Ensemble learning of deep learning and traditional machine learning approaches for skin lesion segmentation and classification

Article *in* Concurrency and Computation Practice and Experience · March 2022

DOI: 10.1002/cpe.6907

5 authors, including:

Adil Humayun Khan
Prince Mohammad bin Fahd University
24 PUBLICATIONS **260** CITATIONS

Dayang Nurfatimah Awang Iskandar
University of Malaysia, Sarawak
45 PUBLICATIONS **602** CITATIONS

Jawad Al-Asad
Prince Mohammad bin Fahd University
34 PUBLICATIONS **273** CITATIONS

Hiren Mewada
125 PUBLICATIONS **652** CITATIONS

Received: 21 June 2021 | Revised: 12 February 2022 | Accepted: 14 February 2022
DOI: 10.1002/cpe.6907

RESEARCH ARTICLE

WILEY

Ensemble learning of deep learning and traditional machine learning approaches for skin lesion segmentation and classification

Adil H. Khan[1,2] | **Dayang NurFatimah Awang Iskandar**[1] | **Jawad F. Al-Asad**[2] | **Hiren Mewada**[2] | **Muhammad Abid Sherazi**[3]

[1]Faculty of Computer Science and Information Technology, Universiti Malaysia Sarawak, Sarawak, Malaysia

[2]Electrical Engineering Department, Prince Mohammad Bin Fahd University (PMU), Al-Khobar, Saudi Arabia

[3]School of Electrical Engineering and Computer Science (SEECS), National University of Sciences and Technology (NUST), Islamabad, Pakistan

Correspondence
Adil Humayun Khan, Electrical Engineering Department, Prince Mohammad Bin Fahd University (PMU), Al Khobar 31952, Saudi Arabia.
Email: aadilhk@gmail.com

Summary

Melanoma is a type of a skin cancer or lesion which has the detrimental ramifications on the human health but with early diagnosis it can be cured easily. The actual identification of skin lesion is very challenging because of factors like a very minute difference between lesion and skin and it is very difficult to differentiate among skin cancer types due to visual comparability. Hence an autonomous system for the diagnosis of true skin cancer type is very useful. In this article, we took the leverage of ensemble learning by combining the features of deep learning architectures with traditional features extraction approaches. For segmentation, we have two pipelines for the feature extraction. We extract the features through traditional split and merge approach as well as from deep learning algorithms of contextual encoding along with the attention mechanism. Later we combine the features of both architectures and predict the segmented region through intersection over union mechanism. After that segmented region is classified into three types of skin lesion using hybrid features of Alex-Net and VGG-16 through the transfer learning approach. The evaluation has been performed using the ISIC and PH2 datasets for which achieved segmentation accuracy is 97.8% and 96.7%, respectively. Moreover, hybrid classification network able to attain the 98.2% accuracy.

KEYWORDS

classification, convolution neural network, deep learning, ensemble learning, feature fusion, segmentation

1 | INTRODUCTION

Melanoma is considered as the deadliest and malignant type of skin lesion. The fatality rate of such skin lesion is reasonably high around the world. According to a survey in 2019 by American Cancer Society nearly 91,000 people are diagnosed with melanoma skin cancer and 7000 people out of this die in an year.[1] There are many factors for increasing cases of skin cancer but ozone layer degradation and tanning are considered as main causes.[2] If detected early then this disease can be cured easily by a simple surgery of excision by dermatologist.[3] The two generalized types of skin cancer are benign skin cancer and malignant skin cancer. There are several benign skin cancer types but our focus in this research will be on two most common types which are seborrheic keratosis (SK) and nevus. Melanoma is the most dangerous skin cancer belongs to malignant skin cancer type. SK is a benign skin cancer emerged from the first layer of the skin in non-melanocytic cells[4] whereas nevus derived from melanocytic cells of first layer of the skin.[5] The third type melanoma is also derived from melanocytic cells but carries the danger of penetrating within the skin layers and other body parts through blood circulation.[6] Dermoscopic images are captured through dermatoscope through the process of optical magnification or liquid immersion. Such dermoscopic images assist in identification of different morphological features of skin lesions so that dermatologist can diagnose actual skin lesion type and start treatment accordingly.[7] This whole process of actual skin lesion detection vastly dependent on dermatologist and maybe be affected by inexperience or human error.[8] Moreover, the physical appearance of skin lesion types is very similar, for example, many incorrect diagnoses are reported in past due to similarity between melanoma and SK due similar color and shape.[9] Similarly, there is resemblance between melanoma and a typical nevus and people with higher numbers of moles are at increased risk of developing melanoma. Though other factors are involved as well in developing melanoma but still higher number of nevi or moles will be a risk factor of turning in to melanoma.[9,10] Hence, an early and accurate diagnosis of skin cancer type is exceedingly important for treatment of this disease. A computerized image analysis, also known as computer aided design (CAD) systems, can improve visual interpretation and reduce the diagnosis error.[11] Some commonly used methods are ABCD based on asymmetry, border, color, diameter, and so forth,[12] 7 point checklist and CASH algorithm.[13] An automated system to detect skin lesion type consist of three main stages: segmentation, feature extraction, and classification. In segmentation, the cancerous part is separated from the background of dermoscopic image. In next stage, the features of this segmented foreground are extracted then classification is performed with the help of extracted features to detect true skin lesion type. Such automated systems to detect skin lesion using computer vision based methods are presented in References 14-17.

1.1 | Motivation and contribution

Although existing techniques in literature exhibited improvement in skin lesion detection, still there are challenges in this research area which needs to be addressed. In traditional systems, preprocessing is performed at initial stage to remove different artifacts from images which introduces extra computation. Irregular and nonuniform skin lesion shapes also degrade segmentation performance. As mentioned earlier that skin lesion types are very similar in color, size, shape, and so forth which makes the task of distinct feature extraction very challenging and for classification we need conclusive and prominent feature for respective skin lesion types. Also most of the research has been done for the binary classification of skin lesion; melanoma or benign.[18-21] In this article, we have made the following contributions

- Proposed ensemble learning based segmentation architecture, which combines the features of traditional machine learning approaches with the convolution neural network (CNN) architecture.
- A hybrid classification model by fine-tuning the features of AlexNet and VGG-16 architecture.
- Transfer learning techniques have been utilized to train the deep learning based architecture for both classification and segmentation.

The rest of this article is organized as follows; related work is presented in Section 2 and proposed system is discussed in Section 3. The datasets are described in Section 4, experiments and evaluation are presented in Section 5. Finally, the conclusions are given in Section 6.

2 | RELATED WORK

Techniques presented in literature for the segmentation of skin lesion can be divided in to the three different groups. This includes traditional machine learning approaches, deep learning techniques, and the ensembling of the features of different techniques of one approach. We first discuss the traditional machine learning techniques from the literature. Histogram thresholding based methods discriminate the lesion from background on the basis of color space features.[22] Unsupervised clustering based methods have also been used to segment the cancerous part from the medical images.[23,24] Edge based methods use edge operators while region based techniques use region split and merge algorithm for segmentation.[25,26] Most recently supervised learning techniques are used for the segmentation process, in which classifiers such as artificial neural networks, support vector machines (SVMs), decision tree, and so forth are trained.[27-29] These learning techniques also proved successful in feature extraction and classification such as CNNs, deep learning based algorithms.[30,31] Extracted features from dermoscopic images plays key role in classifying true lesion type. Usual features of skin lesion include color, size, shape, histogram, contour, border, and so forth a detailed survey on feature extraction and selection is done in Reference 32. In Reference 27, proposed algorithm used full resolution convolution networks (FrCN) without any preprocessing or artifacts removal to secure resolution based features. Analysis performed on two publically available datasets. This proposed method exhibit superior segmentation performance as compare to other deep learning based methods like fully convolution networks (FCN), U-net, SegNet, and DL based methods. In Reference 33, skin lesion is detected on the basis of color evaluation of cancerous part by incorporating QuadTree based method. It is shown in this research that melanoma lesions contain more colors with higher frequency of pink, black, and blue-grey as compare to benign lesions. Four classifier were evaluated from which SVM provide best results of 93%. In Reference 34, transfer learning based method incorporated with deep neural network for the classification. Transfer learning is implemented on AlexNet while using softmax as classification layer.

Skin lesion was classified between melanoma, nevus, and SK while accuracy of 95% achieved on publically available dataset. A hybrid of machine learning and statistical is proposed in Reference 19 to extract the features, where VGG and AlexNet models are used to extract optimized color features and later classification is done through deep CNN based model. Author claimed segmentation and classification accuracy of 90% and 92%, respectively. Khan et al.[35] first preprocess the input image using the local color-controlled histogram intensity values and then extract the features using 10 layers CNN architecture. They fused the features using the multiset maximum correlation analysis (MMCA) and classify these features with the help of kernel extreme learning machine (KELM) classifier. From the past few years, researcher have now focused on combining the features of different architectures of same approach like combining features of traditional machine learning approaches or deep learning approaches. In Reference 36, a hybrid of deep learning and Newton–Raphson which is an iteration controlled based method is proposed. Segmentation is performed through region based neural networks with feature extraction through deep learning and selection by Newton–Raphson based method. These features were given input to multilayer forward neural networks and secure accuracy of almost 94% on publically available dataset. In Reference 37, boundary of the skin lesion is segmented through deep learning based method FrCN. This method extracted critical class features which were given input to multiple CNN based classifiers such as DenseNet, Inception, ResNet, and Inception-ResNet-v2. In Reference 38, region growing based method is applied for the segmentation on preprocessed images after hair removal. Feature extraction is done through local and color morphological features. Deep learning based recurrent neural network model is used for the classification. In Reference 21, recent deep learning based models are used for feature extraction, using transfer learning. Segmented part of the image is subjected to model for the generation of vector with fused features. Optimization of fused features is performed through principle component analysis. Multiple classifiers based on K-nearest neighbor are used for binomial classification. In Reference 20, deep semantic segmentation with CNN based classification is performed. Open neural networks and squeeze Net model is used for localization of skin lesion. At second stage, semantic segmentation model based on convolution, batch normalization, ReLU, and softmax layers is used for segmenting the foreground. Features are extracted through ResNet-18 based model which are passed to an optimized SVM based classifier. Context encoding module is also used recently for sematic segmentation in imaging based applications.[39] This module generates class dependent features to highlight distinct parts in the image and better segmentation accuracy is achieved as compare to FCN based framework.

Currently, to the best of our knowledge most of the work in the literature is either used the traditional machine learning approaches for segmentation as well as classification, or they utilized the deep learning algorithms. In this work, we have taken the leverages of combining the features of both approaches of traditional machine learning as well as the deep learning through ensemble learning. Ensemble learning approach is based on the amalgamation of the features of both machine learning and deep learning.

3 | PROPOSED SYSTEM

3.1 | Segmentation

Image segmentation is a complex computer vision problem that splits the complete image into multiple regions based on the incorporated content of the image. Image segmentation of different objects provides the compact contours and masks of objects and background. We can consider an image as the collection of thousands of pixels and the overall target of any segmentation network is to assign the individual pixel into one of the classified object or the background. In this article, we have focused on merging traditional machine learning based segmentation with state-of-the-art deep learning method to utilize the power of both techniques. The power of explainability from traditional method is combined with the excellent abstraction power of deep neural networks for huge amount of data. We employed split and merge method and context encoding network (EncNet) from deep learning paradigm. The origin of above-mentioned techniques is derived from region growing method that is why we considered combining these two would be more appropriate. The basic contribution of the article is as follows: This methodology section highlights the details of feature extraction using region growing, split and merge, and context encoding techniques. Ultimately, we combined the region extracted from split and merge method with the features from context encoding network.

3.1.1 | Split and merge method

Split and merge method based image segmentation is an image processing approach to segment out the objects, which is skin lesion in this scenario, from dermoscopic image.[26,40] This process starts from dividing the complete image into multiple small segments that are called quadrants. Afterwards, these quadrants are concatenated together to create the segmented image. We specifically employed integral method for finding the homogeneity in the image. Let us first explain the integral images that is the basic part of split and merge method.

3.1.2 | Integral images

The integral image is also known as the summed area table; effectively used for computing sums of pixel intensities in a squared region. Let us consider an input image I with dimensions of width and height as $(w \times h)$ having a starting pixel as $(0,0)$ on the top left of the overall image and the $(w-1, h-1)$ on the bottom of the right side. The pixel at (x,y) position is denoted by $I(x,y)$ for x as column and y as the row number. The integral image of shape $(w \times h)$ is denoted by Equation (1). The integral cell at (x,y) is donated by $S(x,y)$.

$$S(x,y) = \sum_{x<x' \leq y}^{U} I(x', y'), \tag{1}$$

where (x', y') are current pixel location during calculation of (x,y).

3.1.3 | Sum of intensities in a region

The integral image extracted from above calculation is employed for computing the pixel intensities summation for a particular squared region. The computation time is constant and independent of the area of the squared image involved. Let us consider multiple points on the image that are $w = (i_0, j_0)$, $x = (i_1, j_0)$, y=(i_0, j_1), and $z = (i_1, j_1)$, where i_i shows the coordinates of the image as shown in Figure 1. The overall summation of pixel values can be described as follows in Equation (2).

$$S = S(i_1, j_1) - S(i_0 - 1, j_1), S(i_1, j_0 - 1), S(i_0 - 1, j_0 - 1). \tag{2}$$

The split and merge method divide the whole image and find out the regions where the concerned object is present. After that, we summed up the intensities of these splitted regions using Equation (2). Figure 1 describes those summed intensities.

3.1.4 | Statistical computation

We calculated arithmetic mean and variance of the squared region computed in previous section. Let us consider the following equation to find the sums of powered intensities from (i_0, j_0) to (i_1, j_1). All the intensity values are first powered by r as shown in Equation (3).

$$S_r = \sum_{j=j_0}^{j_1} \sum_{i=i_0}^{i=i_0} . \tag{3}$$

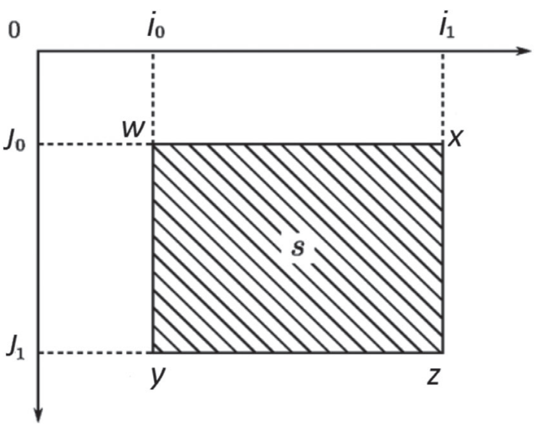

FIGURE 1 Summed of intensities after split and merge method

Moreover, we have utilized the equation of mean and variance of integral image based on squared region. The value of mean is represented as μ and the variances as σ^2 as shown in Equations (4) and (5).

$$\text{Mean} = \mu = \frac{S_1}{N}, \tag{4}$$

$$\text{Variance} = \sigma^2 = \frac{S_2}{N} - \left(\frac{S_2}{N}\right)^2. \tag{5}$$

3.1.5 | Splitting process

After this region is sub divided into further regions. This step breaks the overall image into homogeneous regions of equal width and height. Complete picture is broken into four quad sub-images. If the quad image is not homogeneous in nature then it is further splitted into quads. This method is executed in loop until the quads are homogeneous. Considering this situation, a particular threshold is introduced as a tolerance level. The intensity variation allowed can be determined by considering variation threshold σ^2 for every particular image individually.

Tree building process: Before further splitting the image, the input image is processed in a way to enable subdivision easier. This method tries to make the image in square shape along with the sides that are power of two. This power method enables that the image is divided at every tree level and output images are in power of two. If the original dimensions of the input image does not satisfy the power of two condition then image is padded to meet the requirement $(2^n, 2^n)$, where the parameter n is the positive whole number. A special value for empty pixel is computed by checking the remaining pixel values. These padded pixels are different from original pixels to make these extra pixels excluded from the segmentation process.

Further splitting process: After upsizing of the image using empty pixels. The division process is started from complete image by first checking the homogeneity of the complete image. Considering the dimensions of the image as 2^n, the number of intensities would be n. Following this, the overall homogeneity of the image is determined by checking the intensities of the pixels with respect to the threshold parameter σ^2. To determine whether an image is homogenous or not, its intensity variation is compared with the maximum allowed variance and its variance should be below threshold. If this condition is not met then the image is subdivided into four portions with each side of 2^{n-1}. This step is repeated in loop for all four sub-regions until all the regions meet the homogeneity condition or $n = 0$ level is attained, means no more subdivision is possible.

3.1.6 | Merging process

The integral split and merge method after splitting the complete image into multiple homogenous regions tries to merge the splitted segments of the image into some bigger representations that are further assigned to a single label of the class. The region should be connected. This process takes three considerations while constructing the region: (1) classes count in the image are determined by automatic means of connected region count, (2) homogeneous regions are smooth gradients, (3) disjoint regions are assigned different classes.

Merging conditions: In the tree, all homogenous regions are represented by the individual tree node. Each node can be combined with the other tree node based on specific merging condition. The conditions of the merging process are defined as: (1) unclassified region would be merged by a classified region, (2) the merging region should be small or equal to the already classified region, (3) the overall difference between mean of intensities of both the regions should be less than σ which is automatically determined threshold. At the end, the combining region should be connected vertically or horizontally. There can be a scenario where we have a region that cannot be assigned to any existing classified regions based on threshold, in that case a new class label is created for that particular region. On the other hand, a single region can be matched to multiple regions based on defined conditions, in that case all the neighboring regions can have same class that is assigned to new region. If the neighboring regions have different classes, then these regions are further reassigned to a single class.

3.1.7 | Class selection

Initially, when the process of merging begins, no region has any class assigned to it. In this case, when a region does not have any candidate to attach with then a group structure g is created and all the neighboring nodes that satisfy the condition of merging are connected with that group structure as shown in Figure 2. Whenever a fresh node is created, on the same time a new class structure is defined and all the respective groups are attached to it. The class maintains the list of all associated groups. The node array is also created that is sorted based on region size, means initially all classes have largest possible regions and small regions are attached later during process of merging. The extra number of small regions may cause issues for the corresponding application. If the pixel counts in a region assigned to a class are less than a particular threshold M, then the complete class

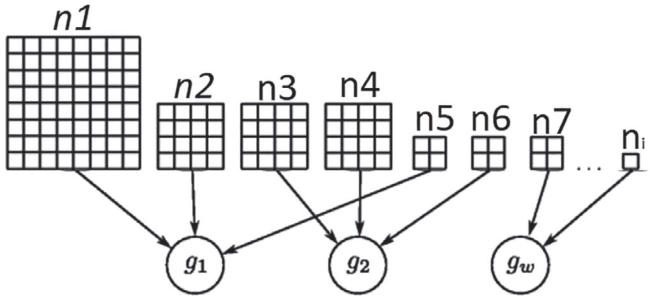

FIGURE 2 Group structure formation from nodes

is removed from the segmented image. These declassified regions are attached to other class (may be background) using region growing method as depicts in Figure 3.

The second type of network that we employed is context encoding network that utilizes the power of CNN for finding the segmented region. Later, we combined both the results based on mask branch.

3.1.8 | Contextual encoding network

Image segmentation task is quite different from the image classification as it assigns more than one labels in a single image. Furthermore, as compared to the object detection model, which gives the boundary boxes of multiple objects, it provides more precise information related to the visual contents. Segmentation is classified into the two different categories; first one is the semantic segmentation and other one is the instance segmentation. In semantic segmentation pixel is classified based on the particular classes present in the image. For example, if there is a car, bus, tree, and so forth present in the images. It uniquely identifies and separate each class based on the pixel. Whereas, in instance segmentation, distinct objects of the same class have been separated by giving the different colors. In this proposed work, we have utilized the semantic segmentation technique to segment out the input image. In recent times, a remarkable progress has been seen in the pixel labeling based on the spatial resolution using the fully convolutional neural network. In our proposed methodology, we have taken the advantage of the latest deep learning based architecture of semantic segmentation.[39] This architecture analyses the impact of the global contextual information on the sematic segmentation. It is based on the contextual encoding module which captures the semantic features of the scene present in the image along with finding the class dependent feature map. In the next section, detailed information has been described of the proposed architecture which is based on the contextual encoding for semantic segmentation.

3.1.9 | Contextual encoding for semantic segmentation

Contextual information is really important in any type of semantic segmentation. To extract the information about the objects, present in the image, a pre-trained CNN has been utilized. This network is trained on the ImageNet[41] dataset. This dataset contains the variety of the images related to the different objects and categories. An encoding layer has introduced in the architecture to extract the statistics of the global sematic features. The output which we get from the encoding layer, termed as the encoded semantics. Different scaling factors for different objects present in image are predicted by using the contextual information. This process helps to identify the class dependent feature map. The defined encoding layer learns the pattern of the contextual information from the dataset and outputs the encoder containing the rich information related to the contextual information. The encoding layer picks up the features map in the shape of the $C \times H \times W$ features which is defined as $F = (f_1, f_2, f_3, \dots, f_N)$. Here C presents

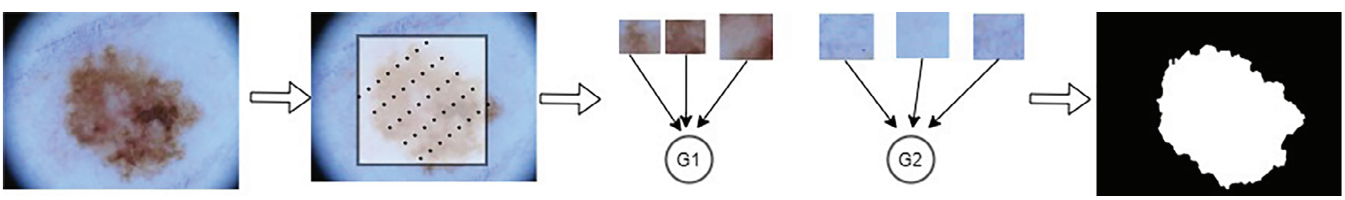

FIGURE 3 Segmentation pipeline based on split and merge method

the number of channels, whereas H and W represent the height and width, respectively. Furthermore, F contains the total number of features. These features are learnable and trained on the codebook $C = c_1, c_2, c_3, \ldots, c_k$ using the K number of code words present in it. It also contains the smoothing factor for detected objects as $S = s_1, s_2, s_3, \ldots, s_k$. This encoding layer outputs the residual by combining with the weights. This has been calculated using Equation (6) using the residual r_{ik} as follows;

$$E_{ik} = \frac{\exp(-\beta||r_{ik}||^2)}{\sum_{j=1}^{K} \exp(-\beta||r_{ik}||^2)}, \tag{6}$$

where E represents the encoding, r depicts residuals, and s shows the smoothing factor. This residual output obtained after applying the batch normalization and ReLU activation layer. This process helps to reduce the feature dimensionality in order to achieve the good results. Figure 4 shows the complete architecture for sematic segmentation using the convolutional neural network.

3.1.10 | Attention mechanism on feature map

In order to utilize the encoded semantics, which we get through the encoding layer, scaling factors against each class or object's feature map have been predicted. This will give us the idea or feedback to whether to emphasize or not on a particular feature map to identify the class dependability. This process has been achieved by applying the fully connected layer on the top of the encoding layer. After that sigmoid activation function has been applied, the algorithm predicts the scaling factors of the feature map. This strategy of getting feedback against feature map is taken through the inspiration of the work done by the style transfer technique. They tuned the feature map scale and generate the statistics.[42] The complete architecture of context encoding has been shown in Figure 4.

3.1.11 | Loss function for semantic encoding

In most of the semantic segmentation algorithms, the loss is calculated by using the cross entropy loss function. It has been calculated by measuring the difference of the pixel's value from the input image and ground truth label. The drawback of this loss function is elimination of global semantic contextual information, which latterly affects the overall performance and accuracy of the algorithm. So, to overcome this problem, proposed architecture has also utilized the semantic encoding loss function that also calculates the global semantic information. This is considerably noted, that additional loss function takes very less computation cost. The additional fully connected layers above the semantic encoding layer calculates the loss value from the prediction of the object presence in the visual scene. Unlike the cross entropy loss function of per pixel, it considers both large and small objects equally. Practically, by considering the global semantic information it has been found that we obtain good results on the sematic segmentation of the small objects.

3.1.12 | Context encoding architecture

In this section, we specify the detail of the contextual encoding architecture. The contextual encoding architecture have been developed using the pre-trained ResNet.[43] The dilated network strategy has been proposed to take the advantage of the pre-trained network. The contextual encoding module have been developed just before the CNN gives the prediction as depicted in Figure 4. To improve the performance and regularizing the

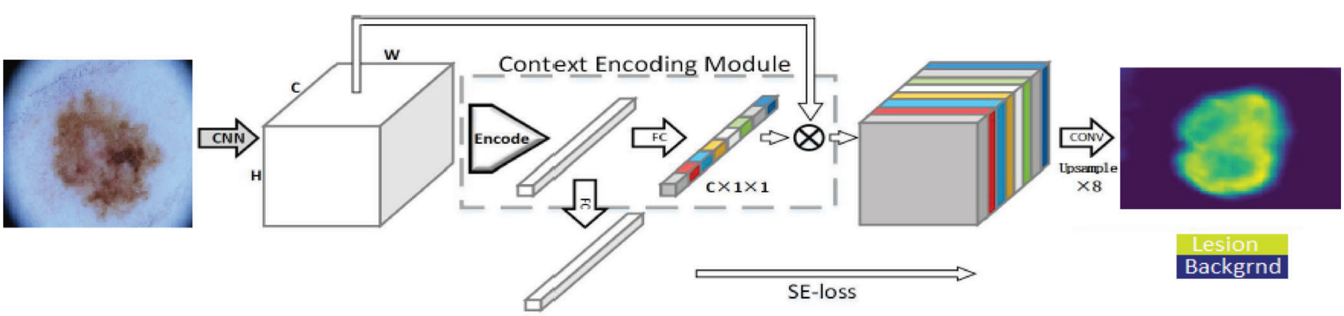

FIGURE 4 Context encoding architecture for sematic segmentation

training process of the context encoding module, a separate layer has been introduced to minimize the semantic encoding loss. This layer takes the semantic encoded features as an input and predicts the object in those features. As both context encoding module and semantic encoding loss are very light weight, so we place additional semantic encoding loss layers. This will help to reduce more semantic encoding loss. This loss function is much similar but much cheaper as compared to the auxiliary loss.[44] Ground truth mask of semantic encodings loss have been extracted using the ground truth labels mask annotations without providing any additional information. The proposed contextual encoding architecture is very flexible and easily inserted into the already existed fully CNN architecture. This process does not require any modification or necessary adjustment along with the supervision on the original framework. If we look at the computational cost, then it increases only marginally to the original computation cost of the utilized CNN architecture. We have set the batch size of 32, and for learning rate we utilized dynamic strategy. Initially, we set the learning rate 10^{-1} and it is decreased with 10^{-1} based on the loss and accuracy of training and validation data. The final step of the whole segmentation process is to combine the results from both the streams. We have employed a simplest technique to combine the results from both techniques. We exploited intersection over union (IoU) for combining traditional results with deep learning results. This approach provides the overlap between two areas that is considered as final segmented region provided their IoU is more than 0.8, meaning very few portion is left out from both segments. Figure 5 shows the approach used for selecting the overlapping region, form the features which we got through traditional machine learning approach and deep learning approach.

3.2 | Classification of the segmented region

After the segmentation of the disease, selected regions contain the disease are obtained. Now, the next task is to classify the detected region. This article proposed the novel approach of concatenating the convolution features of two pre-trained architectures which are Alex-Net and VGG-16 as shown in Figure 6. Both of these architectures are pre trained on the ImageNet dataset. This dataset contains the 15 million images of 1000 classes. Our proposed system will classify the images in to three skin lesion classes, that is, melanoma, nevus, and SK. In deep learning based CNN architectures, one class required at least 10,000 instances to train the model from the scratch. Since, this defined dataset contains the limited amount of the instances per class, so to achieve the good results transfer learning mechanism have been utilized. In both of described architectures, the input images are preprocessed into the 224×224 size. Our first pre-trained architecture is AlexNet. This architecture contains the total eight layers, which contains five convolution layers and three fully connected layers. Convolutions and max pooling have been performed on the filter size of 11×11, 5×5, and 3×3. In addition to that, each convolution layer also contains the ReLU activation and drop out function. Optimization of the weights done with the stochastic gradient descent. The proposed architecture in this article only picks the convolution features from the last convolution layer and rest of the fully connected network have been discarded. On the other hand, VGG-16 consists of total 16 layers. In VGG-16, there are six different blocks. The first five blocks contain the convolution layers whereas the sixth block has the three fully connected layers. The proposed classification architecture only utilized the convolution blocks for feature extraction. Each convolution layers utilized the small size filters of 3×3 and 2×2 in whole architecture. Each convolution layer also consists of the max pooling layer and ReLU activation function. This architecture has also been pre-trained on the ImageNet dataset. Likewise, to AlexNet, from this architecture, we picked the convolution features from the last convolution block. To classify our given dataset, the algorithm contains the linear classification model using the concatenated convolution features of both architectures. After that, there are three fully connected layers. These layers contain the 512, 256, and 128 nodes, respectively. Each fully connected layer has adjacent ReLU activation function along with the drop out layer. This will prevent the model to go towards the overfitting.

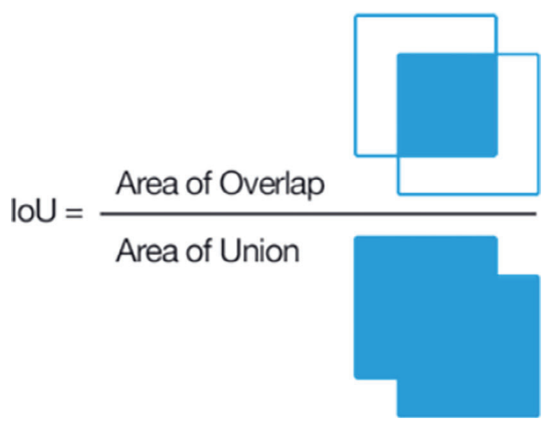

FIGURE 5 Intersection over union for selecting the segmented region

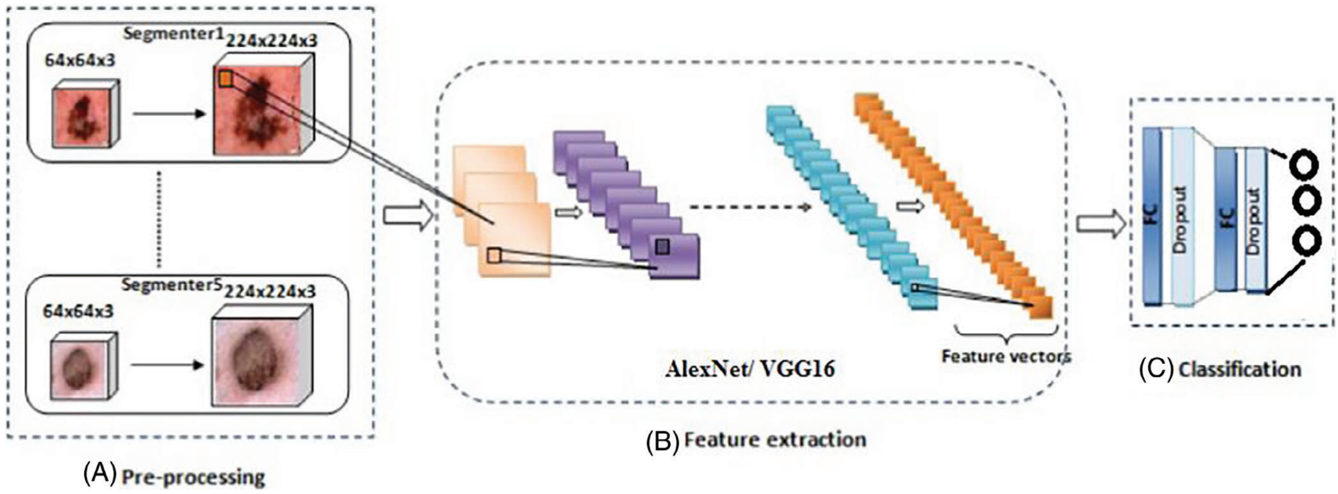

FIGURE 6 Classification module using the concatenated features of CNN architectures

Cross entropy loss function has been utilized to find out the difference between the actual and the predicted value. Whereas, stochastic gradient descent have been utilized for the weights optimization with the momentum weight of 0.9. The learning rate has been changed dynamically. Initially, we have set 10^{-1}, and it have been increase with the -1 factor depending upon the loss value. We have set the batch size of 32. At last, we have three nodes and the softmax classification function gives us the probability against each class present in our dataset. The workflow of the proposed methodology is described in Figure 7. Initially, the image is passed into the two different architectures of segmentation. The above architecture is based on traditional machine learning features, and the below one is based on the deep learning. The output result of both architectures is then passed to the IoU. From IoU, we got the segmented image, then it is passed to the classification network to find out the label of the segmented region.

4 | DATASETS

In this research, we have used two datasets to train and evaluate the performance of proposed system. First one is International Skin Imaging Collaboration (ISIC) which is largest publicity available dataset with considerable amount of instances.[45] It contains three different classes having label of melanoma, nevus and SK. From 2000 total images; 254 images belong to SK class, 1372 images are nevus, and 374 instances of melanoma. Apart from that, this dataset also provides the annotation of the segmented region. In this dataset, against each image which is in JPEG form, there is also a segmented region for ground truth label. Moreover, there is also a CSV file along with the dataset which provide the ground truth labels for the classification task. This dataset is used for the evaluation of both segmentation and classification. Second dataset is PH2[46] which has been used for the evaluation of segmentation only. This dataset includes 200 dermoscopic images along with clinical explanations. From these 200 images; 40 images are malignant melanomas, 80 normal nevi, and remaining 80 are atypical nevi. The segmentation results are evaluated by comparing the outcome of proposed method with the ground truth segmented images from two benchmark datasets ISIC and PH2. For classification type of skin lesion, proposed method results are compared with class labels provided by ISIC dataset.

5 | EXPERIMENTS AND EVALUATION

The evaluation of both segmentation and classification has been performed using the qualitative and quantitative measures. The proposed architectures have been implemented using the deep learning library tensor-flow in the python language. Whereas, dataset splits into the 80% training portion and 20% images on validation side. Training and validation graphs of both accuracy and loss for segmentation and classification has been shown in Figures 8 and 9, respectively. Both architectures are trained till the convergence of the loss.

Left side of Figure 8 shows the training and validation accuracy of the contextual encodings. The accuracy is started with 10.2% at 100th epochs, after that it ramped exponentially and ends on 98.25% training accuracy and 97.21% validation accuracy. Similarly, right side of Figure 8 depicts the training and validation loss, both of these values decreases gradually which depicts the robustness of used hyper parameters in our proposed architecture. Moreover, similarly left side of Figure 9 shows the training and validation accuracy along with the proposed hybrid classification archi-

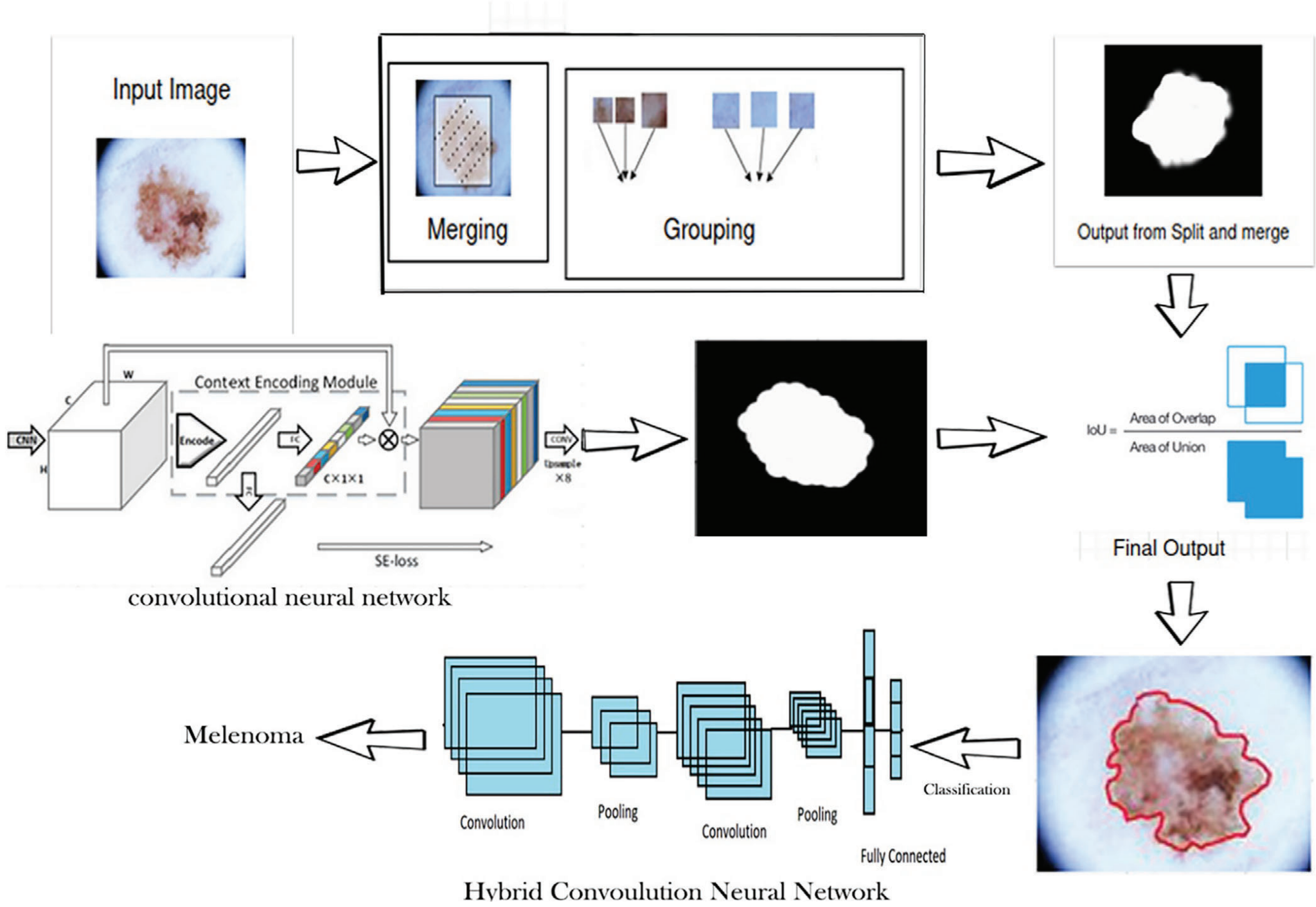

FIGURE 7 Proposed architecture of the system for skin lesion segmentation and classification

tecture. Whereas right side of Figure 9 shows the training and validation loss on the classification architecture. In both cases, convergence has been done moderately. We have achieved 98.40% training accuracy and 97.01% validation accuracy on our proposed classification network. We have utilized the early stop function to stop the training. Early stop functions help to stop the model's training going toward the over-fitting. We find the suitable training curves for both accuracy and loss at 5000 epochs for segmentation and 100 epochs for classification network with the help of early stop function. Early stop function has been utilized in both of these architectures.

5.1 | Segmentation evaluation

As mentioned earlier that both quantitative and qualitative analysis has been performed, Figures 10–12 present visual results presented by proposed segmentation technique whereas quantitative results are presented in Tables 1–4. Figure 10 depicts the segmentation results which our algorithm produced by taking the intersection over union of the split and merge method with the contextual encoding. In Figure 10, first columns represent the input images, second and third one shows output result after the split and merge method and contextual encodings, respectively. Whereas the last column depicts the output image after taking the intersection over union of the split and merge method with the contextual encoding. It is clear from the visual results presented in Figure 10 that cancerous part of the image is segmented in almost all the images. Moreover, Figures 11 and 12 show the results of our proposed model segmentation with the ground truth labels of ISIC and Ph2 dataset, respectively. The blue rounded line depicts the ground truth label and the red line shows the region which our hybrid model predicts. It has been clearly observed from all three figures that algorithm almost clearly identify the effected region with a minimal error.

The evaluation matrix for quantitative analysis of segmentation includes accuracy (ACC), true negative rate (TNR), true positive rate (TPR), Dice similarity coefficients (DSC), and Jaccard similarity coefficients (JSC). We have evaluated pixel level accuracy for both foreground and the

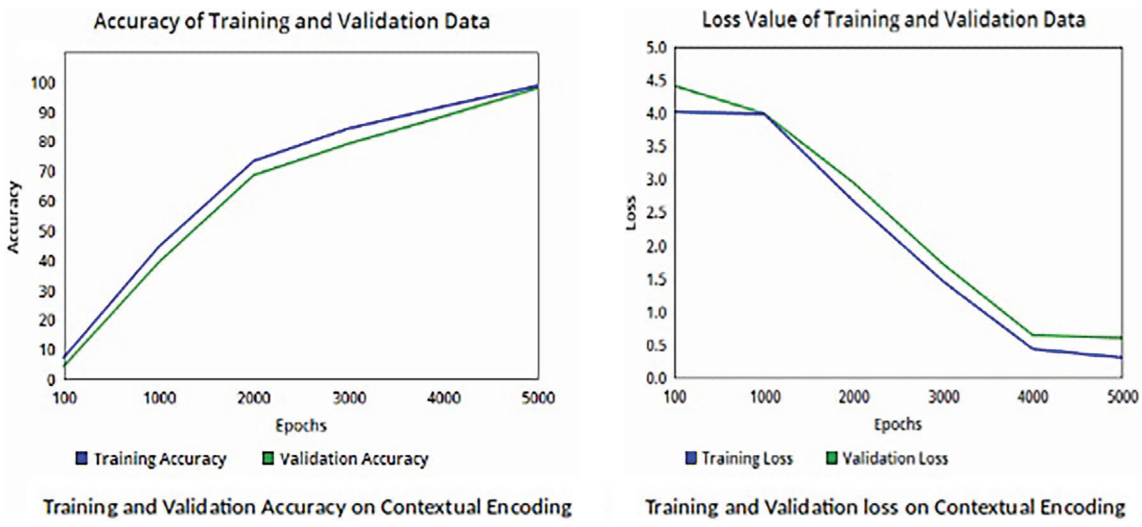

FIGURE 8 Training and validation for accuracy and loss on contextual encoding in segmentation

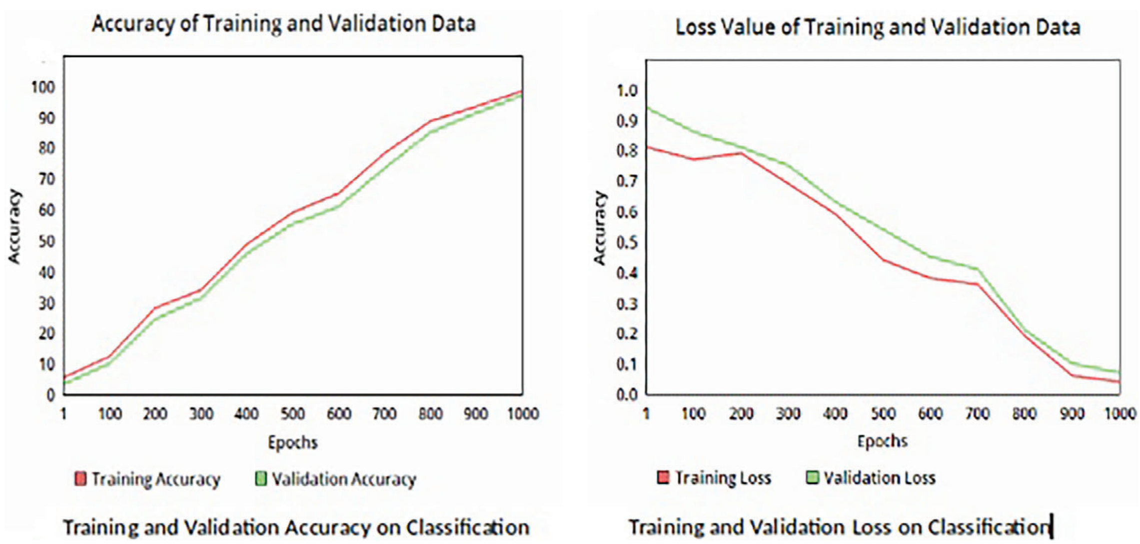

FIGURE 9 Training and validation for accuracy and loss on classification

background using Equation (7).

$$A_{ccseg}\ (\%) = \frac{N_{TP} + N_{TN}}{N_{TP} + N_{TN} + N_{FP} + N_{FN}}. \tag{7}$$

The number of background and foreground pixels that are correctly classified is represented by N_{TP} and N_{TN}, respectively, similarly, background and foreground pixels that are wrongly classified are shown by N_{FP} and N_{FN}, respectively. We have also measured the accuracy of our proposed architecture using the TPR or sensitivity and TNR or specificity. TPR is calculated by dividing true positive values with the total samples in the evaluation as presented in Equation (8).

$$\text{TPR}\ (\%) = \frac{N_{TP}}{N_{TP} + N_{FN}}. \tag{8}$$

Similarly, TNR calculated by dividing the true negative values with the total instances as shown in Equation (9).

$$\text{TNR}\ (\%) = \frac{N_{TN}}{N_{TN} + N_{FP}}. \tag{9}$$

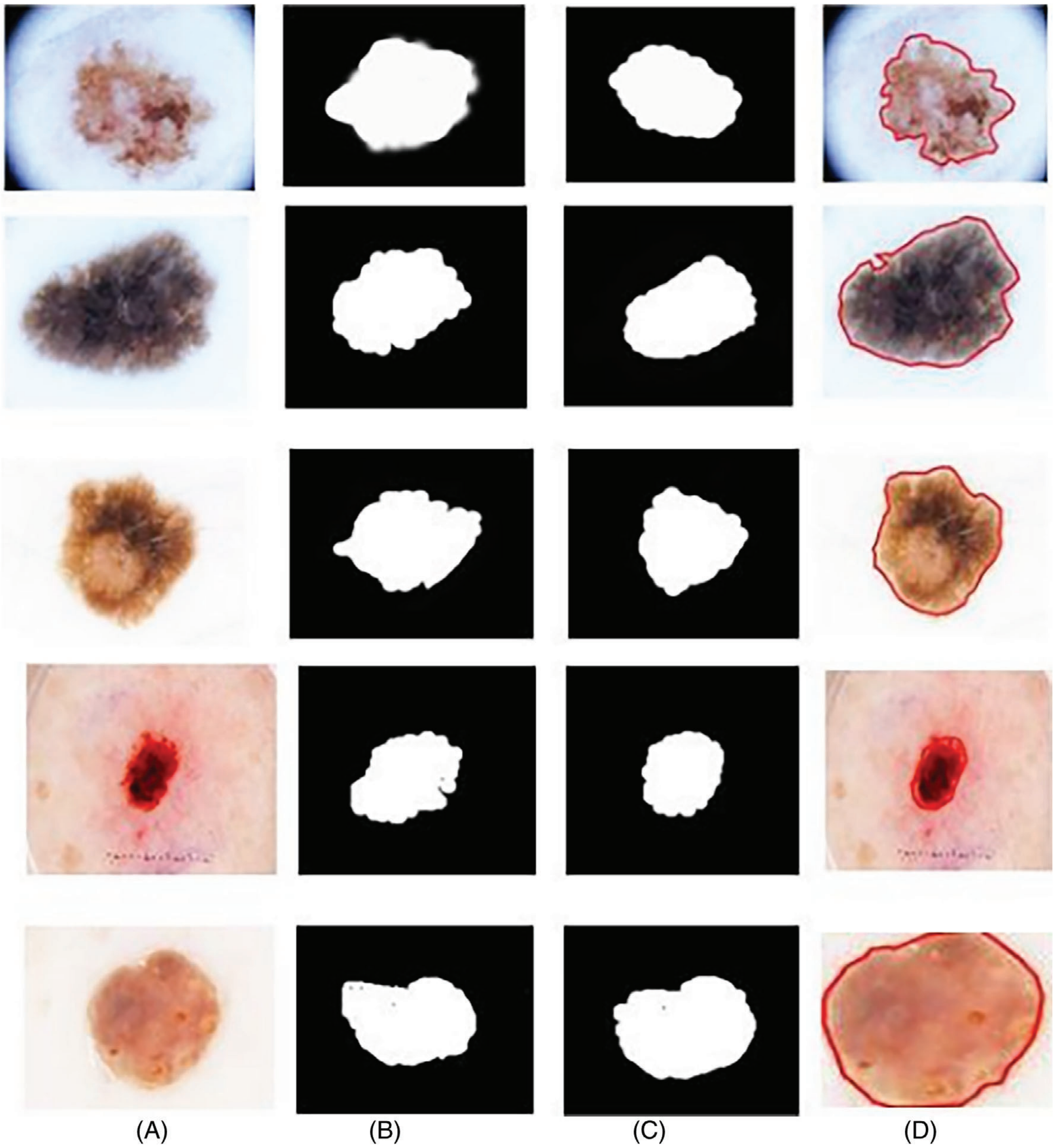

FIGURE 10 Segmentation results by proposed method: (A) Input image for the segmentation process, (B) split and merge method, (C) contextual encoding, and (D) segmented image after IoU

DSC is the measure of overlapped segmented image part and JSC is the percentage of similarity between segmented images. DCS and JSC presented in Equations (10) and (11), respectively.

$$\text{DSC}\ ,(\%) = \frac{2N_{\text{TP}}}{2N_{\text{TP}} + N_{\text{FP}} + N_{\text{FN}}}, \tag{10}$$

$$\text{JSC}\ (\%) = \frac{N_{\text{TP}}}{N_{\text{TP}} + N_{\text{FP}} + N_{\text{FN}}}. \tag{11}$$

The confusion matrix on segmentation for both datasets ISIC and PH2 dataset have been shown in Tables 1 and 2, respectively. Usually, a confusion matrix is generated by taking the average of each N_{TP}, N_{TN}, N_{FP}, N_{FN} with the number of actual positive and negative pixels, respectively. Our proposed segmentation method accurately classifies the foreground pixels with true positive accuracy of 98.2% and background pixels with true negative accuracy of 97.4% as shown in Table 1. The results given in Table 1 shows that more background pixels are classified as foreground pixels as compare to foreground pixels classified as background. These results are quite demanding as no information related to melanoma regions will be dropped. Likewise, segmentation results on PH2 dataset is presented in Table 2 which shows that 97.1% of foreground pixels are classified

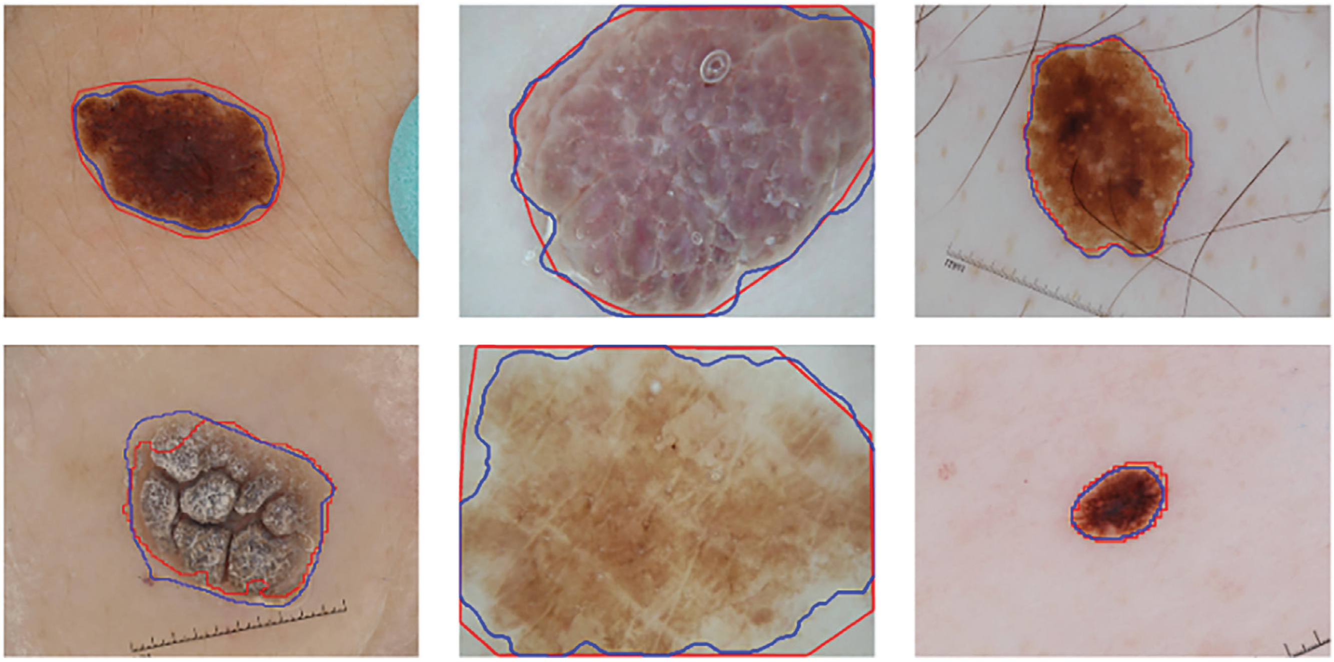

FIGURE 11 Predicted segmentation mask (red line) versus ground truth masks (blue line) on ISIC dataset

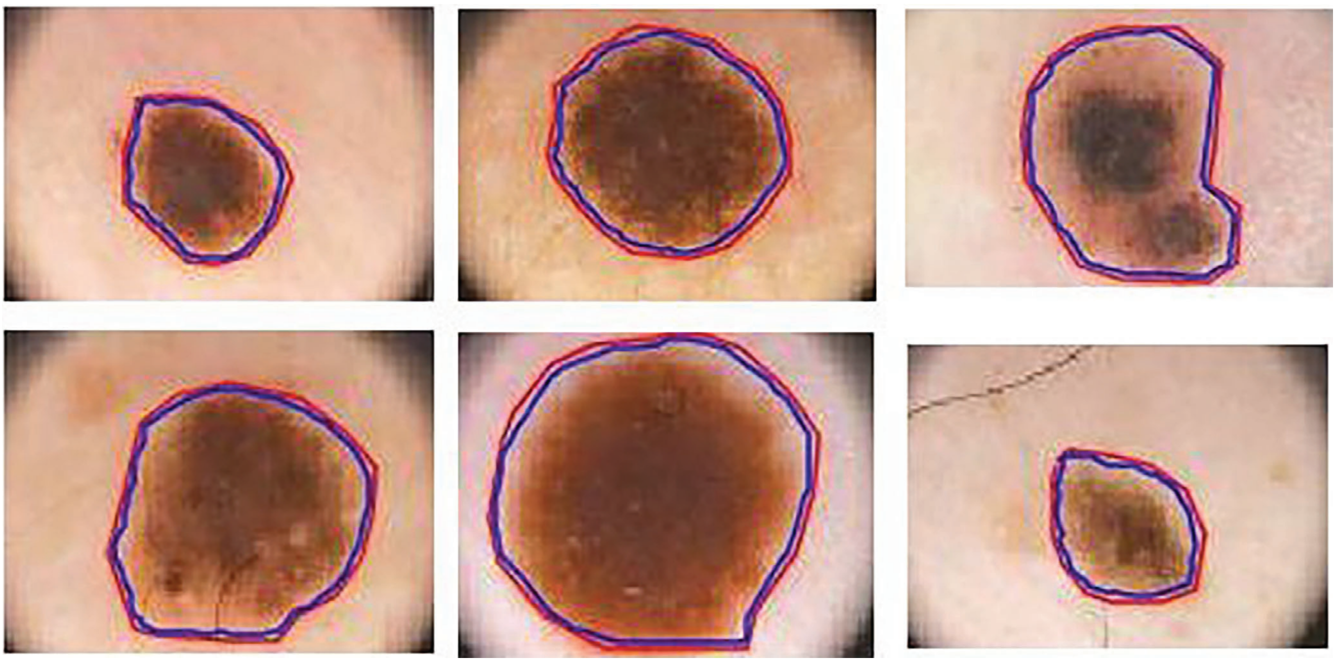

FIGURE 12 Predicted segmentation mask (red line) versus ground truth masks (blue line) on PH2 dataset

correctly whereas 2.9% of foreground pixels are misclassified as background pixels, resulting very minor loss of useful information. Similar results were observed with ISIC dataset images that is more background pixels are classified as foreground pixels.

Moreover proposed technique is compared with state of art segmentation methods and results are presented in Tables 3 and 4 for ISIC and PH2 dataset, respectively. The proposed model achieved highest accuracy with respect to all benchmark techniques, that is, 97.8% and 96.7% for ISIC and PH2 dataset, respectively. Moreover, it has been clearly observed from Tables 3 and 4, that on five different evaluation measures our proposed outperformed the other compared techniques. The benchmark algorithms used in both tables are based on traditional machine learning and latest neural networks based techniques with reasonable accuracies. In Table 3, the evaluation on ISIC dataset has been performed using the utilized

TABLE 1 Confusion matrix on ISIC dataset

ISIC	Foreground	Background
Foreground	98.2%	1.8%
Background	2.6%	97.4%

TABLE 2 Confusion matrix on PH2 dataset

PH2	Foreground	Background
Foreground	97.1%	2.9%
Background	3.7%	96.3%

TABLE 3 Comparison of proposed segmentation technique with benchmark techniques on ISIC dataset

Paper	Accuracy (%)	TNR (%)	TPR (%)	DSC (%)	JSC (%)
47	91.9%	98.5%	80.5%	81.2%	79.3%
18	96.0%	91.3%	83.4%	84.2%	81.1%
48	94.6%	80.3%	87.4%	85.4%	83.2%
19	94.4%	90.0%	92.0%	83.2%	81.9%
49	93.8%	96.4%	87.0%	85.1%	82.4%
50	77.5%	78.2%	76.8%	82.4%	76.3%
51	92.9%	98.1%	81.5%	84.2%	76.6%
52	93.6%	97.8%	82.0%	85.4%	77.9%
53	93.7%	99.0%	80.2%	84.8%	76.9%
54	93.6%	98.5%	82.1%	85.0%	76.8%
Proposed	97.8%	98.2%	97.4%	87.1%	79.2%

TABLE 4 Comparison of proposed segmentation technique with benchmark techniques on PH2 dataset

Paper	Accuracy (%)	TNR (%)	TPR (%)	DSC (%)	JSC (%)
55	93.6%	91.9%	80.5%	81.2%	76.3%
56	89.6%	97.2%	80.2%	–	–
50	87.0%	89.0%	85.0%	82.4%	76.3%
51	90.0%	94.3%	80.5%	85.4%	74.3%
52	92.1%	95.2%	81.3%	85.2%	76.2%
53	93.2%	96.3%	81.3%	81.5%	77.4%
54	94.2%	97.2%	83.2%	83.3%	77.3%
57	91.9%	91.3%	92.5%	–	–
58	92.2%	95.1%	80.8%	–	–
Proposed	96.7%	97.1%	96.3%	88.1%	81.2%

segmentation network. We have achieved highest values on accuracy, TPR and DSC, whereas difference for TNR and JSC is also minimal. Furthermore, in Table 4, we achieved the state of the art results on accuracy, TNR, DSC, TPR, and JSC as compared to the mentioned algorithms. ROC curves for segmentation on both utilized dataset have been depicted in Figure 13.

5.2 | Classification evaluation

We have performed experimentation on three different neural network architectures as specify in Table 6. First, we finetuned the AlexNet on ISIC dataset. We achieved the 93.2% accuracy on the validation data. Whereas, for the second experimentation, we choose the VGG-16. It contains the more layers and learning parameters as compared to the AlexNet. We are able to obtain the 94.5% accuracy on this architecture. Since, there is not too much difference in terms of accuracy in both of these described architectures. So, we decided to utilize the hybrid features of both of these neural network architecture. On our proposed hybrid architecture we are able to achieve the 98.2% accuracy on the validation data of the ISIC dataset. Evaluation matrix used for classification evaluation includes accuracy, precision, recall, F1, false negative rate (FNR), and response time. These evaluation parameters are defined in Equations (12)–(15).

$$\text{Precision }(\%) = \frac{N_{TP}}{N_{TP} + N_{FP}}, \tag{12}$$

$$\text{Recall }(\%) = \frac{N_{TN}}{N_{TN} + N_{FP}}, \tag{13}$$

$$F_1\ (\%) = \frac{2 \times (\text{Recall} \times \text{Precision})}{\text{Recall} + \text{Precision}}, \tag{14}$$

$$\text{FNR }(\%) = \frac{N_{FN}}{N_{TP} + N_{FN}}. \tag{15}$$

For the evaluation of the proposed hybrid classification architecture on the three different classes present into the ISIC dataset, we have utilized the confusion matrix. The confusion matrix for the proposed architecture of the classification network have been shown in Table 5.

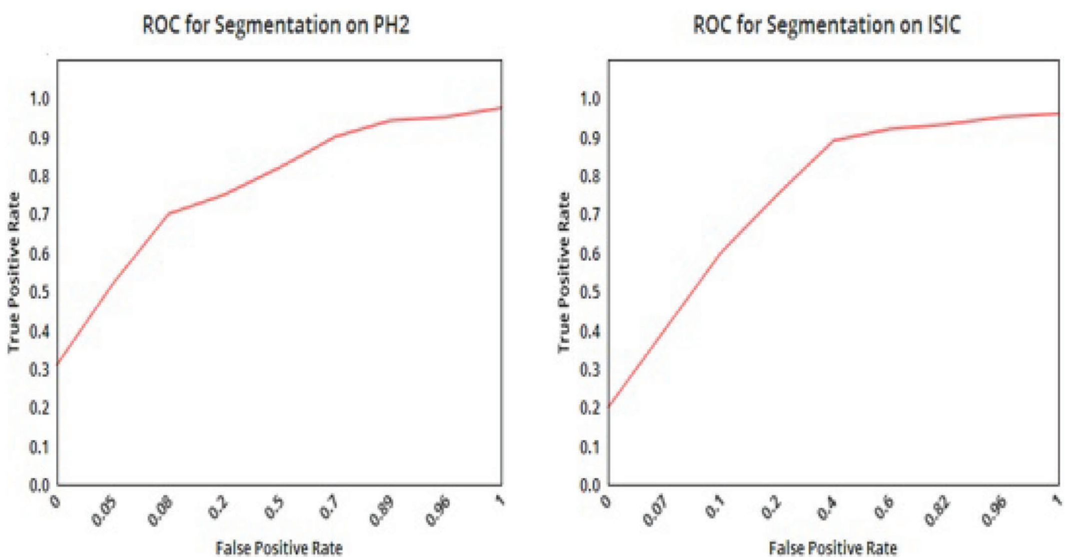

FIGURE 13 ROC for segmentation on PH2 and ISIC dataset

TABLE 5 Confusion matrix on ISIC dataset for proposed classification network

ISIC	Melanoma	Nevus	SK
Melanoma	98.2%	1.8%	1.8%
Nevus	2.6%	97.4%	1.8%
SK	2.6%	97.4%	1.8%

TABLE 6 Classification results on purposed techniques

Classifier	Accuracy (%)	FNR (%)	Precision (%)	Recall (%)	F1-score (%)
AlexNet	93.2%	6.8%	93.0%	93.4%	93.1%
VGG-16	94.5%	5.5%	94.2%	94.8	94.4
Hybrid AlexNet and VGG-16	98.2%	1.8%	98.0%	98.5%	98.2%

TABLE 7 Comparison of classification techniques with benchmark techniques

Paper	Accuracy (%)	TNR (%)	TPR (%)	Response time (s)	Dataset
19	96.5%	97.0%	96.5%	16.46	ISIC
18	93.7%	95.0%	93.3%	38.61	ISIC
59	72.1%	–	–	–	ISIC
60	87.0%	–	–	–	ISIC
Proposed	98.2%	98.3%	97.5%	22.44	ISIC

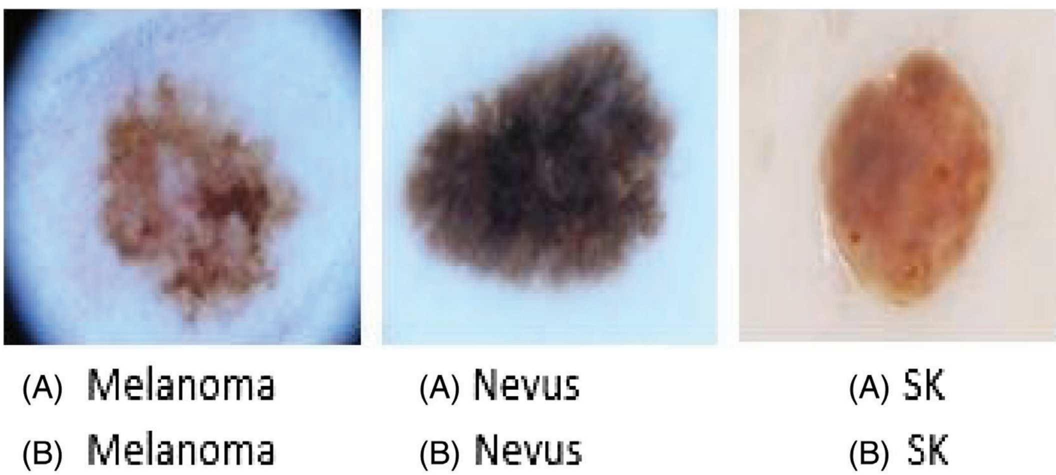

FIGURE 14 Classification results on ISIC dataset: (A) True label and (B) predicted label

Table 6 illustrates the evaluation results done using the ISIC dataset on different and proposed classifier architectures. These experiments are performed using evaluation parameters as shown in Equations (12)–(15). It has been clearly observed from Table 6 that our proposed hybrid architecture achieved the 98.2% accuracy which is far better than the other architecture's and main reason of this improvement was combined features.

Moreover, we also compared our proposed architecture technique with the benchmark techniques in Table 7. We have achieved state of the art results on the ISIC dataset at our proposed hybrid architecture. Although performance of algorithm presented in Reference 19 is better with respect to response time but overall accuracy is still less than proposed method. Furthermore, some predictions have also been displayed in Figure 14. Figure 14A,B represents the predicted label and the ground truth label, respectively. The correct predictions show the strength of our proposed classification architecture. The ROC for the classification task on different thresholds using the ISIC dataset have been shown in Figure 15.

Since our proposed approach utilized the amalgamated features through different machine learning and deep learning algorithms, so it is quite computationally expensive. We have utilized Nvidia 1080 Ti GPU for training purpose. Although, we achieved the good accuracy, but for the future directions we can do the optimization for our proposed architecture. Moreover, this algorithm has huge impact in medical industry for automatic skin lesion detection and classification.[60] So, for medical industry point of view, if we want to add some new class into our Dataset, then we have to train from the scratch which is computationally expensive and time taking. So, we can extend this research in few shot learning direction.

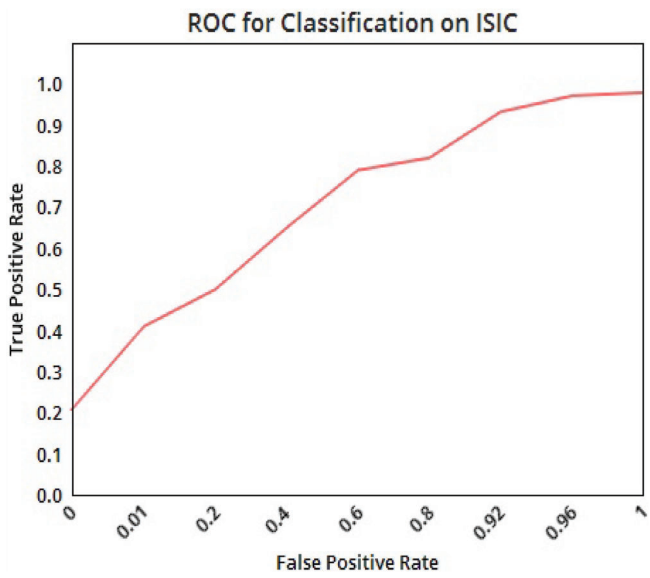

FIGURE 15 ROC for classification on ISIC dataset

6 | CONCLUSION

In this article, the proposed architecture segment out and classifies the skin lesion. We have proposed the novel architecture based on the ensemble learning of hand-crafted features with the automated feature selection. Our algorithm based on the two steps. In the first step, we combine the features using split and merge method with the contextual encoding along with the attention mechanism. The outcome for segmentation process has been obtained using the intersection over union. Whereas, in the second step, classification has been done with hybrid features of the fine-tuned AlexNet and VGG16. We have achieved the remarkable good results through the novel approach of combining the features using two different techniques. We are not only combining the features of machine learning and deep learning but we did also the amalgamations of different architectures of convolutional neural networks. ISIC and PH2 datasets have been used for experimentation. To check the learning during training, we closely monitored the training and validation accuracy along with the loss values. We have achieved 97.8% and 96.7% accuracy for segmentation on ISIC and PH2 dataset, respectively and 98.3% accuracy on classification using ISIC dataset. For evaluation purpose, specificity, accuracy, sensitivity, Jaccard index score, dice coefficient, and confusion matrix have been utilized. In most of the evaluation techniques, our proposed algorithms out-performed some state-of-the-art methods. Overall, it has been concluded that proposed ensemble learning approach outweigh the individual latest statistical as well as deep learning-based algorithms for segmentation and classification in term of accuracy. In future, we have a plan to extend our work by enhancing the dataset using data augmentation techniques or by utilizing the few shot approaches of deep learning.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in International Skin Imaging Collaboration (ISIC) at https://www.isic-archive.com/#!/topWithHeader/wideContentTop/main.

ORCID

Adil H. Khan https://orcid.org/0000-0003-1676-362X
Hiren Mewada https://orcid.org/0000-0002-3579-8708

REFERENCES

1. Miller KD, Nogueira L, Mariotto AB, et al. Cancer treatment and survivorship statistics, 2019. *CA Cancer J Clin*. 2019;69(5):363-385.
2. Vasicek BE, Szpunar SM, Manz-Dulac LA. Patient knowledge of sunscreen guidelines and frequency of physician counseling: a cross-sectional study. *J Clin Aesthet Dermatol*. 2018;11(1):35.
3. Mocellin S, Nitti D. Cutaneous melanoma in situ: translational evidence from a large population-based study. *Oncologist*. 2011;16(6):896.
4. Wollina U. Seborrheic Keratoses–The most common benign skin tumor of humans. clinical presentation and an update on pathogenesis and treatment options. *Open Access Maced J Med Sci*. 2018;6(11):2270.
5. Wang X, Liu Q, Wu L, Nie Z, Mei Z. Risk of non-melanoma skin cancer in patients with psoriasis: an updated evidence from systematic review with meta-analysis. *J Cancer*. 2020;11(5):1047.
6. Halpern AC, Marghoob AA, Sober AJ, Mar V, Marchetti MA. Clinical presentations of melanoma. *Cutaneous Melanoma*. Springer, Cham; 2020;107-144. https://link.springer.com/referenceworkentry/10.1007/978-3-030-05070-2_9#citeas

7. Janda M, Horsham C, Vagenas D, et al. Accuracy of mobile digital teledermoscopy for skin self-examinations in adults at high risk of skin cancer: an open-label, randomised controlled trial. *Lancet Digit Health*. 2020;2(3):e129-e137.
8. Reinehr CPH, Bakos RM. Actinic keratoses: review of clinical, dermoscopic, and therapeutic aspects. *An Bras Dermatol*. 2020;94:637-657.
9. Oliveira A, Arzberger E, Massone C, Carrera C, Zalaudek I. Verrucous melanoma simulating melanoacanthoma: dermoscopic, reflectance confocal microscopic and high-definition optical coherence tomography presentation of a rare melanoma variant; 2016.
10. Gandini S, Sera F, Cattaruzza MS, et al. Meta-analysis of risk factors for cutaneous melanoma: I. Common and atypical naevi. *Eur J Cancer*. 2005;41(1):28-44.
11. Moreno-Ramírez D, Raya-Maldonado J, Morales-Conde M, et al. Increasing frequency of seborrheic keratosis diagnoses as a favorable consequence of teledermatology-based skin cancer screening: a cross-sectional study of 34,553 patients. *Am J Clin Dermatol*. 2017;18(5):681-685.
12. Marghoob NG, Liopyris K, Jaimes N. Dermoscopy: a review of the structures that facilitate melanoma detection. *J Osteopath Med*. 2019;119(6):380-390.
13. Zaqout I. Diagnosis of skin lesions based on dermoscopic images using image processing techniques. *Pattern RecognitSelect Methods Appl*. 2019. https://doi.org/10.5772/intechopen.88065
14. Celebi ME, Iyatomi H, Stoecker WV, et al. Automatic detection of blue-white veil and related structures in dermoscopy images. *Comput Med Imaging Graph*. 2008;32(8):670-677.
15. Saba T, Al-Zahrani S, Rehman A. Expert system for offline clinical guidelines and treatment. *Life Sci J*. 2012;9(4):2639-2658.
16. Nida N, Irtaza A, Javed A, Yousaf MH, Mahmood MT. Melanoma lesion detection and segmentation using deep region based convolutional neural network and fuzzy C-means clustering. *Int J Med Inform*. 2019;124:37-48.
17. Ullah H, Saba T, Islam N, et al. An ensemble classification of exudates in color fundus images using an evolutionary algorithm based optimal features selection. *Microsc Res Tech*. 2019;82(4):361-372.
18. Afza F, Khan MA, Sharif M, Rehman A. Microscopic skin laceration segmentation and classification: a framework of statistical normal distribution and optimal feature selection. *Microsc Res Tech*. 2019;82(9):1471-1488.
19. 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 Systems*. 2019. http://doi.org/10.1111/exsy.12497
20. Anjum MA, Amin J, Sharif M, Khan HU, Malik MSA, Kadry S. Deep semantic segmentation and multi-class skin lesion classification based on convolutional neural network. *IEEE Access*. 2020;8:129668-129678.
21. Akram T, Lodhi HMJ, Naqvi SR, et al. A multilevel features selection framework for skin lesion classification. *HCIS*. 2020;10(1):1-26.
22. Peruch F, Bogo F, Bonazza M, Cappelleri VM, Peserico E. Simpler, faster, more accurate melanocytic lesion segmentation through meds. *IEEE Trans Biomed Eng*. 2013;61(2):557-565.
23. Masood A, Al-Jumaily A. Orientation sensitive fuzzy C means based fast level set evolution for segmentation of histopathological images to detect skin cancer. Proceedings of the International Conference on Hybrid Intelligent; 2018:501-510; Springer, Cham.
24. Pereira PM, Fonseca-Pinto R, Paiva RP, et al. Dermoscopic skin lesion image segmentation based on local binary pattern clustering: comparative study. *Biomed Signal Process Control*. 2020;59:101924.
25. Abbas Q, Celebi ME, Fondón García I, Rashid M. Lesion border detection in dermoscopy images using dynamic programming. *Skin Res Technol*. 2011;17(1):91-100.
26. Correa-Tome FE, Sanchez-Yanez RE. Integral split-and-merge methodology for real-time image segmentation. *J Electron Imaging*. 2015;24(1):013007.
27. Al-Masni MA, Al-Antari MA, Choi MT, Han SM, Kim TS. Skin lesion segmentation in dermoscopy images via deep full resolution convolutional networks. *Comput Methods Prog Biomed*. 2018;162:221-231.
28. Murugan A, Nair SAH, Kumar KS. Detection of skin cancer using SVM, random forest and KNN classifiers. *J Med Syst*. 2019;43(8):1-9.
29. Kang C, Yu X, Wang SH, et al. A heuristic neural network structure relying on fuzzy logic for images scoring. *IEEE Trans Fuzzy Syst*. 2020;29(1):34-45.
30. Li Y, Shen L. Skin lesion analysis towards melanoma detection using deep learning network. *Sensors*. 2018;18(2):556.
31. Khatami A, Nazari A, Khosravi A, Lim CP, Nahavandi S. A weight perturbation-based regularisation technique for convolutional neural networks and the application in medical imaging. *Expert Syst Appl*. 2020;149:113196.
32. Javed R, Rahim MSM, Saba T, Rehman A. A comparative study of features selection for skin lesion detection from dermoscopic images. *Netw Model Anal Health Inf Bioinform*. 2020;9(1):1-13.
33. Mahmouei SS, Aldeen M, Stoecker WV, Garnavi R. Biologically inspired QuadTree color detection in dermoscopy images of melanoma. *IEEE J Biomed Health Inform*. 2018;23(2):570-577.
34. Hosny KM, Kassem MA, Foaud MM. Classification of skin lesions using transfer learning and augmentation with Alex-net. *PLoS One*. 2019;14(5):e0217293.
35. Khan MA, Sharif M, Akram T, Damaševičius R, Maskeliūnas R. Skin lesion segmentation and multiclass classification using deep learning features and improved moth flame optimization. *Diagnostics*. 2021;11(5):811.
36. Khan MA, Sharif M, Akram T, Bukhari SAC, Nayak RS. Developed Newton-Raphson based deep features selection framework for skin lesion recognition. *Pattern Recogn Lett*. 2020;129:293-303.
37. Al-Masni MA, Kim DH, Kim TS. Multiple skin lesions diagnostics via integrated deep convolutional networks for segmentation and classification. *Comput Methods Prog Biomed*. 2020;190:105351.
38. Divya D, Ganeshbabu T. Fitness adaptive deer hunting-based region growing and recurrent neural network for melanoma skin cancer detection. *Int J Imaging Syst Technol*. 2020;30(3):731-752.
39. Zhang H, Dana K, Shi J, et al. Context encoding for semantic segmentation. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2018:7151-7160; IEEE.
40. Chaudhuri D, Agrawal A. Split-and-merge procedure for image segmentation using bimodality detection approach; 2010.
41. Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L. Imagenet: a large-scale hierarchical image database. Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition; 2009:248-255; IEEE.
42. Huang X, Belongie S. Arbitrary style transfer in real-time with adaptive instance normalization. Proceedings of the IEEE International Conference on Computer Vision; 2017:1501-1510; IEEE.
43. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2016:770-778; IEEE.

44. Zhao H, Shi J, Qi X, Wang X, Jia J. Pyramid scene parsing network. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2017:2881-2890.
45. Reiter O, Chousakos E, Kurtansky N, et al. Association between the dermoscopic morphology of peripheral globules and melanocytic lesion diagnosis. *J Eur Acad Dermatol Venereol*. 2021;35(4):892-899.
46. Mendonca T, Celebi M, Mendonca T, Marques J. Ph2: a public database for the analysis of dermoscopic images. *Dermoscopy Image Anal*. 2015:419–439. https://www.researchgate.net/profile/Catarina-Barata-2/publication/300467916_PH2_A_Public_Database_for_the_Analysis_of_Dermoscopic_Images/links/570fa8cf08ae170055bde6db/PH2-A-Public-Database-for-the-Analysis-of-Dermoscopic-Images.pdf
47. Jaisakthi SM, Mirunalini P, Aravindan C. Automated skin lesion segmentation of dermoscopic images using GrabCut and K-means algorithms. *IET Comput Vis*. 2018;12(8):1088-1095.
48. Sultana NN, Mandal B, Puhan NB. Deep residual network with regularised fisher framework for detection of melanoma. *IET Comput Vis*. 2018;12(8):1096-1104.
49. Xie F, Yang J, Liu J, Jiang Z, Zheng Y, Wang Y. Skin lesion segmentation using high-resolution convolutional neural network. *Comput Methods Prog Biomed*. 2020;186:105241.
50. Hasan MK, Dahal L, Samarakoon PN, Tushar FI, Martí R. DSNet: automatic dermoscopic skin lesion segmentation. *Comput Biol Med*. 2020;120:103738.
51. Wen H. II-FCN for skin lesion analysis towards melanoma detection; 2017. arXiv preprint arXiv:1702.08699.
52. Yuan Y. Automatic skin lesion segmentation with fully convolutional-deconvolutional networks; 2017. arXiv preprint arXiv:1703.05165.
53. Bi L, Kim J, Ahn E, Feng D. Automatic skin lesion analysis using large-scale dermoscopy images and deep residual networks; 2017. arXiv preprint arXiv:1703.04197.
54. Jahanifar M, Tajeddin NZ, Gooya A, Asl BM. Segmentation of lesions in dermoscopy images using saliency map and contour propagation. arXiv; 2017.
55. Fan H, Xie F, Li Y, Jiang Z, Liu J. Automatic segmentation of dermoscopy images using saliency combined with Otsu threshold. *Comput Biol Med*. 2017;85:75-85.
56. Pennisi A, Bloisi DD, Nardi D, Giampetruzzi AR, Mondino C, Facchiano A. Skin lesion image segmentation using delaunay triangulation for melanoma detection. *Comput Med Imaging Graph*. 2016;52:89-103.
57. Hu K, Niu X, Liu S, et al. Classification of melanoma based on feature similarity measurement for codebook learning in the bag-of-features model. *Biomed Signal Process Control*. 2019;51:200-209.
58. Abayomi-Alli OO, Damasevicius R, Misra S, Maskeliunas R, Abayomi-Alli A. Malignant skin melanoma detection using image augmentation by oversampling in nonlinear lower-dimensional embedding manifold. *Turk J Electr Eng Comput Sci*. 2021;29(SI-1):2600-2614.
59. Yap J, Yolland W, Tschandl P. Multimodal skin lesion classification using deep learning. *Exp Dermatol*. 2018;27(11):1261-1267.
60. Alom MZ, Aspiras T, Taha TM, Asari VK. Skin cancer segmentation and classification with NABLA-N and inception recurrent residual convolutional networks; 2019. arXiv preprint arXiv:1904.11126.

How to cite this article: Khan AH, Awang Iskandar DNF, Al-Asad JF, Mewada H, Sherazi MA. Ensemble learning of deep learning and traditional machine learning approaches for skin lesion segmentation and classification. *Concurrency Computat Pract Exper*. 2022;e6907. doi: 10.1002/cpe.6907