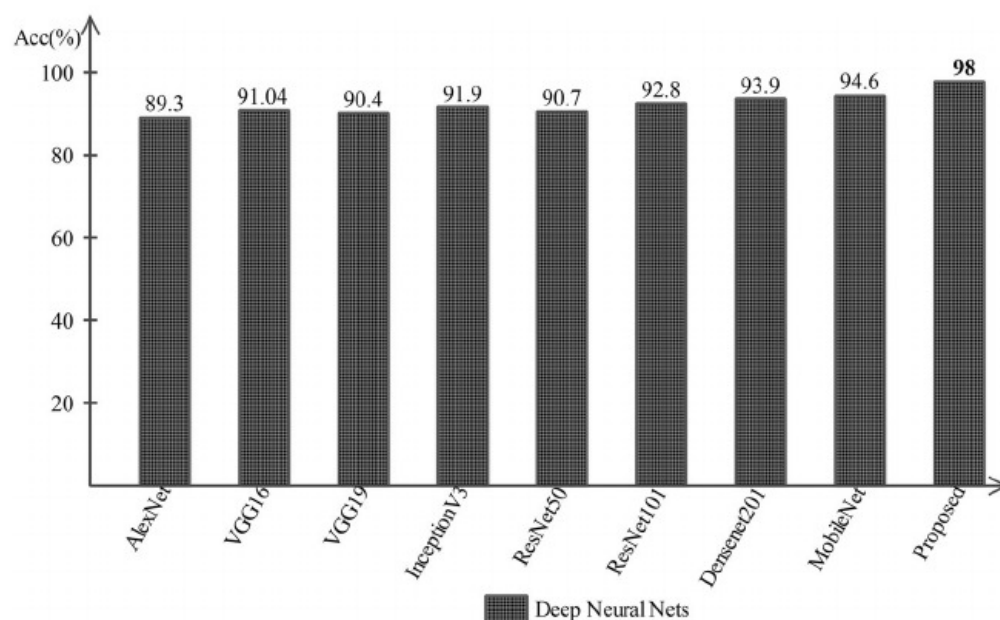
FIGURE 14   Proposed method accuracy compared with individual deep neural nets. Acc, accuracy

TABLE 6   Proposed multiclass skin lesion classification for ISBI2019 data set

| Classifier | Evaluation protocols | | | | |
	Sen (%)	Prec (%)	F1-Score (%)	Acc (%)	Time (s)
MCELM	**79.62**	**88.25**	**83.71**	**89.0**	122.621
MCSVM	69.18	86.80	76.99	86.5	141.214
GN Bayes	68.90	77.70	73.04	82.1	108.06
W-KNN	78.56	80.38	79.46	85.0	**97.114**
Softmax	79.46	87.08	83.10	88.3	102.176
EBT	71.48	83.70	77.11	85.1	123.230

Note: Bold values represents best values.

Abbreviations: Acc, accuracy; EBT, ensemble baggage tree; GN, Gaussian naïve; MCELM, multiclass extreme learning machine; MCSVM, multiclass support vector machines; Prec, precision; Sen, sensitivity; W-KNN, weighted *k*-nearest neighbors.

TABLE 7   Stepwise propose method accuracy and time comparison for ISBI2019 data set

| Method | Features stream | | | Measure | |
	S1SL	S2SL	Final	Acc (%)	Time (s)
MCELM				82.6	**96.686**
				83.1	104.523
				89.0	122.620

Note: Bold values represents best values.

Abbreviations: Acc, accuracy; MCELM, multiclass extreme learning machine; S1SL, Stream 1 selected feature; S2SL, Stream 2 selected feature.

From this table, it is showed that the proposed accuracy is significantly better compared with both streams (can be verified through Figure 15). In addition, the proposed accuracy is also compared with other neural nets, as illustrated in Figure 16. This figure shows the significance of the proposed accuracy.

6.1 | Analysis and comparison

In Section 5, results are discussed and showed that the proposed two-stream architecture works better for the MCELM classifier. Also, this architecture gives better performance for imbalanced data sets. The comparison of the proposed architecture for each data set is also compared with other neural nets as illustrated in Figures 12, 14, and 16. Moreover, we also conducted an analysis to analyze the performance of the proposed architecture on MCELM. For this purpose, we performed 100 iterations and get three values—minimum accuracy, average accuracy, and maximum accuracy, as plotted in Figure 17. In this figure, it is described that the proposed method the accuracy for HAM10000 and ISBBI2018 is consistent as compared with accuracy plotted of ISIC2019. Overall, we can say that the proposed method is useful for huge changes in features and imbalanced data sets.

Table 8 presents the results of different relation coefficient values. In this table, it is shown that the proposed method performance is better for 0.4. The other values, such as 0.3 and 0.5, also give the second and third best accuracies. Overall, this table concludes that the increases in
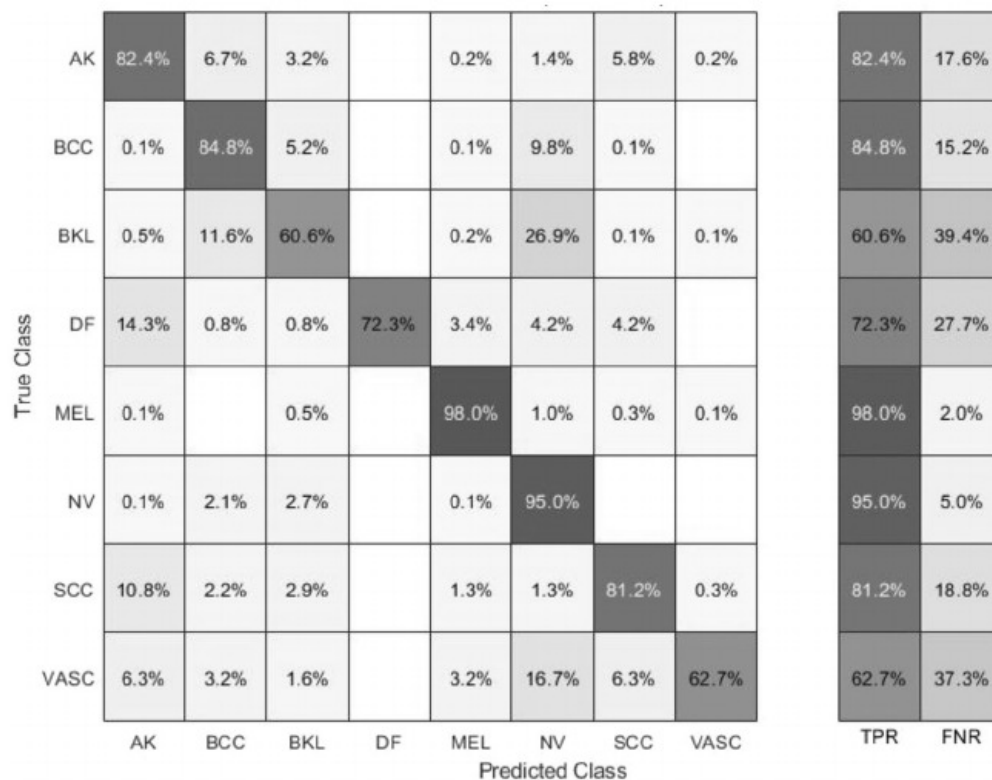
FIGURE 15 Confusion matrix of MCELM for ISBI2019 data set using the proposed methodology. akiec, actinic keratoses; bcc, basal cell carcinoma; bkl, benign keratosis; df, dermatofibroma; MCELM, multiclass extreme learning machine; mel, melanoma; nv, nevi; vasc, vascular [Color figure can be viewed at wileyonlinelibrary.com]
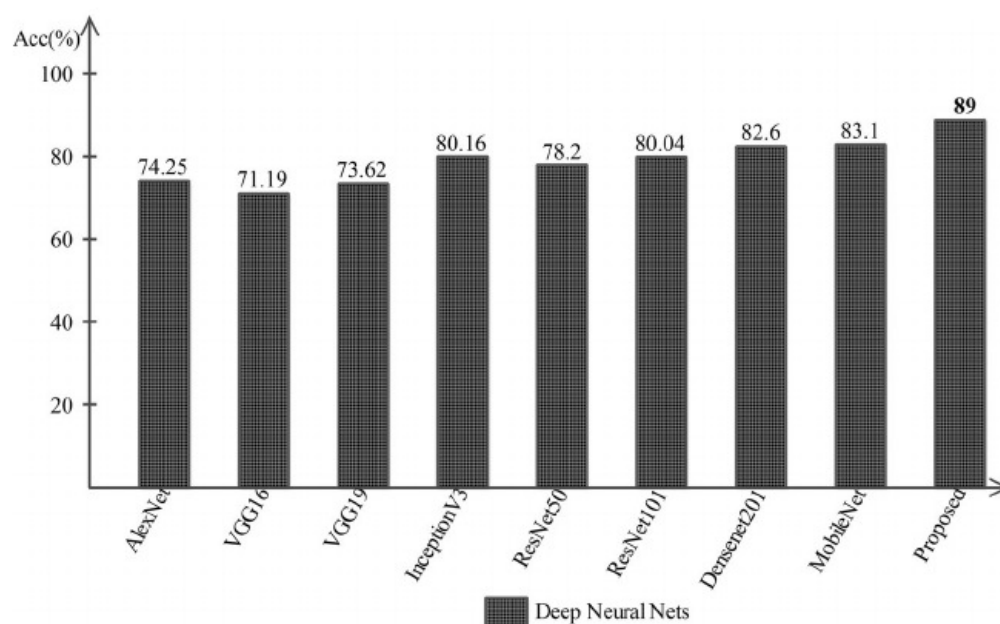
FIGURE 16 Proposed method accuracy compared with individual deep neural nets for ISBI2019. Acc, accuracy
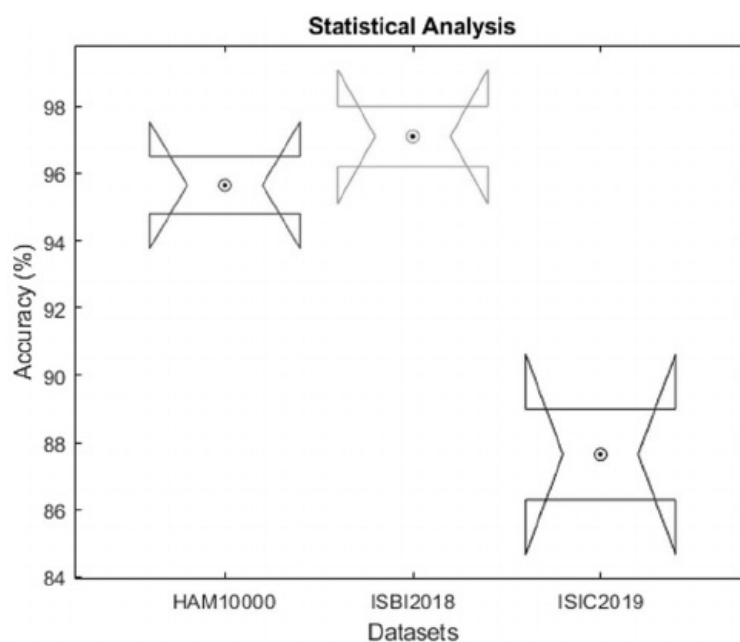

FIGURE 17 Iterations-based analysis of MCELM accuracy on selected data sets. MCELM, multiclass extreme learning machine [Color figure can be viewed at wileyonlinelibrary.com]

TABLE 8 Propose classification results using different relation coefficient values

|	Relation coefficient value					
Data set	0.2	0.3	0.4	0.5	0.6	0.7
HAM10000	93.4	94.7	**96.5**	94.3	93.6	93.9
ISIC2018	96.1	95.0	**98.0**	96.4	95.8	95.3
ISIC2019	87.8	85.6	**89.0**	88.4	87.1	86.0

Note: Bold values represents best values.

the value of the relation coefficient are decreases in results. Hence, the value 0.4 is showing better results for all three selected data sets.

Figures 18 and 19 show the feature-based comparison of the proposed method. Figure 18 illustrates the accuracy-based comparison of individual extracted deep features, best selected deep features, and final fused features. In this figure, the difference among individual features and best-selected feature is almost 4%–5%. Moreover, accuracy difference among individual and final fused feature is almost 6%–7%. Figure 19 illustrates that the features before reduction consumes more time as compared with feature after reduction. Overall, these figures (Figures 18 and 19) show the importance of feature reduction in terms of accuracy and time.

Table 9 presents the comparison of MMFO feature selection algorithm with a few other existing selection techniques, such as Auto-Encoders, Genetic Algorithm (GA), Particle Swarm Optimization (PSO), and Ant Colony. In this table, the S1SL and S2SL represent DenseNet201 and MobileNetV2 best-selected feature vectors. The auto-encoder-based features are represented by S1 Auto-Encoder and S2 Auto-Encoder. On these two selected vectors, the achieved accuracies are $(88.4, 90.3)$, $(79.3, 80.4)$, and $(87.6, 88.7)$, respectively. Second, GA is applied for best feature selection and obtained an accuracy for both vectors (S1-GA and S2-GA) of $(86.5, 89.1)$, $(76.5, 78.4)$, and $(84.3, 84.7)$, respectively. From these accuracy values, it is noted that the performance of GA is low compared with Auto-Encoder feature selection. Similarly, PSO- and ant colony optimization (ACO)-based feature selection results are presented and achieved lower accuracy as compared with GA and Auto-Encoder. Overall, the proposed MMFO selection algorithm gives better results. The main motivation behind the feature selection through MMFO is two-step process; however, the disadvantage is an increase in computational time.

In the last, we compared the proposed method accuracy with some recent techniques as presented in Table 9. On the basis of this table, it is showing that the proposed two-stream deep architecture accuracy is improved as compared with existing techniques (Table 10).
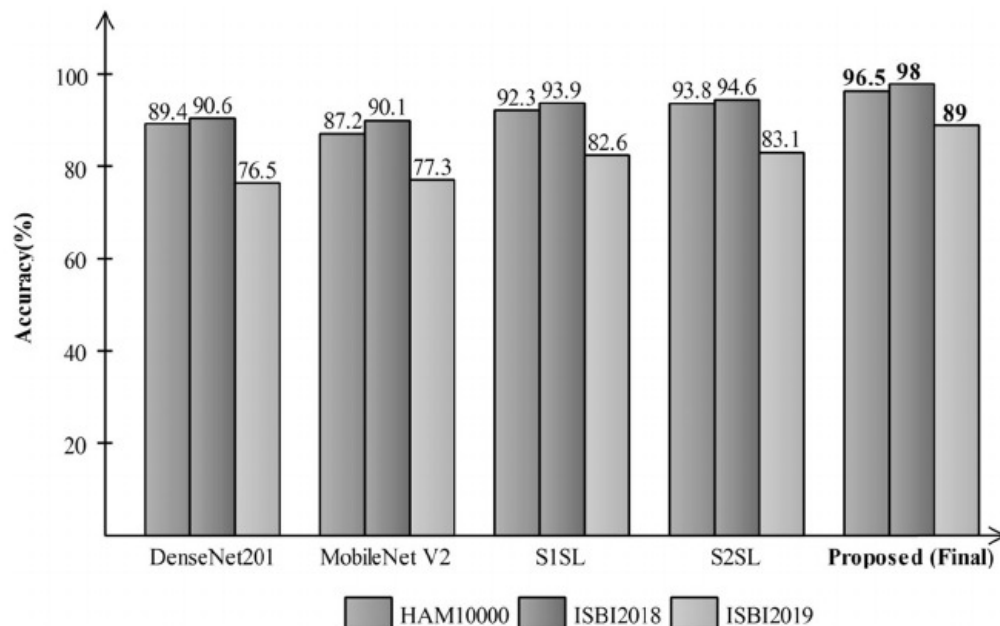
FIGURE 18 Accuracy-based comparison of selected features (S1SL and S2SL) with without feature reduction (DenseNet201 features and MobileNetV2 features). S1SL, Stream 1 selected feature; S2SL, Stream 2 selected feature [Color figure can be viewed at wileyonlinelibrary.com]
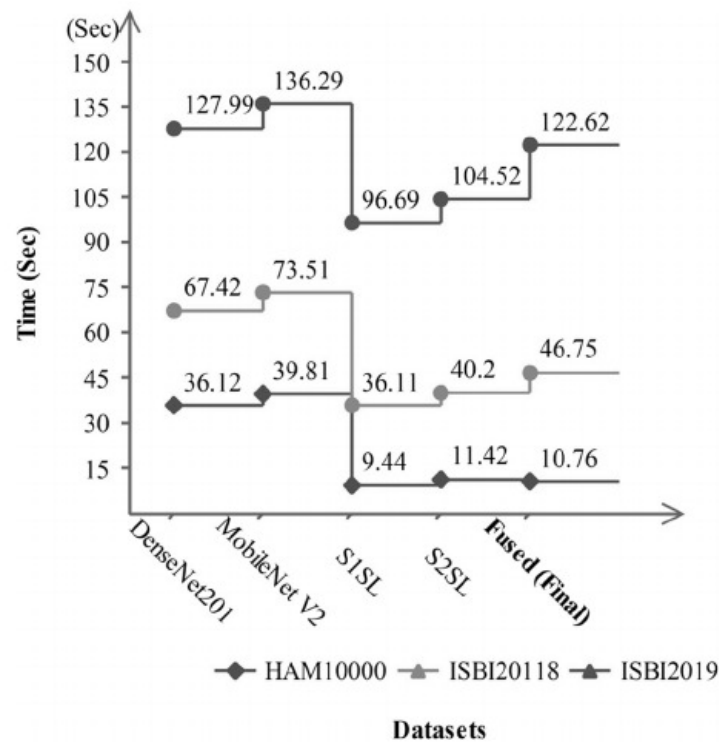
FIGURE 19 Time-based comparison of selected features (S1SL and S2SL) with without feature reduction (DenseNet201 features and MobileNetV2 features). S1SL, Stream 1 selected feature; S2SL, Stream 2 selected feature [Color figure can be viewed at wileyonlinelibrary.com]

TABLE 9 Comparison of proposed MMFO feature selection technique with some other techniques

	Data set		
Feature selection technique	**ISBI2018 (%)**	**ISBI2019 (%)**	**HAM10000 (%)**
S1SL	**93.9**	**82.6**	**92.3**
S1 Auto-Encoder	88.4	79.3	87.6
S2SL	**94.6**	**83.1**	**93.8**
S2 Auto-Encoder	90.3	80.4	88.7
S1-GA	86.5	76.5	84.3
S2-GA	89.1	78.4	84.7
S1-PSO	88.8	77.9	82.9
S2-PSO	85.5	79.3	83.1
S1-ACO	84.4	75.8	83.8
S2-ACO	87.2	74.4	84.5

Note: S1 represents DenseNet201 deep feature vector; S2 represents MobileNetV2 deep feature vector. Bold values represents best values.

Abbreviations: ACO, ant colony optimization; GA, Genetic Algorithm; MMFO, modified moth–flame optimization; PSO, Particle Swarm Optimization; S1SL, Stream 1 selected feature; S2SL, Stream 2 selected feature.

TABLE 10 Proposed two-stream architecture accuracy comparison with existing techniques

Method	Data set	Evaluation measures (accuracy, %)
[56]	HAM10000	85.8
[57]	HAM10000	87.7
[58]	HAM10000	93.0
[59]	HAM10000	92.70
[60]	HAM10000	92.83
[61]	ISBI2018	89.05
[62]	ISBI2018	92.4
[63]	ISBI2018	93.4
Proposed	**ISBI2018**	**98.0**
Proposed	**ISBI2019**	**89.0**
Proposed	**HAM10000**	**96.5**

Note: Bold values represents best values.

7 | CONCLUSION

Data analytics in healthcare is one of the most recent domains, where ML concepts have been effectively implemented to get the productive results. In this study, we presented two-stream deep learning architecture by employing fusion-based contrast enhancement dermoscopic images. The proposed architecture is trained on imbalanced skin classes instead of balance classes. In the first stream, a fine-tuned pretrained DenseNet201 model is trained using transfer learning and optimized features using a modified MFO algorithm. In the second stream, a fine-tuned pretrained Mobile-NetV2 is employed and trained using TL. The extracted features of this stream are also optimized using modified MFO and select the best features. The selected features of both streams are fused by employing a new approach name Parallel Max Correlation Coefficient. Finally, the MCELM classifier is employed for the final classification. We compare the results of each stream. Also, we conducted a comparison for both selected vectors from both streams. However, the fusion of both vectors improves the classification accuracy. Furthermore, the feature selection process minimizes the computational time and also increases the classification rate. This process can be useful in the clinics for dermatologists as a second opinion. However, it has been observed that the fusion process improves the accuracy, but on the other side increases the computational burden. Similarly, the addition of the segmentation step further refines the classification results, which was not the focus of this study. Therefore, as future work, both limitations will be addressed by adding improved segmentation and feature fusion steps.

CONFLICT OF INTERESTS
The authors declare that there are no conflict of interests.

ORCID
Muhammad Attique Khan ⓘ https://orcid.org/0000-0001-6763-2123
Muhammad Sharif ⓘ https://orcid.org/0000-0002-9214-4292

Tallha Akram ⓘ https://orcid.org/0000-0003-4578-3849
Seifedine Kadry ⓘ https://orcid.org/0000-0002-4939-9900
Ching-Hsien Hsu ⓘ https://orcid.org/0000-0002-2440-2771

REFERENCES

1. Song L, Lin JP, Wang ZJ, Wang H. An end-to-end multi-task deep learning framework for skin lesion analysis. *IEEE J Biomed Health Inf.* 2020;24:2912-2921.
2. Andersson K, Waterboer T, Kirnbauer R, et al. Seroreactivity to cutaneous human papillomaviruses among patients with nonmelanoma skin cancer or benign skin lesions. *Cancer Epidemiol Prev Biomarkers.* 2008;17: 189-195.
3. Simon S. (ed.). Facts & figures 2019: US cancer death rate has dropped 27% in 25 years, 2019 (/cancer/acs-medical-content-and-news-sta.html).
4. Chaturvedi SS, Gupta K, Prasad PS. Skin lesion analyser: an efficient seven-way multi-class skin cancer classification using MobileNet. In: Hassanien A, Bhatnagar R, Darwish A, eds. *Advanced Machine Learning Technologies and Applications. AMLTA 2020. Advances in Intelligent Systems and Computing.* Vol 1141. Springer; 2021. https://doi-org.ezproxy.bu.edu/10.1007/978-981-15-3383-9_15
5. Argenziano G, Giacomel J, Zalaudek I, et al. A clinico-dermoscopic approach for skin cancer screening: recommendations involving a survey of the International Dermoscopy Society. *Dermatol Clin.* 2013;31: 525-534.
6. Nasir M, Attique Khan M, Sharif M, Lali IU, Saba T, Iqbal T. An improved strategy for skin lesion detection and classification using uniform segmentation and feature selection based approach. *Microsc Res Tech.* 2018;81:528-543.
7. Barata C, Celebi ME, Marques JS. Explainable skin lesion diagnosis using taxonomies. *Pattern Recognit.* 2020;110:107413.
8. Conforti C, Giuffrida R, Vezzoni R, Resende FS, di Meo N, Zalaudek I. Dermoscopy and the experienced clinicians. *Int J Dermatol.* 2020;59:16-22.
9. Kittler H, Pehamberger H, Wolff K, Binder M. Diagnostic accuracy of dermoscopy. *Lancet Oncol.* 2002;3: 159-165.
10. Robinson JK, Ortiz S. Use of photographs illustrating ABCDE criteria in skin self-examination. *Arch Dermatol.* 2009;145:332-333.
11. Soyer HP, Argenziano G, Zalaudek I, et al. Three-point checklist of dermoscopy. *Dermatology.* 2004;208: 27-31.
12. Argenziano G, Fabbrocini G, Carli P, De Giorgi V, Sammarco E, Delfino M. Epiluminescence microscopy for the diagnosis of doubtful melanocytic skin lesions: comparison of the ABCD rule of dermatoscopy and a new 7-point checklist based on pattern analysis. *Arch Dermatol.* 1998;134:1563-1570.
13. Stacke K, Eilertsen G, Unger J, Lundstrom C. Measuring domain shift for deep learning in histopathology. *IEEE J Biomed Health Inf.* 2020;25:325-336.
14. Dissanayake T, Fernando T, Denman S, Sridharan S, Ghaemmaghami H, Fookes C. A robust interpretable deep learning classifier for heart anomaly detection without segmentation. *IEEE J Biomed Health Inf.* 2020; 25:2162-2171.
15. Doi K. Computer-aided diagnosis in medical imaging: historical review, current status and future potential. *Comput Med Imaging Graph.* 2007;31:198-211.
16. Guaragnella C, Rizzi M. Simple and accurate border detection algorithm for melanoma computer aided diagnosis. *Diagnostics.* 2020;10:423.
17. Panayides AS, Amini A, Filipovic ND, et al. AI in medical imaging informatics: current challenges and future directions. *IEEE J Biomed Health Inf.* 2020;24(2020):1837-1857.
18. Al-Masni MA, Kim D-H, Kim T-S. Multiple skin lesions diagnostics via integrated deep convolutional networks for segmentation and classification. *Comput Methods Programs Biomed.* 2020;190:105351.
19. Mahbod A, Schaefer G, Wang C, Dorffner G, Ecker R, Ellinger I. Transfer learning using a multi-scale and multi-network ensemble for skin lesion classification. *Comput Methods Programs Biomed.* 2020;193:105475.
20. Elia S, D'Angelo G, Palmieri F, et al. A machine learning evolutionary algorithm-based formula to assess tumor markers and predict lung cancer in cytologically negative pleural effusions. *Soft Comput.* 2019;24:1-13.

21. Khan MA, Majid A, Akram T, et al. Classication of COVID-19 CT scans via extreme learning machine. *Comput Mater Continua*. 2021;68:1003-1019.

22. D'Angelo G, Palmieri F. Enhancing COVID-19 tracking apps with human activity recognition using a deep convolutional neural network and HAR-images. *Neural Comput Appl*. 2021:1-17.

23. D'Angelo G, Tipaldi M, Glielmo L, Rampone S. Spacecraft autonomy modeled via Markov decision process and associative rule-based machine learning. In: *2017 IEEE International Workshop on Metrology for Aerospace (MetroAeroSpace)*; 2017:324-329.

24. D'Angelo G, Palmieri F. Knowledge elicitation based on genetic programming for non destructive testing of critical aerospace systems. *Future Gener Comput Syst*. 2020;102:633-642.

25. Xie Y, Zhang J, Xia Y, Shen C. A mutual bootstrapping model for automated skin lesion segmentation and classification. *IEEE Trans Med Imaging*. 2020;39:2482-2493.

26. Sikkandar MY, Alrasheadi BA, Prakash N, Hemalakshmi G, Mohanarathinam A, Shankar K. Deep learning based an automated skin lesion segmentation and intelligent classification model. *J Ambient Intell Hum Comput*. 2020;12:3245-3255.

27. Mahbod A, Tschandl P, Langs G, Ecker R, Ellinger I. The effects of skin lesion segmentation on the performance of dermatoscopic image classification. *Comput Methods Programs Biomed*. 2020;197:105725.

28. Khan MA, Akram T, Sharif M, Javed K, Rashid M, Bukhari SAC. An integrated framework of skin lesion detection and recognition through saliency method and optimal deep neural network features selection. *Neural Comput Appl*. 2019;32:15929-15948.

29. Feng R, Liu X, Chen J, Chen DZ, Gao H, Wu J. A deep learning approach for colonoscopy pathology WSI analysis: accurate segmentation and classification. *IEEE J Biomed Health Inf*. 2020.

30. Adegun A, Viriri S. Deep learning techniques for skin lesion analysis and melanoma cancer detection: a survey of state-of-the-art. *Artif Intell Rev*. 2020;54:811-841.

31. Al Nazi Z, Abir TA. Automatic skin lesion segmentation and melanoma detection: transfer learning approach with U-net and DCNN-SVM. In: *Proceedings of International Joint Conference on Computational Intelligence*; 2020:371-381.

32. Garg S, Jindal B. Skin lesion segmentation using k-mean and optimized fire fly algorithm. *Multimedia Tools Appl*. 2020;80:7397-7410.

33. Miglani V, Bhatia M. Skin lesion classification: a transfer learning approach using EfficientNets. In: *International Conference on Advanced Machine Learning Technologies and Applications*; 2020:315-324.

34. Iqbal I, Younus M, Walayat K, Kakar MU, Ma J. Automated multi-class classification of skin lesions through deep convolutional neural network with dermoscopic images. *Comput Med Imaging Graph*. 2021; 88:101843.

35. Cano E, Mendoza-Avilés J, Areiza M, Guerra N, Mendoza-Valdés JL, Rovetto CA. Multi skin lesions classification using fine-tuning and data-augmentation applying NASNet. *PeerJ Comput Sci*. 2021;7:e371.

36. López-Leyva JA, Guerra-Rosas E, Álvarez-Borrego J. Multi-class diagnosis of skin lesions using the fourier spectral information of images on additive color model by artificial neural network. *IEEE Access*. 2021;9: 35207-35216.

37. Lima DM, Rodrigues-Jr JF, Brandoli B, Goeuriot L, Amer-Yahia S. DermaDL: advanced convolutional neural networks for computer-aided skin-lesion classification. *SN Comput Sci*. 2021;2:1-13.

38. Hameed N, Shabut AM, Ghosh MK, Hossain MA. Multi-class multi-level classification algorithm for skin lesions classification using machine learning techniques. *Expert Syst Appl*. 2020;141:112961.

39. Hassan SR, Afroge S, Mizan MB. Skin lesion classification using densely connected convolutional network. In: *2020 IEEE Region 10 Symposium (TENSYMP)*; 2020:750-753.

40. Khan MA, Sharif M, Akram T, Bukhari SAC, Nayak RS. Developed Newton–Raphson based deep features selection framework for skin lesion recognition. *Pattern Recognit Lett*. 2020;129:293-303.

41. 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*. 2019:e12497.

42. Khan MA, Muhammad K, Sharif M, Akram T, de Albuquerque VHC. Multi-class skin lesion detection and classification via teledermatology. *IEEE J Biomed Health Inf*. 2021.

43. Tschandl P, Rosendahl C, Kittler H. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Sci Data*. 2018;5:180161.

44. Codella N, Rotemberg V, Tschandl P, et al. Skin lesion analysis toward melanoma detection 2018: a challenge hosted by the international skin imaging collaboration (ISIC). 2019. arXiv preprint arXiv:1902.03368.

45. Kassem MA, Hosny KM, Fouad MM. Skin lesions classification into eight classes for ISIC 2019 using deep convolutional neural network and transfer learning. *IEEE Access*. 2020;8:114822-114832.

46. Chen S-D, Ramli AR. Contrast enhancement using recursive mean-separate histogram equalization for scalable brightness preservation. *IEEE Trans Consum Electron*. 2003;49:1301-1309.

47. Lee C-Y, Xie S, Gallagher P, Zhang Z, Tu Z. Deeply-supervised nets. In: *Artificial Intelligence and Statistics*; 2015:562-570.

48. Glorot X, Bordes A, Bengio Y. Deep sparse rectifier neural networks. In: *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*; 2011:315-323.

49. LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. *Proc IEEE*. 1998;86:2278-2324.

50. Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift. 2015. arXiv preprint arXiv:1502.03167.

51. Mirjalili S. Moth–flame optimization algorithm: a novel nature-inspired heuristic paradigm. *Knowl-Based Syst*. 2015;89:228-249.

52. Shehab M, Abualigah L, Al Hamad H, Alabool H, Alshinwan M, Khasawneh AM. Moth–flame optimization algorithm: variants and applications. *Neural Comput Appl*. 2020;32:9859-9884.

53. Yuan Y-H, Sun Q-S, Zhou Q, Xia D-S. A novel multiset integrated canonical correlation analysis framework and its application in feature fusion. *Pattern Recognit*. 2011;44:1031-1040.

54. Huang G-B, Zhou H, Ding X, Zhang R. Extreme learning machine for regression and multiclass classification. *IEEE Trans Syst Man Cybern Part B (Cybern)*. 2011;42:513-529.

55. Martins B, Ferreira D, Neto C, Abelha A, Machado J. Data mining for cardiovascular disease prediction. *J Med Syst*. 2021;45:1-8.

56. Huang HW, Hsu BWY, Lee CH, Tseng VS. Development of a light-weight deep learning model for cloud applications and remote diagnosis of skin cancers. *J Dermatol*. 2020;48(3):310-316.

57. Thurnhofer-Hemsi K, Domínguez E. A convolutional neural network framework for accurate skin cancer detection. *Neural Process Lett*. 2020:1-21.

58. Le DN, Le HX, Ngo LT, Ngo HT. Transfer learning with class-weighted and focal loss function for automatic skin cancer classification. 2020. arXiv preprint arXiv:2009.05977.

59. Mohamed EH, El-Behaidy WH. Enhanced skin lesions classification using deep convolutional networks. In: *2019 Ninth International Conference on Intelligent Computing and Information Systems (ICICIS)*; 2019: 180-188.

60. Chaturvedi SS, Tembhurne JV, Diwan T. A multi-class skin cancer classification using deep convolutional neural networks. *Multimedia Tools Appl*. 2020;79:28477-28498.

61. Shahin AH, Kamal A, Elattar MA. Deep ensemble learning for skin lesion classification from dermoscopic images. In: *2018 9th Cairo International Biomedical Engineering Conference (CIBEC)*; 2018:150-153.

62. Almaraz-Damian J-A, Ponomaryov V, Sadovnychiy S, Castillejos-Fernandez H. Melanoma and nevus skin lesion classification using handcraft and deep learning feature fusion via mutual information measures. *Entropy*. 2020;22:484.

63. Zhang J, Xie Y, Xia Y, Shen C. Attention residual learning for skin lesion classification. *IEEE Trans Med Imaging*. 2019;38:2092-2103.