Discover Applied Sciences

Research

Multiclass skin lesion classification using deep learning networks optimal information fusion

Muhammad Attique Khan[1] · Ameer Hamza[1] · Mohammad Shabaz[2] · Seifeine Kadry[3] · Saddaf Rubab[4] · Muhammad Abdullah Bilal[5] · Muhammad Naeem Akbar[6] · Suresh Manic Kesavan[7]

Received: 15 April 2024 / Accepted: 27 May 2024
Published online: 30 May 2024

Abstract
A serious, all-encompassing, and deadly cancer that affects every part of the body is skin cancer. The most prevalent causes of skin lesions are UV radiation, which can damage human skin, and moles. If skin cancer is discovered early, it may be adequately treated. In order to diagnose skin lesions with less effort, dermatologists are increasingly turning to machine learning (ML) techniques and computer-aided diagnostic (CAD) systems. This paper proposes a computerized method for multiclass lesion classification using a fusion of optimal deep-learning model features. The dataset used in this work, ISIC2018, is imbalanced; therefore, augmentation is performed based on a few mathematical operations. After that, two pre-trained deep learning models (DarkNet-19 and MobileNet-V2) have been fine-tuned and trained on the selected dataset. After training, features are extracted from the average pool layer and optimized using a hybrid firefly optimization technique. The selected features are fused in two ways: (i) original serial approach and (ii) proposed threshold approach. Machine learning classifiers are used to classify the chosen features at the end. Using the ISIC2018 dataset, the experimental procedure produced an accuracy of 89.0%. Whereas, 87.34, 87.57, and 87.45 are sensitivity, precision, and F1 score respectively. At the end, comparison is also conducted with recent techniques, and it shows the proposed method shows improved accuracy along with other performance measures.

Article highlights

- The most prevalent cause of skin lesions is UV radiation.
- To diagnose skin lesions with less effort, dermatologists are increasingly turning to machine learning (ML) techniques.
- A computerized method for multiclass lesion classification using a fusion of optimal deep-learning model features is proposed.
- The experimental procedure produced an accuracy of 89.0%.

Keywords Skin cancer · Dermoscopy · Augmentation · Deep learning · Optimization · Fusion

✉ Mohammad Shabaz, bhatsab4@gmail.com; Muhammad Attique Khan, attique.khan@hitecuni.edu.pk; Ameer Hamza, ameer.hamza@hitecuni.edu.pk; Seifeine Kadry, seifedine.kadry@noroff.no; Saddaf Rubab, srubab@sharjah.ac.ae; Muhammad Abdullah Bilal, mabilal.bese19seecs@seecs.edu.pk; Muhammad Naeem Akbar, chmuhammadnaeemakbar@gmail.com; Suresh Manic Kesavan, sureshmanic@nu.edu.om | [1]Department of Computer Science, HITEC University, Taxila, Pakistan. [2]Model Institute of Engineering and Technology, Jammu, J&K, India. [3]Department of Applied Data Science, Noroff University College, Oslo, Norway. [4]Department of Computer Engineering, College of Computing and Informatics, University of Sharjah, 27272, Sharjah, United Arab Emirates. [5]Department of CS, SEECS NUST, Islamabad, Pakistan. [6]National University of Sciences and Technology (NUST), Islamabad 46000, Pakistan. [7]Department of Electrical and Communication Engineering, National University of Science and Technology, Muscat, Sultanate of Oman.

1 Introduction

Skin cancer is a widespread, deadliest, and considered a most common type of cancer, which has increased remarkably in the last few years [1]. The abnormal proliferation of skin cells that can become part of the body is known as skin cancer. Harmful Ultraviolet (UV) rays of the sun and the use of UV tanning beds damage human skin with time and cause the growth of cancer cells [2]. A skin lesion is a serious and deadliest form of cancer, which is the irregular development of cells that affects any part of the human body [3]. Melanoma is among the worst types of skin cancer, with a survival rate of about 5% [4]. The expansion of skin cancer has increased rapidly in the last few years. In the USA, almost 5.4 million people are diagnosed with skin cancer each year. Melanomas are responsible for around 35% of all skin cancer mortality, with over ten thousand deaths occurring annually in the United States alone [5]. As per the World Health Organization (WHO), skin cancer accounts for 1/3 (one-third) of all malignancies identified each year, with more than 0.13 million melanomas and 3.0 million non-melanomas identified worldwide [6, 7]. According to the United States Cancer Society (ACS), about a million Americans live with melanoma. There were 196,060 melanomas registered in 2020, with 100,350 invasive cases and 95,710 noninvasive instances [8]. In 2019, there were 192,310 new cases recorded, including 96,480 invasive and 95,380 noninvasive instances [9].

Skin cancer may be properly treated if detected in the early stages [1]. Manual techniques of visual examination and detection of lesion images like dermoscopy are time-consuming, subjective, complicated, and vulnerable [10]. Dermatologists use dermoscopy (a visual examination method) to diagnose malignant lesions, which are difficult to distinguish because of different textured skin or injury [11]. To attain more accuracy in minimum time and help inexperienced dermatologists, different computerized techniques of deep learning are introduced by computer vision researchers to detect skin cancer [12]. Computerized techniques are usually implemented in hospitals to increase identification precision. Computer-based approaches, such as computer vision (CV), have recently improved and are now used by clinicians for early diagnosis [13]. One single evaluation with multiple components can complicate the process recognition, and all these approaches can be used to diagnose skin cancer effectively [14].

Convolutional neural network (CNN) is being used by some researchers to recognize and characterize skin lesions [15, 16]. The Newton Raphson (IcNR) protocol to select the best functions for lesion location and isolation is controlled and more expeditious in detecting skin cancer. A regionally directed neural convolutional network is used in many additional techniques (RCNN). The average accuracy of ISBI2016 was 94.5%, and the ISBI2017 data set was 93.4% [17]. Hua Ng et al. [18] used a computerized color normalization method to improve the low contrast of dermoscopic images. Any preprocessing action must be taken to delete items; hence, for this reason, blue plates, skin lines, hairs, dermoscopic gel, and vascular vessels use segmentation and classification processes. Muhammad et al. [19] presented segmentation and sorting procedures for skin lesions. A MASK R-CNN (Region-Based Convolutional Neural Network) based technique is deployed for the segmentation of lesions with the final mask generation map for ResNet50, and a feature pyramid network (FPN) is to be used. MASK R-CNN is Avant-garde in terms of image segmentation. This deep neural network variation identifies objects in a picture and creates a rising segmentation mask for each. Three datasets (i.e., PH2, ISBI2016, and ISIC2017) are used for verifying the segmentation process. The grouping consists of the architecture of a 24-layer neural network depending on the higher features shown. Hsin et al. [20] built a reliable, feasible, easy-to-use platform to encourage automatic diagnostic algorithms. It examined patient medical records and clinical photos before creating models for remote diagnostic apps on cloud platforms and mobile devices. The lightweight model was introduced to classify skin lesions followed by these steps. The accuracy is 85.8% in the seven-class HAM10000 data set and 72.1% in the five-classes KCGMH data set respectively.

The presence of artifacts such as hairs, bubbles, and noise can be the reason for inapt feature extraction that later decreases classification accuracy. Classifying several skin malignant tumors into an exact class is a complicated and challenging task due to the maximum similarity among various lesions [21]. The chance of a higher sample class increases when the dataset object classes are unequal. The morphology, color, and texture of multiclass skin lesion types are identical, and they share common features [22]. Those features are later classified into an incorrect skin class. Several essential features are discarded during the feature extraction stage, which might lead to a misclassification problem. Multi-properties features are fused into one matrix for improved accuracy in the fusion stage, however, there's a significant chance that some numerous redundant features are also included. Later, this type of difficulty elongates the calculation time. The following are contributions:

- It is evident from the above discussion that a technique is needed that can effectively classify skin cancer types.

- Preprocessing step improved by using phases such as horizontal flipping, translation, and rotation are applied to enlarge the datasets. Moreover, the dull razer method is utilized to clean the skin lesion region.
- Two pre-trained models DarkNet19 and MobileNetV2, are used to retrieve significant features from input images.
- Features optimization is performed using the FireFly algorithm. Which is a heuristic search optimization technique that removes redundant features and selects the most prominent features.
- Three advanced classifiers are used for classifiers.
- In a nutshell, a comprehensive technique to classify skin cancer types with feature optimization is proposed.

In the next section we have explained the proposed methodology in detail followed by the results and analysis in Sect. 3. Finally, conclusions have been drawn in Sect. 4.

2 Proposed methodology

A novel end-to-end automated method for the classification of multiclass skin lesions is proposed in this work. This paper uses the ISIC (International Skin Imaging Collaboration) 2018 dataset to assess the suggested approach. To train and create balanced datasets, we first carried out data augmentation, which can enhance the functionality and output of machine learning (ML) models. A large and sufficient dataset improves the performance and accuracy of a machine learning model. Next, using transfer learning (TL), two pre-trained models (DarkNet-19 and MobileNetv2) are adjusted and trained. After features are retrieved, optimization is done, and then a modified threshold serial technique is utilized to fuse the data. Finally, we use classifiers, such as SVM, K closest neighbor, decision trees, artificial neural networks, and CNN's. The flow diagram of the proposed model is presented in Fig. 1.

2.1 Dataset augmentation

Data augmentation is a typical data enhancement technique in machine learning (ML) and especially in deep learning (DL) that increases the training data [23]. Deep learning methods perform more effectively when the quantity of training data is increased [24]. In this work, we have used the ISIC 2018 dataset, which includes 7 skin classes named AK, BCC, BKL, DF, MEL, NV, and VASC having 327, 514, 1099, 115, 1113, 6705, and 142 images respectively in the training dataset. This variation of images makes the CNN (Convolutional Neural Network) tanning model more complex. Therefore, we perform image augmentation concerning the skin class having the maximum number of images. In this dataset, 6705 is the maximum number of images of only one skin class; the remaining classes have fewer than 1100 images. As a result,

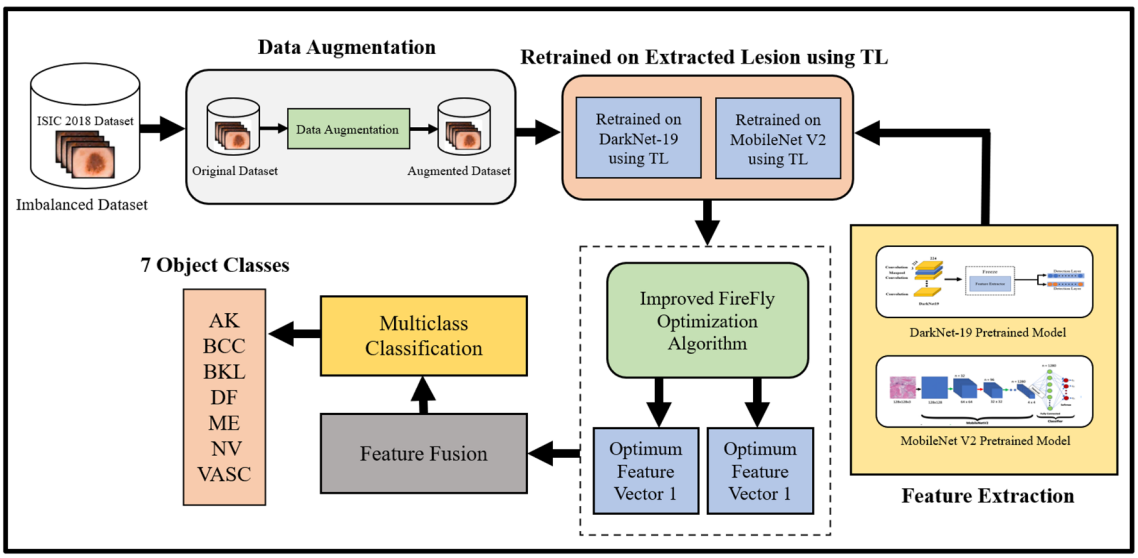

Fig. 1 Deep learning architecture proposed for the classification of skin lesions

we use flipping or rotating procedures to equalize the number of photographs for each skin class by taking an average number of images from each class. Mathematically the performed operations are defined as below:

Let us have an input image matrix of dimension 256×256 denoted $\tilde{A}x, y$ of *Xth* row and *Yth* column. Where $\tilde{A}x, y \in K^{x,y}$. The rows $x = \{1,2, \ldots \tilde{a}\}$ and column $y = \{1,2, \ldots .\dot{b}\}$, where are 3 columns of channels. The RGB nature of the source image is used for three separate flip operations for augmentation. The indices of the original image are updated after this process.

$$\tilde{A}^V = \tilde{A}_x(\tilde{a} + 1 - x)y \tag{1}$$

where $\widetilde{A}^v$ denotes the vertical flip image.

$$\tilde{A}^H = \tilde{A}_x(\dot{b} + 1 - y) \tag{2}$$

where $\widetilde{A}^H$ denotes the horizontal flip image.

$$\tilde{A}^T = \tilde{A}_{y,x} \tag{3}$$

where $\widetilde{A}^T$ indicates that the original image has been transposed. These operations are performed till the quantity of each object classes is equal. Figure 2 shows the flipped image where visualization of the image is changed after the flipped operation. However, it should be noted that only the positions of A and B pixels are altered.

Rotation at 90 degrees, up down (UD), and left right (LR) are shown in Fig. 2, moreover Eqs. 1,2 and 3 represent these operations mathematically.

2.2 Deep learning features

2.2.1 DarkNet19

In this proposed methodology, we use the DarkNet-19 model, which is a pre-trained CNN model. It is a new generation network that has only recently been built that completes the process with the transfer learning approach. This architecture features a 19-layer network topology, and the network accepts 256×256 pixel images as input. The YOLOv2 dataset serves as the foundation for this convolutional neural network model. In conjunction with the VGG model, it makes use of a 3×3 filter. Following each phase of pooling, there are two more channels available. The network employs 1×1 filters to reduce the feature representation and global average pooling to provide predictions. DarkNET has excellent real-time object detection performance. Compared to popular CNNs, the network approaches object detection as a straightforward regression problem. The flowchart for deep CNN feature extraction is displayed in Fig. 3.

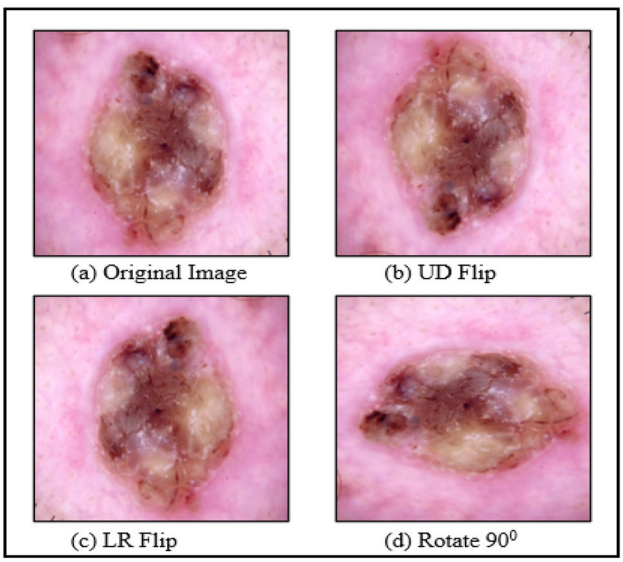

Fig. 2 A visual illustration of flipped operations performed for augmentation

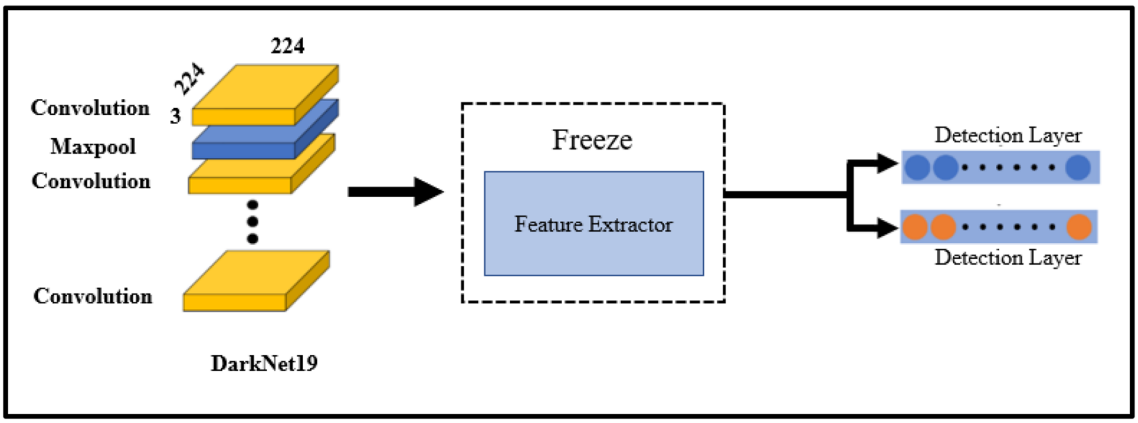

Fig. 3 The workflow for feature extraction from deep CNN

2.2.2 MobileNetV2

It is a deep neural network model that has been optimized for portability and resource constraints. It is based on an upside down residual structure with links to the bottleneck layer [25]. It consists of 153 layers and the input layer has the dimension of $224 \times 224 \times 3$. Two types of blocks are used with the parameters of stride1 and stride2. Every block has three types of layers which are used to decrease the sample size. The first layer is the convolutional layer with a linear activation function named ReLU6. The second convolutional layer is utilized for dimensionality reduction. The last layer is the convolutional layer but without a nonlinearity function. Each layer has two functions namely activation and batch normalization. The last layer has only one batch normalization. The performance will get slow because of the utilization of ReLU6 [26]. A convolutional block of MobileNetV2 is illustrated in Fig. 4. Nineteen residual bottleneck layers come after a convolution layer with thirty-two filters in the basic MobileNetV2 design. Table 1 presents a comprehensive architecture.

For skin lesion classification, we employed pre-trained CNN model named as MobileNet-V2. The original architecture is fine-tuned for this, and the last layer is deleted. Afterward, a new layer with 7 skin classes is introduced.

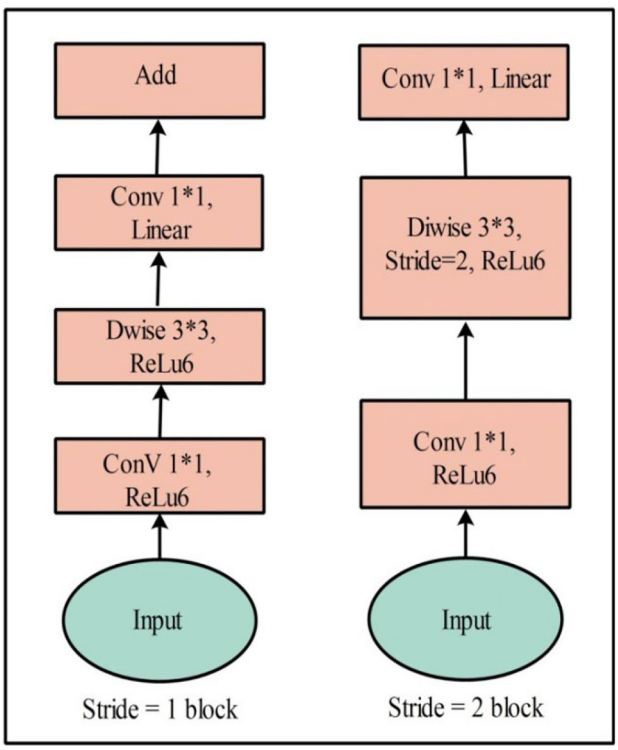

Fig. 4 MobileNetV2 convolutional blocks [26]

Table 1 MobileNetV2 convolutional blocks

Input	Operator	T	C	N	S
$224^2 \times 3$	2d Conv	–	32	1	2
$112^2 \times 32$	Bottleneck architecture	1	16	1	1
$112^2 \times 16$	Bottleneck architecture	6	24	2	2
$56^2 \times 24$	Bottleneck architecture	6	32	3	2
$28^2 \times 32$	Bottleneck architecture	6	64	4	2
$14^2 \times 64$	Bottleneck architecture	6	96	3	1
$14^2 \times 96$	Bottleneck architecture	6	160	3	2
$7^2 \times 60$	Bottleneck architecture	6	320	1	1
$7^2 \times 320$	conv2d 1 × 1	–	1280	1	1
$7^2 \times 1280$	avgpool7 × 7	–	–	1	–
1 × 1 × 1280	conv2d 1 × 1	–	K	–	

Our main objective labels will be based on these classes. Thereafter, transfer learning is used to transfer the original model's information to a specific model, resulting in a new tailored CNN model. The 20 and 55 layers are frozen of darknet19 and mobilenetv2 to reduce the training time. The TL procedure is described in Fig. 5. In this figure, it is shown that the original models are trained on the ImageNet dataset. After the fine-tuning process, the skin-updated dataset has been employed for the training process and later utilized for feature extraction. The average pool layer is used for feature extraction for both models. The dimensions of extracted features are N × 1000 and N × 1280 respectively. The extracted is utilized for feature selection using a feature optimization algorithm. The trained parameters of both models are 3.5M for mobileNetv2 and 20.8M parameters for darknet19.

The capacity to identify significant visual patterns and traits connected to various forms, phases, and clinical consequences makes features collected from skin cancer images employing deep learning based approaches clinically relevant. Clinical decision-making can benefit from these traits in a number of ways, including diagnosis, prognosis, treatment planning, and individualized patient care.

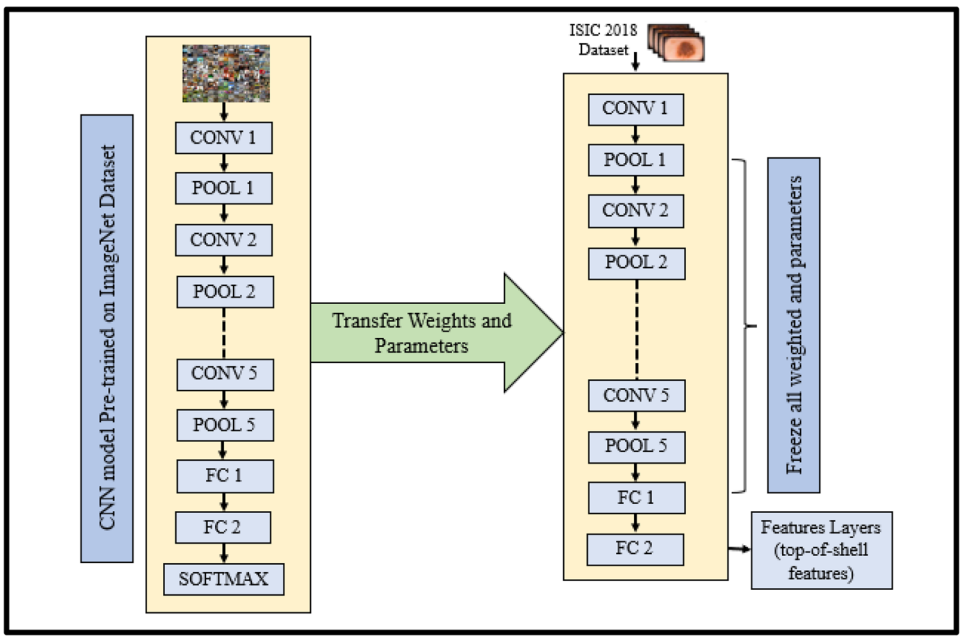

Fig. 5 Visual architecture of feature extraction via TL

2.3 Features optimization

The feature selection algorithm is the hybrid working of slime mould (SMA) and firefly algorithm. The SMA is a population-based optimization technique inspired by the natural oscillation style of slime mould. This is a distinctive scientific process of drawing a path for food and creating propagation feedback (positive and negative) based on oscillation. Consider, the set of solutions $\hat{S}$ and fitness value for each solution is computed in SMA and the best solution $\hat{S}_b$ is determined. The solutions are then updated using the SMA operators $\hat{S}_b$. The slime mould algorithm (SMA) consists of 3 stages which are discussed in the following subsections.

2.3.1 Food approach stage

The following formula is defining the first stage of SMA.

$$\hat{S}t+1 = \begin{cases} \hat{S}b(t) + Vb \times (W.\hat{S}a(t) - \hat{S}b(t)) \; r < p \\ \hat{S}c.\hat{S}t \; r \geq p \end{cases} \tag{4}$$

where $\hat{S}_c$ denotes a parameter that reduces from 1 to 0, the current loop is defined by t, $\hat{S}_a$ and $\hat{S}_b$ are two solutions chosen randomly. ρ is defined as:

$$\rho = tanh|S(i)\, DF| \tag{5}$$

DF represents the best fitness value as $\hat{S}(i)$ denotes the fitness value of $X_i = \{i = 1, 2 \ldots n\}$. in equation no. 1 V_b denotes random number which belongs to [-*a*, *a*], where a is:

$$a = arctanh(-(\frac{t}{maxt}) + 1) \tag{6}$$

In addition, *W* denotes the factor weight of the slime mould, and it is defined as below:

$$W(SmellIndex = sort(i)) \tag{7}$$

$$\hat{S}t+1 = \begin{cases} 1 + r\log((bf - \hat{S}(i))/(bf - wf) + 1) \; cond \\ 1 + r\log((bf - \hat{S}(i))/(bf - wf) + 1) \; otherwise \end{cases} \tag{8}$$

$$Smell\ Index = sort(\hat{S}) \tag{9}$$

where *cond* is that $\hat{S}(i)$ ranks the first half of the $\hat{S}$, and $r \in [0,1]$ denotes the random number. w_f and b_f represent the worst and best local fitness values, respectively. The value of the sorted fitness values is *SmellIndex*.

2.3.2 Food wrap process

This stage emulates the process of updating the solutions, and this is formulated as:

$$\hat{S}t+1 = \begin{cases} rand(UB - LB) + LB \; rand < z \\ Xb(t) + vb(WXA(t) - XB(t))r1 < p \\ vcZ(t)r1 \geq p \end{cases} \tag{10}$$

In this equation, UB and LB are the limitations of the search space, and r1 ∈ [0, 1] is the random number.

2.3.3 Oscillation process

Oscillation is performed in this stage which is defined below where V_b and V_c are oscillating in the range $[-a, a]$ and $[-1, 1]$ respectively.

Firefly is a population-based optimization algorithm proposed by Xin-She Yang. FA is a meta-heuristic technique that mimics Firefly behavior. Firefly is attracted to another Firefly since they are unisex. In simple words, Firefly is generally inspired by the flashing behavior of fireflies. To replicate these behaviors, FA used a set of rules (i.e., attractiveness and brightness). The brighter Firefly will attract the less brilliant Firefly in general. The relation (*ß*) between the solution *i* and solution *j* is defined as follows:

$$\beta = \beta_0 \times e(-rm^2) \tag{11}$$

where *r* = coefficient of light absorption, $ß_0 = 1$ is the attractiveness at the initial distance. The $m_{a,b}$ is defined as follows:

$$m_{a,b} = ||Xi - Xj|| = \sqrt{\sum_{k=1}^{d} (Xac - Xbc)^2} \tag{12}$$

The following is the formula for the movement of a FireFly *m* that was drawn to the brighter FireFly *n*:

$$Xm = Xm + \beta \times (Xm - Xn) + r4\times \in m \tag{13}$$

where $r_4 \in [1,0]$ = random value and $Em \in \hat{S}(\mu, \delta)$ = random vector.

By effectively navigating the high-dimensional feature space, the Firefly Algorithm for feature optimization in skin cancer datasets chooses pertinent characteristics that are essential for proper classification. Because of its non-deterministic search approach, a wide range of feature subsets are thoroughly explored, improving classification accuracy. The method effectively finds informative feature subsets while preserving scalability for big datasets by iteratively optimizing an objective function representing classification performance. This increases the interpretability and precision of skin cancer classification models. The size of the best returned features is $N \times 681$ and $N \times 834$. The selected features of this algorithm are used in the final fusion step because there are two different feature vectors which are obtained- (i) DarkNet-19 model features, and (ii) MobileNet-v2 features.

2.4 Features fusion

Since features from different layers and CNN architectures have varied visual qualities, several feature utilization will allow exploiting the strengths for better accuracy of classification. The main goal of this study is to see how the classifier reacts when several ConvNets features are combined. We use a low-level fusion technique to generate a joint feature space by concatenating the collected features directly. Assume that we obtain the resulting feature vector $FV = \{FV_a^c\} or \{FV_b^d\}$, after selecting different layers and applying multiple pre-trained CNN models

$$a = b \in \text{ Selected layers of CNN model)}$$

$$c = d \in \{1,2, \ldots \ldots n\}(\text{Different models})$$

here we have two models, so $c \in (1,2)$. The fused vector using the serial approach is defined as follows:

$\xi = FV_a^c || FV_b^d$ having D dimensions. The resultant vector is further passed in a threshold function that is dependent on the standard deviation.

$$\widetilde{\xi} = \begin{cases} \widetilde{\xi} for \xi(k) \geq sd \\ Ignore,\ Elsewhere \end{cases} \tag{14}$$

The selected features are finally classified using machine learning classifiers. In contrast to the serial-based approach, the threshold-based fusion approach combines decision outputs by aggregating decisions that exceed a threshold. This approach has several benefits, including flexible decision aggregation, resilience to individual errors, improved sensitivity and specificity through threshold optimization, decreased computational complexity,

and flexibility in response to decision confidence levels. This approach is useful for a variety of applications, including pattern recognition and classification, because it guarantees a more nuanced and flexible fusion of information, reduces the impact of individual errors, optimizes performance in accordance with application requirements, expedites decision-making, and takes decision confidence levels into account for better-informed final decisions.

To sum up each stage of the proposed methodology, first vary the training data, data augmentation techniques like as rotation by 90 degrees and flipping both horizontally and vertically are employed on the skin cancer dataset. Two pre-trained models, DarkNet-19 and MobileNetv2, are used for feature extraction in order to obtain rich representations from the augmented photos. Firefly optimization is then used to refine the retrieved features, improving their ability to discriminate. Both serial-based and concatenation fusion procedures are employed to fuse the optimal feature vectors from the two models. The goal of classifying lesions into discrete groups is achieved by applying this fused feature representation to the categorization of skin cancer kinds. To raise the precision and resilience of skin cancer classification, this all-encompassing method combines data augmentation, and feature extraction with pre-trained models, optimization, and fusion techniques.

3 Results and analysis

The proposed method results are presented in this section. Results are computed using a 70:30 approach with tenfold cross-validation. 70% data indicates the training portion and 30% of data are used for the testing process. The hyperparameters for training both models are statically initialized. The hyperparameters are learning rate, epochs, mini-batch size, validation frequency, and optimizer having values are 0.001, 30, 16, 05, and SGDM. Many classifiers such as SVM, KNN, and linear discriminant analysis have been employed for classification results. Results are computed using two different steps. In the first step, results are computed using original serial-based fusion results. In the second step, threshold-based serial fusion results are evaluated. All the experimental process is executed on MATLAB R2023a using a Desktop system based on core i5 configured with 16 of RAM, 256 SSD, and 4 GB NVIDIA GTX 1650 D6 graphic card.

Table 2 presents the classification results of skin cancer using a proposed method on the ISIC 2018 dataset after the original serial fusion approach. The Cubic SVM attained the highest accuracy of 87.5%. Several other performance measures are also computed: Sensitivity rate is 86.27, Precision rate is 86.46, F1-Score is 86.36, FPR is 0.02, and AUC is 0.98. The second best-noted accuracy of this experiment is also 87.5% for Ensemble (Sub Space KNN), but it takes more time than the SVM (Cubic SVM). The rest of the classifiers, such as SVM (Quadratic SVM), SVM (Medium Gaussian SVM), KNN (Fine KNN), KNN (Weighted KNN), SVM (Linear SVM), Ensemble (Bagged Tree), and Linear Discriminant achieved accuracy of 85.9, 85.9, 85.1, 84.3, 80.9, 78.6, 78.3, and 77.1 respectively. The Cubic SVM gives better classification results for multiclass

Table 2 Multiclass Skin Lesion classification results using a proposed method on original, optimized features for the ISIC2018 dataset

Classifier	Sensitivity (%)	Precision (%)	F1-Score	FPR	AUC	Accuracy (%)	Time (S)
SVM (Cubic SVM)	86.27	86.46	86.36	0.02	0.98	**87.5**	**368.83**
Ensemble (Sub Space KNN)	87.26	86.74	87	0.02	0.98	87.5	782.4
SVM (Quadratic SVM)	84.26	84.77	84.51	0.03	0.98	85.9	532.46
SVM (Medium Gaussian SVM)	83.87	85.13	84.5	0.02	0.98	85.9	526.59
KNN (Fine KNN)	85.4	84	84.69	0.03	0.91	85.1	702
KNN (Weighted KNN)	83.37	83.3	83.33	0.03	0.98	84.3	473.76
SVM (Linear SVM)	78	79.01	78.5	0.03	0.96	80.9	571.84
Ensemble (Bagged tree)	75.04	78.89	76.92	0.04	0.95	78.6	317.63
KNN (Cosine KNN)	76.24	75.96	76.1	0.04	0.95	78.3	799.22
Linear discriminant	74.57	74.39	74.48	0.04	0.94	77.1	424.83

Bold values describes the most efficient results

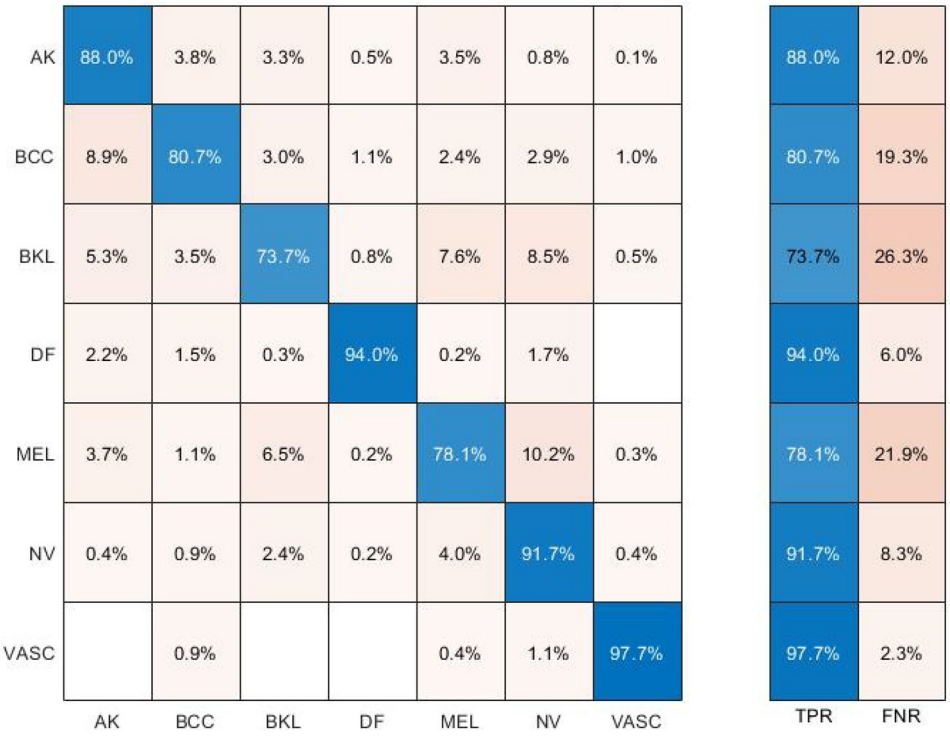

Fig. 6 Confusion matrix of CSVM for proposed method (original serial fusion)

Table 3 Multi Class Skin Lesion fusion results after optimization results of both models

Classifier	Sensitivity (%)	Precision (%)	F1-Score	FPR	AUC	Accuracy (%)	Time (S)
SVM (Cubic SVM)	87.34	87.57	87.45	0.02	0.84	**89.0**	263.71
SVM (Quadratic SVM)	86.81	87.03	86.92	0.02	0.99	88.5	435.79
Linear discriminant	85.16	84.64	84.9	0.02	0.97	86.7	**71.625**
SVM (Medium Gaussian SVM)	83.84	84.97	84.4	0.02	0.98	86.3	741.8
Ensemble (Sub Script Discriminant)	84.29	84.06	84.2	0.02	0.97	86.2	489.87
SVM (Linear SVM)	83.4	83.71	83.55	0.02	0.98	85.7	435.65
Ensemble (Sub space KNN)	83.84	80.76	82.27	0.03	0.96	82.9	693.3
KNN (Fine KNN)	79.23	78.27	78.75	0.03	0.88	80.4	227.7
KNN (Weighted KNN)	76.89	79.5	78.17	0.03	0.96	80.0	327.68
SVM (Coarse Gaussian SVM)	75.21	77.9	76.53	0.04	0.97	79.6	188.1

Bold values describes the most efficient results

problems based on these values. Figure 6 illustrates the confusion matrix of Cubic SVM. This confusion matrix describes the correct predicted rate of each class.

Table 3 shows the classification results of threshold based serial fusion approach. The Cubic SVM attained the highest accuracy of 89.0%. Several other performance measures are also computed: Sensitivity rate is 87.34, Precision rate is 87.57, F1-Score is 87.45, FPR is 0.02, and AUC is 0.84. The second best noted accuracy of this experiment is also 88.5% for SVM (Quadratic SVM). The rest of the classifiers such as, Linear Discriminant, SVM (Medium Gaussian SVM), Ensemble (Sub Script Discriminant), SVM (Linear SVM), Ensemble (Sub Space KNN), KNN (Fine KNN), KNN

Fig. 7 Confusion matrix of CSVM for proposed method (threshold serial fusion)

(Weighted KNN), and SVM (Coarse Gaussian SVM) achieved an accuracy of 86.7, 86.3, 86.2, 85.7, 82.9, 80.4, 80 and 79.6 respectively. Based on these values, the Cubic SVM gives better classification results for multiclass problems. Figure 7 illustrates the confusion matrix of Cubic SVM. This confusion matrix describes the correct predicted accuracy of each class.

Compared to the performance of both steps accuracy, it is noted that the accuracy of the threshold-based serial method is better. In addition, the computational time of threshold serial fusion is minimized due to the reduction of a few features. The comparison is also conducted with [27], which achieved an accuracy of 85.1%. The proposed method attained a better accuracy of 89.0%. Overall, the proposed method obtained improved accuracy.

In Table 2 which presents the results on original data, the lowest computational time is taken by the Cubic SVM classifier as well as highest accuracy of 89% is achieved. After applying the optimization technique, Table 3 demonstrates that Linear Discriminant (LD) classifier completes its work in lowest times as compared to other classifiers. on the other hand, side highest accuracy is achieved by Cubic SVM classifier. In a bigger picture, usage of an optimization technique selects the best contributing features and in turn reducing the overall execution time for almost all classifiers.

4 Conclusion

A fully automated deep learning framework is designed in this work for multiclass skin lesion classification. The ISIC2018 dataset has been utilized in this work and performed augmentation at the initial stage for the imbalance problem. After that, two pre-trained CNN models are fine-tuned and trained on the augmented dataset and obtained trained models. Features were extracted and optimized using a hybrid firefly optimization approach. The major contribution is: The optimal features are first fused using a serial-based technique and obtained an accuracy of 87.5%. Later, a threshold serial fusion approach proposed and an accuracy of 89% is obtained. Which is improved than the recent state-of-the-art techniques. The limitation of the proposed framework was the high training time of proposed deep learning models. In future, a new deep neural network will be designed based on the residual blocks and attention mechanism. Moreover, a feature selection technique will be introduced based on heuristics mechanism inspired from nature. The experimental process will also expand on the new datasets such as ISIC2019 and ISBI2020.

Author contributions Muhammad Attique Khan: Conceptualization, Methodology, Resources, Software, Writing—original draft. Ameer Hamza: Conceptualization, Methodology, Validation, Visualization, Writing—original draft. Mohammad Shabaz: Conceptualization, Data curation, Project administration, Supervision, Validation, Visualization, Writing—original draft. Seifeine Kadry: Conceptualization, Data curation, Formal analysis, Investigation, Visualization, Writing—original draft. Saddaf Rubab: Formal analysis, Investigation, Validation, Writing—original draft, Writing—review & editing. Muhammad Abdullah Bilal: Formal analysis, Validation, Visualization, Writing—original draft, Writing—review & editing. Muhammad Naeem Akbar: Investigation, Validation, Writing—review & editing. Suresh manic Kesavan: Formal analysis, Validation, Visualization, Writing—review & editing.

Funding This research work is self-funded.

Data availability Data used in this research is publicly available at: https://www.kaggle.com/datasets/tschandl/isic2018-challenge-task1-data-segmentation

Declarations

Competing interests The authors declare that they have no competing interests.

Human and animal rights No human or animals were involved in this research.

References

1. Bibi S, Khan MA, Shah JH, Damaševičius R, Alasiry A, Marzougui M, Alhaisoni M, Masood A. 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. https://doi.org/10.3390/diagnostics13193063.
2. Dillshad V, Khan MA, Nazir M, Saidani O, Alturki N, Kadry S. D2LFS2Net: Multi-class skin lesion diagnosis using deep learning and variance-controlled Marine predator optimisation: an application for precision medicine. CAAI Trans Intell Technol Institut Eng Technol (IET). 2023. https://doi.org/10.1049/cit2.12267.
3. Hussain M, Khan MA, Damaševičius R, Alasiry A, Marzougui M, Alhaisoni M, Masood A. SkinNet-INIO: multiclass skin lesion localization and classification using fusion-assisted deep neural networks and improved nature-inspired optimization algorithm. Diagnostics. 2023;13(18):2869. https://doi.org/10.3390/diagnostics13182869.
4. Cummins DL, Cummins JM, Pantle H, Silverman MA, Leonard AL, Chanmugam A. Cutaneous malignant melanoma. Mayo Clinic Proceed. 2006;81(4):500–7. https://doi.org/10.4065/81.4.500.
5. Yousef H, Sharma S. Anatomy, skin (Integument), epidermis. St. Petersburg, FA: StatPearls Publishing LLC; 2018.
6. Gordon R. Skin cancer: an overview of epidemiology and risk factors. Semin Oncol Nurs. 2013;29(3):160–9. https://doi.org/10.1016/j.soncn.2013.06.002.
7. Sethanan K, Pitakaso R, Srichok T, Khonjun S, Thannipat P, Wanram S, Boonmee C, Gonwirat S, Enkvetchakul P, Kaewta C, Nanthasamroeng N. Double AMIS-ensemble deep learning for skin cancer classification. Expert Syst Appl. 2023;234:121047. https://doi.org/10.1016/j.eswa.2023.121047.
8. Yuan J, Li X, Yu S. Global, regional, and national incidence trend analysis of malignant skin melanoma between 1990 and 2019, and projections until 2034. Canc Control. 2024. https://doi.org/10.1177/10732748241227340 5.
9. Hameed N, Ruskin A, Abu Hassan K, Hossain MA (2016) A comprehensive survey on image-based computer aided diagnosis systems for skin cancer. In: 2016 10th International conference on software, knowledge, information management & applications (SKIMA), IEEE. https://doi.org/10.1109/skima.2016.7916221
10. Erkol B, Moss RH, Joe Stanley R, Stoecker WV, Hvatum E. Automatic lesion boundary detection in dermoscopy images using gradient vector flow snakes. Skin Res Technol. 2005;11(1):17–26. https://doi.org/10.1111/j1600-0846200500092.x.
11. Argenziano G, Soyer HP, Chimenti S, Talamini R, Corona R, Sera F, Binder M, Cerroni L, De Rosa G, Ferrara G, Hofmann-Wellenhof R, Landthaler M, Menzies SW, Pehamberger H, Piccolo D, Rabinovitz HS, Schiffner R, Staibano S, Stolz W, et al. Dermoscopy of pigmented skin lesions: results of a consensus meeting via the Internet. J Am Acad Dermatol. 2003;48(5):679–93. https://doi.org/10.1067/mjd.2003.281.
12. Dildar M, Akram S, Irfan M, Khan HU, Ramzan M, Mahmood AR, Alsaiari SA, Saeed AHM, Alraddadi MO, Mahnashi MH. Skin cancer detection: a review using deep learning techniques. Int J Environ Res Public Health. 2021;18(10):5479. https://doi.org/10.3390/ijerph18105479.
13. Afza F, Sharif M, Khan MA, Tariq U, Yong H-S, Cha J. Multiclass skin lesion classification using hybrid deep features selection and extreme learning machine. Sensors. 2022;22(3):799. https://doi.org/10.3390/s22030799.
14. Tahir M, Naeem A, Malik H, Tanveer J, Naqvi RA, Lee S-W. DSCC_Net: multi-classification deep learning models for diagnosing of skin cancer using dermoscopic images. Cancers. 2023;15(7):2179. https://doi.org/10.3390/cancers15072179.

15. Olayah F, Senan EM, Ahmed IA, Awaji B. AI techniques of dermoscopy image analysis for the early detection of skin lesions based on combined CNN features. Diagnostics. 2023;13(7):1314. https://doi.org/10.3390/diagnostics13071314.
16. Patel M. Multi class skin diseases classification based on dermoscopic skin images using deep learning. Int J Next-Generat Comput. 2022. https://doi.org/10.47164/ijngc.v13i2.480.
17 Oliveira RB, Marranghello N, Pereira AS, Tavares JMRS. A computational approach for detecting pigmented skin lesions in macroscopic images. Expert Syst Appl. 2016;61:53–63. https://doi.org/10.1016/j.eswa.2016.05.017.
18. Hewitt B, Yap MH, Ng J, Goyal M (2019) The effect of color constancy algorithms on semantic segmentation of skin lesions. In: Gimi B, Krol A (Eds), Medical imaging 2019: Biomedical applications in molecular, structural, and functional imaging. SPIE. https://doi.org/10.1117/12.2512702
19 Khan MA, Zhang Y-D, Sharif M, Akram T. Pixels to classes: intelligent learning framework for multiclass skin lesion localization and classification. Comput Electr Eng. 2021;90:106956. https://doi.org/10.1016/j.compeleceng.2020.106956.
20 Huang H, Hsu BW, Lee C, Tseng VS. Development of a light-weight deep learning model for cloud applications and remote diagnosis of skin cancers. J Dermatol. 2020;48(3):310–6. https://doi.org/10.1111/1346-8138.15683.
21. Ahmad N, Shah JH, Khan MA, Baili J, Ansari GJ, Tariq U, Kim YJ, Cha J-H. A novel framework of multiclass skin lesion recognition from dermoscopic images using deep learning and explainable AI. Front Oncol. 2023. https://doi.org/10.3389/fonc.2023.1151257.
22 Malik S, Akram T, Awais M, Khan MA, Hadjouni M, Elmannai H, Alasiry A, Marzougui M, Tariq U. An improved skin lesion boundary estimation for enhanced-intensity images using hybrid metaheuristics. Diagnostics. 2023;13(7):1285. https://doi.org/10.3390/diagnostics13071285.
23. Khan MA, Akram T, Zhang Y, Alhaisoni M, Al Hejaili A, Shaban KA, Tariq U, Zayyan MH. <scp>SkinNet-ENDO </scp> Multiclass skin lesion recognition using deep neural network and <scp> Entropy-Normal </scp> distribution optimization algorithm with <scp> ELM </scp>. Int J Imag Syst Technol. 2023;33(4):1275–92. https://doi.org/10.1002/ima.22863.
24. Hu Z, Tan B, Salakhutdinov R, Mitchell T, Xing EP (2019) Learning data manipulation for augmentation and weighting (Version 1). arXiv. https://doi.org/10.48550/ARXIV.1910.12795
25 Jiang S, Li H, Jin Z. A visually interpretable deep learning framework for histopathological image-based skin cancer diagnosis. IEEE J Biomed Health Inform. 2021;25(5):1483–94. https://doi.org/10.1109/jbhi.2021.3052044.
26. Sandler M, Howard A, Zhu M, Zhmoginov A, Chen LC (2018) MobileNetV2: Inverted Residuals and Linear Bottlenecks. In: 2018 IEEE/CVF conference on computer vision and pattern recognition (CVPR), IEEE. https://doi.org/10.1109/cvpr.2018.00474
27 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. https://doi.org/10.1111/exsy.12497.