

RESEARCH

Open Access



A multilevel features selection framework for skin lesion classification

Tallha Akram¹, Hafiz M. Junaid Lodhi¹, Syed Rameez Naqvi^{1*}, Sidra Naeem¹, Majed Alhaisoni², Muhammad Ali¹, Sajjad Ali Haider¹ and Nadia N. Qadri¹

*Correspondence:
rameeznaqvi@ciitwah.edu.pk
¹ Department of Electrical
and Computer Engineering,
COMSATS University
Islamabad, Wah Campus,
Wah Cantonment, Pakistan
Full list of author information
is available at the end of the
article

Abstract

Melanoma is considered to be one of the deadliest skin cancer types, whose occurring frequency elevated in the last few years; its earlier diagnosis, however, significantly increases the chances of patients' survival. In the quest for the same, a few computer based methods, capable of diagnosing the skin lesion at initial stages, have been recently proposed. Despite some success, however, margin exists, due to which the machine learning community still considers this an outstanding research challenge. In this work, we come up with a novel framework for skin lesion classification, which integrates deep features information to generate most discriminant feature vector, with an advantage of preserving the original feature space. We utilize recent deep models for feature extraction, and by taking advantage of transfer learning. Initially, the dermoscopic images are segmented, and the lesion region is extracted, which is later subjected to retrain the selected deep models to generate fused feature vectors. In the second phase, a framework for most discriminant feature selection and dimensionality reduction is proposed, entropy-controlled neighborhood component analysis (ECNCA). This hierarchical framework optimizes fused features by selecting the principle components and extricating the redundant and irrelevant data. The effectiveness of our design is validated on four benchmark dermoscopic datasets; PH2, ISIC MSK, ISIC UDA, and ISBI-2017. To authenticate the proposed method, a fair comparison with the existing techniques is also provided. The simulation results clearly show that the proposed design is accurate enough to categorize the skin lesion with 98.8%, 99.2% and 97.1% and 95.9% accuracy with the selected classifiers on all four datasets, and by utilizing less than 3% features.

Keywords: Skin lesion, Convolutional neural network, Dermoscopy, Deep learning, Feature selection, Transfer learning.

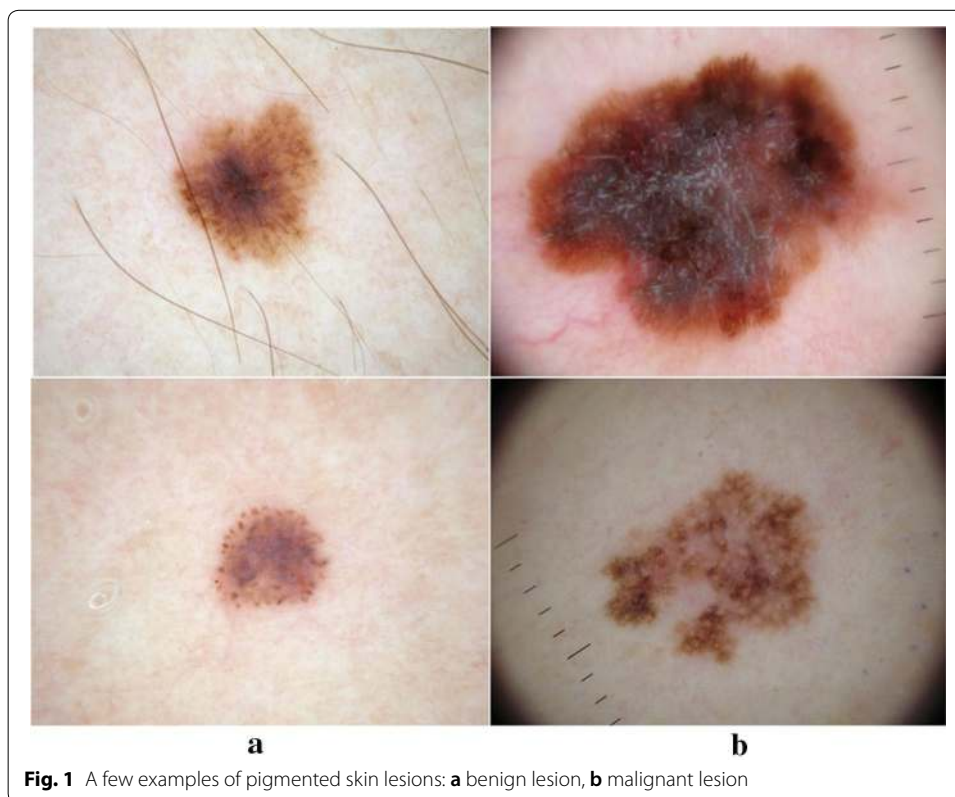
Introduction

Melanoma belongs to the category of inoperable type of skin cancers, and its occurrence rate has increased tremendously over the past three decades [1]. According to statistics provided by the World Health Organization (WHO), almost 132,000 new cases of melanoma are reported each year worldwide. It has been reported [2] that diagnosis of melanoma, in its early stages, significantly increases chances of the patient's survival. Dermatoscopy, also known as dermoscopy is a non-invasive clinical procedure used for

melanoma detection, in which physicians apply gel on the affected skin, prior to examining it with a dermoscope. It allows recognition of sub-surface structures of the infected skin that are invisible to naked eye. With this clinical procedure, the skin lesion is amplified up to 100 times, thereby easing the examination [3].

For the diagnosis of melanoma, dermatologists mostly rely on ABCD rule [4], seven-point checklist [5], and Menzie's method [6]. These aforementioned methods have been formally approved at the 2000 Consensus Net Meeting on Dermoscopy (CNMD) [7], and are widely exploited by the physicians for diagnostics. Even though, these methods of manual inspection have shown improved performance, due to a number of constraints, including a large number of patients, human error and infrastructure etc., they have not proven feasible. Additionally, melanoma, at its initial stages, exhibits a similar type of features like benign lesions, which makes it difficult to recognize; Fig. 1 presents two such examples. Furthermore, physician analysis may also be quite subjective, since it clearly depends on their clinical experience and human vision as well—making the diagnosis procedure quite challenging.

To handle such constraints, there still exists a requirement for an automated system that has a capacity to differentiate melanoma from benign at its very initial stages. Computer-aided diagnosis (CAD) system maybe useful for the physicians to use technological developments in the field of dermoscopy, and it may also provide a second opinion. The CAD systems adopt various machine learning techniques, for example, extracting various features (color, shape, and texture) from each dermoscopic image, followed by applying a state-of-the-art classifier [8, 9]. These classification approaches mostly rely on



the extracted set of features for the training, which are broadly divided into three main levels: low, mid, and higher levels [10]. Various existing classification methods exploit the extracted features by simply concatenating them in order to generate a fused feature vector. Feature fusion methodology, on one hand, increases the classification accuracy by taking into account all the advantages from the host models, but on the other hand increases the computational time and memory requirements [11].

Recently, convolutional neural networks (CNN) [12] have been introduced in this domain, and their models have been widely accepted for feature extraction—leading to improved classification [13, 14]. In such solutions, discriminant deep features are extracted by using set of convolution, pooling and feedforward layers from the images, by embedding a concept of transfer learning (TL) using fine-tuning and features descriptors [15]. To achieve further improvement in the classification results, in terms of overall accuracy (OA), computational time, and memory, feature selection process plays a pivotal role by identifying the most discriminate features. This is something we exploit in the proposed framework, entropy-controlled neighborhood component analysis (ECNCA), for skin lesion classification. The latter exploits the resilience of deep features and utilizes them in the lower dimensions—preserving the original feature space information. We demonstrate that our approach utilizes less than 3% deep features—equivalent to 97.55% average reduction rate, and is substantively superior to state-of-the-art approaches in terms of OA. Most of the existing literature, to the best of our knowledge, does not reduce the deep features to this level.

The exclusive contributions of this work are enumerated below:

- 1 We exploit behavior of the selected layers of deep architectures, including DenseNet 201, Inception-ResNet-v2, and Inception-V3, on the performance of classifiers.
- 2 We propose to fine-tune the existing pre-trained models with smaller learning rate and keep weights of the initial layers frozen to avoid distortion of the complete model. We exploit feature fusion technique, which takes advantage of all the three selected architectures to generate a denser feature space.
- 3 We propose a hierarchical architecture for feature selection and dimensionality reduction, which in the initial step relies upon entropy for feature selection, followed by dimensionality reduction using neighborhood component analysis (NCA).

The rest of the article is organized as follows. In the following section, "[Literature review](#)" section, we present a detailed overview of the existing literature in this domain. "[Mathematical model](#)" section presents the mathematical model, whereas, materials and methods are discussed in "[Materials and methods](#)" section. The proposed framework is detailed in "[Proposed framework](#)" section, and "[Results and discussion](#)" section contains the experimental results and discussions. We conclude the manuscript in "[Conclusion](#)".

Literature review

In literature, several CAD systems [16, 17] have been proposed for melanoma detection, which, to some extent, try to mimic the procedure performed by dermatologists, based on a range of features extracted using machine learning approaches. These

systems mostly follow four primary steps [18]: (1) preprocessing, (2) lesion segmentation, (3) feature extraction and selection, and (4) classification.

Lesion image segmentation is one of the primary steps that have abiding effects [19] on this classification process. Accurate segmentation of a lesion is an arduous task due to a number of reasons; range of lesion sizes, shapes, colors, and skin texture. Secondly, sometimes there exists a smooth transition between skin color and lesion [19, 20]. In addition to that, a few other constraints include specular reflection, presence of hair, falloff towards the edges, and air and immersion-fluid bubbles. Sumithra [21] proposed to initially remove the unwanted hair from lesion prior to applying the segmentation algorithm. Feature extraction was performed subsequently using color and texture features. For the classification both support vector machine (SVM), and K-nearest neighbor (KNN) were used. Similarly, Attia et al. [22] implemented a hybrid framework for hair segmentation by combining convolutional and recurrent layers. They utilized deep encoded features for hair delineation, which are later fed into recurrent layers to inscribe the spatial dependencies among the incoherent image patches. The segmentation accuracy calculated using Jaccard Index is 77.8% in comparison to the existing methods, 66.5%.

Joseph [23] used fast marching and 2D derivative of Gaussian in painting algorithm for hair artifact removal. Cheerla et al. [24] proposed automatic method for segmentation. They used otsu's thresholding for segmentation, and for texture feature extraction local binary patterns (LBP) [25] was utilized. Neural network classifiers were used for classification, which yielded 97% sensitivity and 93% specificity. Hawas et al. [26] proposed an optimized clustering estimation using neutrosophic graph-cut (OCE-NGC) algorithm for skin lesion segmentation. They made use of bio-inspired technique (genetic algorithm), which optimizes the histogram-based clustering procedure, which searches the optimal centroid/threshold values. In the following step, they grouped the pixels by using the generated threshold value using neutrosophic c-means algorithm. Finally, a graph-cut methodology [27] is implemented to segregate the foreground and background regions in the dermoscopic image. Authors claimed to achieve 97.12% average accuracy and 86.28% average Jaccard values. Similarly, [28] implemented a novel scheme (transform domain representation-driven CNN) for skin lesion segmentation. They trained the model from scratch and successfully managed to cope with the constraints including small data set, artifact removal, excessive data augmentation, and contrast stretching. Authors claimed to achieve 6% higher Jaccard index and a less training time on a publicly available ISBI 2016 and 2017 datasets. Euijoon et al. [29] proposed a saliency [30] based segmentation algorithm, in which detection of background was based on spatial layout including color and boundary information. To minimize detection error, they implemented Bayesian framework.

Features play a vital role in classification, which are extracted by following local, global or local–global scenarios [7]. Barata et al. [31] adopted a local–global method for detecting melanoma from dermoscopic images. Local methods were applied to extract features using bag-of-words, whilst, global methods were explored for the classification of skin lesions. Promising results were achieved in terms of greater sensitivity and specificity. Abbas et al. [32] suggested a perceptually oriented framework for border identification—combining the strengths of both edge and region based segmentation. Later,

a hill-climbing [33] approach was efficiently utilized to identify the region-of-interest (ROI), followed by an adaptive threshold mechanism to detect the optimal lesion border.

Chatterjee et al. [34] proposed a cross-correlation based technique for feature extraction with an application to skin lesion classification. The authors considered both spatial and spectral features of lesion region based on visual coherency using cross-correlation technique. kernel patches are later selected based on the skin disease categories, which are later classified using proposed multi-label ensemble multi-class classifier. The acquired sensitivities of a set of classes including nevus, melanoma, BCC and SK diseases are 99.01%, 98.7%, 98.87%, and 99.41%. Lei et al. [35] proposed a lesion detection and recognition methodology—built on a multi-scale lesion-biased representation (MLR) and joint reverse classification. This proposed algorithm takes advantage of scales and rotations to detect lesion, compared to the conventional single rotation method. Omer et al. [36] provided a unique solution for skin lesion segmentation using global thresholding based on color features. As a following feature extraction step, they utilized 2D fast Fourier transform (2D-FFT) and 2D discrete Fourier transform (2D-DFT). Mahbod et al. [37] introduced an ensemble technique by combining inter and intra-architecture of CNN. The extracted deep features from each CNN network are later utilized in classification using multi-SVM classifiers. The proposed method proved to be robust in terms of feature extraction, fusion and classification for skin lesion images. Kahn et al. [18] presented a techniques for classification of skin lesion using probabilistic distribution, and for feature selection entropy based method was used. Al-masni et al. [38] investigated a set of deep frameworks both for segmentation and classification. Initially, they implemented a full resolution convolution network for lesion segmentation. Later, the lesion regions are used to extricate the features using multiple deep architectures including Inception-ResNet-v2, and DenseNet 201. Proposed framework is trained on three datasets, ISIC 2016, ISIC 2017, and ISIC 2018, to achieving the promising results. Similarly, a pool of researchers [39–41] are utilizing deep frameworks to detect multiple abnormalities with an application to skin lesion classification.

From the detailed review, it is concluded that various existing methods show improved performance on dermoscopic images, but the following conditions were already satisfied:

- 1 High contrast distinctness between the lesion area and the surrounding region.
- 2 Color uniformity inside the lesion area.
- 3 Marginal existence or absence of different artifacts including dark corners, hair, color chart, to name but a few.

Therefore, considering the aforementioned conditions, our primary focus is to develop a technique which efficiently handles the negation of given conditions.

Mathematical model

Given a dermoscopic image database, we are required to assign a label to each and every image—belonging to a class of either benign or malignant. Let us consider $D \subset \mathbb{R}^{(r \times c \times p)}$ be a demoscopic image, $\psi = \psi(j) | j \in \mathbb{R}$ be a formally specified image dataset, where $\left((\psi_1(j), \dots, \psi_k(j) \subset \psi) \in \mathbb{R} \right)$ are the pixel values of k-channels. The number of classes \mathbb{C} is provided by the user, therefore a class is discriminated as $\tilde{\psi}$ —a modified version of ψ ,

interpreted as $\tilde{\psi} : \psi \rightarrow \tilde{\psi}$. The modeling of ψ to achieve output $\tilde{\psi}$ is described in terms of:

$$\tilde{\psi} \triangleq (\psi^f, \psi^{fs}, \kappa(\psi^{fs})) \in \mathbb{Z}^3 \tag{1}$$

where ψ^f represents the extracted features after applying transfer learning, ψ^{fu} represents the fused features from fully connected layers of different architectures, and $\kappa(\psi^{fu})$ is the selected features' representation after processing through a hierarchical structural design.

Materials and methods

Convolutional neural networks

CNN are one of the most powerful deep feedforward neural network models used for object detection and classification [42]. In CNN, all neurons are connected to a set of neurons in the next layer in a feedforward fashion. The CNN's basic architecture, as given in Fig. 2, incorporates three primary sub-blocks, comprising convolution, pooling, and fully connected layers.

- 1 *Convolution layer* A fundamental unit in the CNN architecture, called convolution layer, is supposed to detect and extract local features from an input image sample $X_p^{(r \times c \times p)}$, where $r = c$ for a square input. Let us consider an input image sample, $X_p = \{x_1, x_2, \dots, x_n\}$, where n represents size of the training dataset. For each input image, the corresponding output is $y_p = \{y_1, y_2, \dots, y_n\}$, where $y_p \in \{1, 2, \dots, C\}$, C represents the number of classes. Convolution layer includes a kernel that slides across the input image as $X^{(r \times c \times p)} * H^{(r' \times c' \times p)}$, and local features $f \in f_l$ are extracted using the following relation:

$$\mathbb{F}_i^l = \sigma \left(\sum_{i=1}^n x_i^{l-1} \times \delta_i^l + b_i^l \right) \tag{2}$$

where \mathbb{F}_i^l provides feature map output for the layer, l ; $\omega_i^l + b_i^l$ are the trainable parameters for layer, l ; $\delta(\cdot)$ represents an activation function.

- 2 *Pooling layer* Addition of a pooling layer is another substantial concept in CNN, which is considered to be a non-linear down sampling technique. It is a meaningful combination of two fundamental concepts, max pooling and convolution. Here

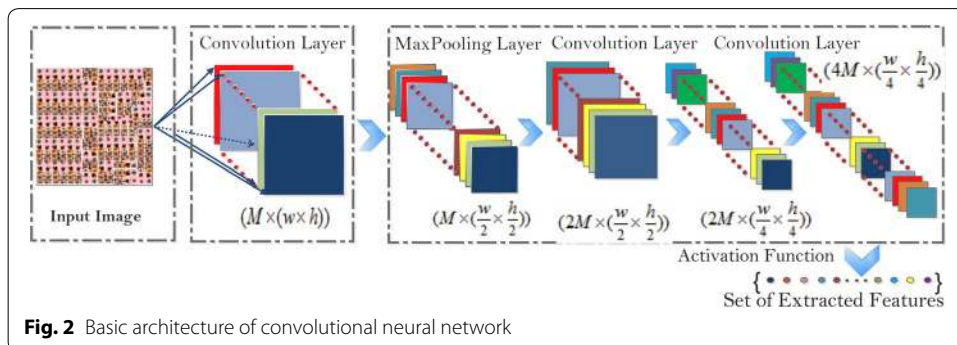


Fig. 2 Basic architecture of convolutional neural network

max-pooling step extracts a set of maximum responses with an objective of feature reduction, as well as robustness against noise and variations. Configuration of max-pooling is represented with the help of the following equation:

$$R_i^l = \max(z_{2i-1}^{l-1}, z_{2i}^{l-1}), \quad l = 2 \zeta \forall \zeta \in \mathbb{R} \tag{3}$$

- 3 *Fully Connected Layer* Convolution and pooling layers are followed by a fully connected feedforward layer, *FC*. It follows the same principle of traditional fully connected feedforward network having set of inputs and output units. This layer extracts responses based on features' weights calculated from the previous layer.

$$V_j^l = \text{Sig} \left(\sum_{i=1}^n x_i^{l-1} \times \omega_{ji}^l \times b_i^j \right) \tag{4}$$

Transfer learning

Conventional algorithms work by making an assumption that the feature characteristics of both training and testing data are quite identical and can be comfortably approximated [43]. Several pretrained models are trained on natural images, and hence not suitable for the specialized applications. Additionally, data collection for the real world applications is a tedious task. Therefore, TL is a solution to provide accurate classification with a limited number of training samples. This concept is briefly defined as a system's capability to transfer the skills and knowledge learnt while solving one class of problems to a different class of problems, (source–target relation), Fig. 3. The real potential of TL may be best leveraged when the target and source domain datasets are highly disparate in size, such that target domain dataset is significantly smaller than the source domain dataset [44]. Given a source domain, $D_S = \{(x_1^S, y_1^S), \dots, (x_i^S, y_i^S), \dots, (x_n^S, y_n^S)\}$, where $(x_n^S, y_n^S) \in \mathbb{R}$; with specified learning tasks, L_S , and target domain $D_T = \{(x_1^T, y_1^T), \dots, (x_i^T, y_i^T), \dots, (x_m^T, y_m^T)\}$ having learning task L_T , $(x_n^T, y_n^T) \in \mathbb{R}$. Let $((m, n) | (n \ll m))$ be a training data size and y_1^D and y_1^T are their respective labels. The fundamental function of TL is to boost the learning capability of the target function D_T —utilizing the knowledge gained from the source D_S and the target D_T .

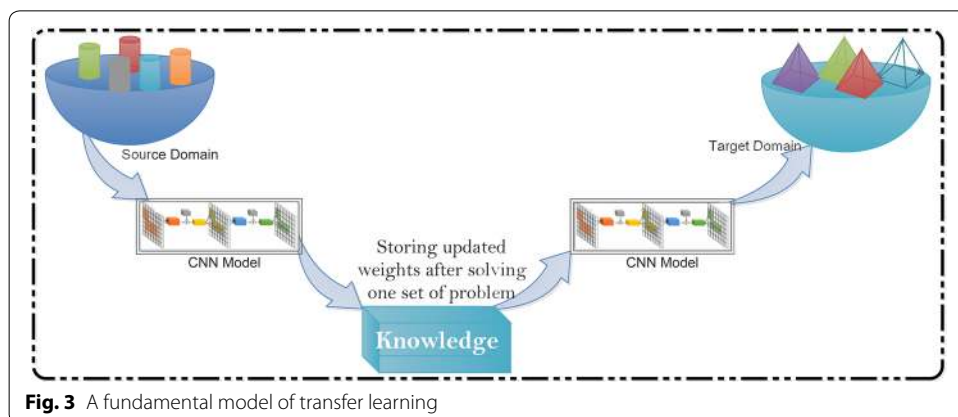


Fig. 3 A fundamental model of transfer learning

Pre-trained CNN models

Several researchers have proposed set of CNN architectures for computer vision applications like segmentation and classification, etc. [53, 54]. In this work, we utilize three widely used pre-trained models for features extraction including Inception-V3, Inception-ResNet-V2 and DenseNet-201. The selection of these models is on the basis of their performance in terms of their Top-1 accuracy, Table 1.

Inception-V3

Inception-V3 is trained on ImageNet database. It comprises two fundamental units: feature extraction and classification. Inception-V3 employs inception units that allow the framework to escalate the depth and width of a network, but also lower the computational parameters.

Inception-ResNet-V2

Inception-ResNet-V2 is an extension of inception-V3, and is also trained on ImageNet database. In its core, it combines the inception with ResNet module. The remaining connections allow bypasses in the model to make the network behave more robustly. Inception-Resnet-v2 fuses the computational adeptness of the Inception units with the optimization leverage contributed by the residual connections.

DenseNet-201

DenseNet 201 is also trained on ImageNet database. It is designed on a more sophisticated connectivity pattern that iteratively integrates all output features in a regular feed-forward fashion. Moreover, it mitigates the vanishing-gradient problem, reduces number of input/functional parameters, and strengthens feature propagation.

Dataset

In this work, we have performed our simulations on four publicly available datasets:

Table 1 Pretrained deep models description: following a yearly sequence

| Model | Year | Source | Top-1 accuracy (ImagNet) (%) |
|--------------------|-----------|-----------|------------------------------|
| AlexNet | 2012 [45] | BVLC | 57.1 |
| VGG16 | 2014 [46] | Oxford | 70.5 |
| VGG19 | 2014 [46] | Oxford | 71.3 |
| Inception V1 | 2015 [47] | Google | 67.9 |
| SqueezeNet | 2016 [48] | DeepScale | 59.5 |
| ResNet 50 | 2016 [49] | MSR | 75.3 |
| ResNet 101 | 2016 [49] | MSR | 76.4 |
| DenseNet201 | 2016 [50] | – | 77.0 |
| Inception V2 | 2016 [51] | Google | 72.2 |
| Inception V3 | 2016 [51] | Google | 76.9 |
| Inception V4 | 2017 [52] | Google | 80.2 |
| InceptionResNet V2 | 2017 [52] | Google | 79.79 |

- 1 *PH²*: This dataset is composed of 200 RGB images, classified as 160 benign and 40 melanoma. These images were collected at the Hospital Pedro Hispano, Matosinhos during clinical examination with the help of dermoscope [55]. The ground truth is also provided, which is segmented manually with the help of physicians; classified as normal, atypical nevus (benign) or melanoma.
- 2 ISIC-MSK: The second dataset used in this research is International Skin Imaging Collaboration (ISIC) [56]. This dataset contains 225 RGB dermoscopic images, acquired from various international hospitals with the help of different devices.
- 3 ISIC-UDA: It is another subdataset of ISIC. We have collected 557 images having 446 training and 111 testing samples from ISIC-UDA dataset.
- 4 ISBI-2017: ISBI-2017 [57] is another publicly available dataset used for characterization of skin cancer in dermoscopic images. It contains 2750 images, with 2200 training and 550 testing samples. The ISBI-2017 dataset has three disease classes: melanoma, keratosis and benign; however, since keratosis is a common benign skin condition, we have divided the samples into two: malignant and benign.

Manual annotations of all datasets, discussed above, by dermatologists have been provided as ground truths for the evaluation purposes. Repartition of above mentioned datasets is shown in Table 2. Note that we have divided the target dataset into two sets with pre-defined 80% for training and 20% for testing. The training set comprises a combination of training set (70%)—used to train the models, and the validation set (10%) for models' evaluation/fine tuning.

Proposed framework

In dermoscopy, cancer classification is still an outstanding challenge, which is efficiently dealt with by the proposed design; discussed below. Most of the constraints enumerated in "Literature review" section are successfully undertaken, and a cascaded framework is proposed, which comprises four fundamental blocks: preprocessing, lesion segmentation, feature extraction and selection, and labeling/classification. Figure 4 summarizes the adopted methodology.

Preprocessing

The preprocessing step copes with image imperfections introduced at the initial step of acquisition, by eliminating multiple artifacts, such as hair or ruler markings. Contrarily, their presence may affect segmentation, which, in turn, leads to an inaccurate classification. Ideally, the collected image should be free from these artifacts, however, due to certain complications, its strenuous to remove the hair. Therefore, an algorithmic

Table 2 Splitting datasets into training and testing samples

| Dataset | Total images | Training/validation set | Testing set |
|-----------------------|--------------|-------------------------|-------------|
| <i>PH²</i> | 200 | 160 | 40 |
| ISIC MSK | 225 | 180 | 45 |
| ISIC UDA | 557 | 446 | 111 |
| ISBI 2017 | 2750 | 2200 | 550 |

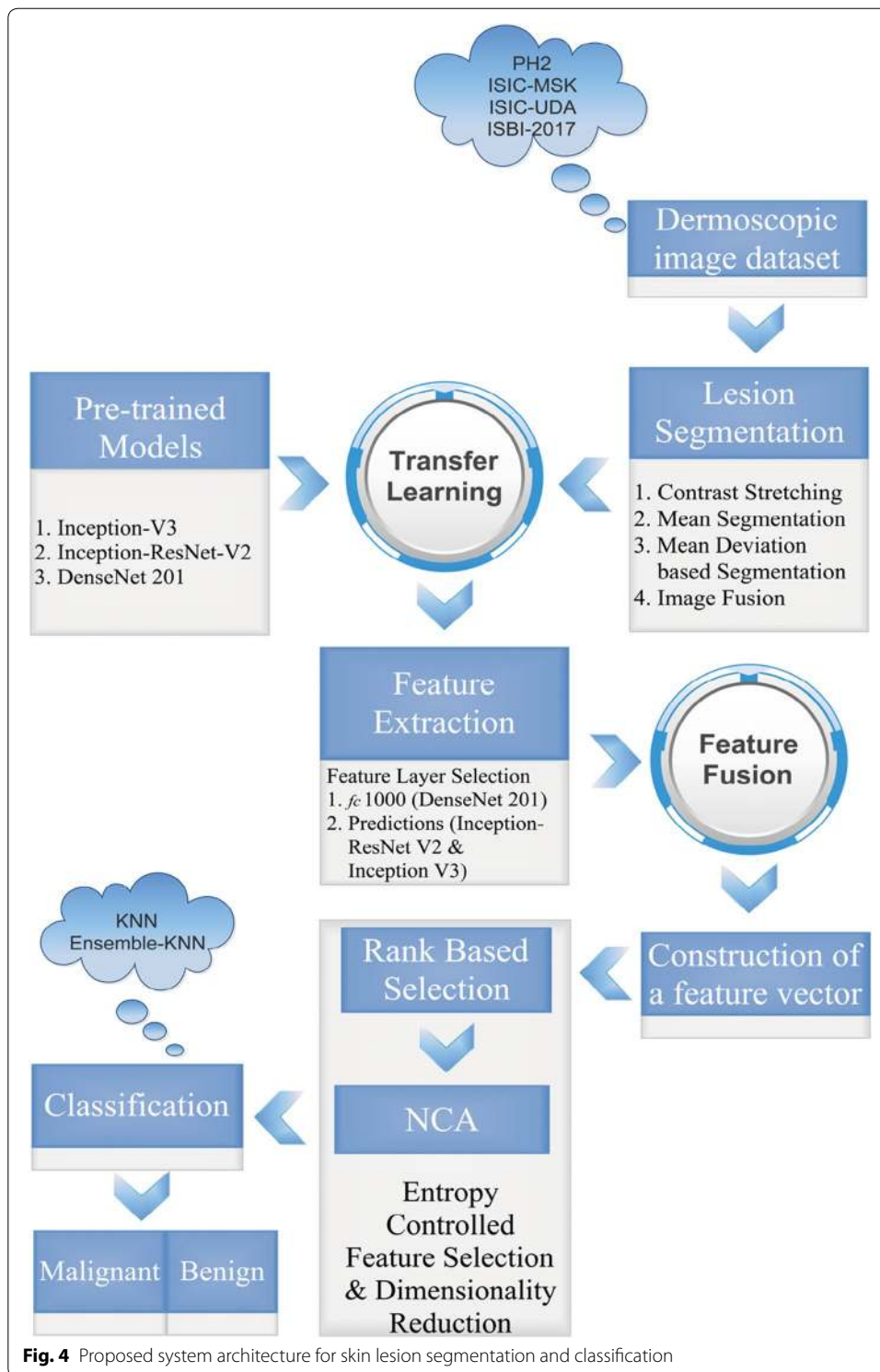


Fig. 4 Proposed system architecture for skin lesion segmentation and classification

approach, rather than the latter, is preferably followed. In this work, a widely used software, Dull Razor [58], is utilized, which is capable of localizing the hair and extricate them by implementing bilinear interpolation. Additionally, it also implements an adaptive median filter to smoothen the replaced hair pixel.

Lesion/image segmentation

Segmentation is one critical step that plays its primary role in classification of the skin lesion. In addition to solving various problems, including color variations, hair presence, and lesion irregularity, a robust segmentation method has a capacity to identify infected regions with improved accuracy. Once the images have been transformed to keep the same aspect ratio, the following two steps are performed in turn to complete the segmentation process:

- 1 Contrast stretching, to make lesion (foreground) region distinct compared to the background.
- 2 Segment the lesion region based on mean and mean deviation based segmentation procedure.

The immediate objective behind implementing contrast stretching scheme is to make foreground (lesion region) maximally differentiable compared to the background. Additionally, introduction of this pre-processing step refines images to much extent which leads to improved classification accuracy [59]. Initially, each channel of a three dimensional RGB image ($I_{\mathbb{D}} \in \mathbb{R}^{r \times c \times p}$) is processed independently to make foreground region visually distinguishable. A series of interlinked steps needs to be followed by each channel; those steps are enumerate below:

- 1 Initially, gradients are computed for each single channel using Sobel–Feldman operator, with a fixed kernel size of (3×3) .
- 2 Divide each channel into equal sized blocks (4, 8, 12, ...), and rearrange them in a descending order. Now weights are assigned to each block according to gradient magnitude.

$$W_{\xi} = \begin{cases} w_b^1 & \text{if } I_s(x, y) \leq \xi_1; \\ w_b^2 & \text{if } \xi_1 < I_s(x, y) \leq \xi_2; \\ w_b^3 & \text{if } \xi_2 < I_s(x, y) \leq \xi_3; \\ w_b^4 & \text{otherwise.} \end{cases} \quad (5)$$

where $w_b^i (i = 1, \dots, 4)$ is a weight coefficient and ξ represents threshold values against computed gradient.

- 3 Compute the overall weighted gray value against each block

$$W_g(b) = \sum_{k=1}^4 \xi_b^i n_k(b) \quad (6)$$

where $n_k(b)$ represents number of gray pixels encased in block k .

To get improved results, few aspects are stringently considered; (a) standard block size, (b) optimized weight criteria, and (c) selection of regions with maximum

information. Upon assiduous examination of dermoscopic images, regions with maximum information (lesion) are in the possible range of 25% to 75%. Therefore, worst case is considered and we partition the image 12 basic cells, with a ratio of 8.3%. Later, based on the criteria of maximum information these cells are selected (summation of pixels against each cell). Finally, according to edge points, weights are assigned for each block, E_p^c .

$$C_{wi} = \frac{E_p^c}{E_{max}^c} \quad (7)$$

where E_{max}^c represents cells with maximum edges. An addition of post log operation further refines the channel [18], $I_c(x, y)$, compared to original, $I_s(x, y)$.

$$I_c^l = C \times \log(\beta + I_c(x, y)) \quad (8)$$

where β is chosen to be 3 by following a trial and error method.

Addition of a contrast stretching block facilitates segmentation step in extracting lesion area with improved accuracy. The probabilistic methods (mean segmentation and mean deviation based segmentation) are applied independently on a same image which are later subjected to image fusion in the following step.

Mean segmentation is calculated using:

$$I(\mu) = \frac{1}{(1 + (\frac{\mu}{I_c}))^\zeta} + \frac{1}{2\mu} + C \quad (9)$$

$$I_\mu^{MS} = \begin{cases} 1 & \text{if } I(\mu) \geq \varphi_{thresh}; \\ 0 & \text{otherwise.} \end{cases} \quad (10)$$

where φ_{thresh} is Otsu's threshold, ζ is a scaling factor—selected to be 7 by following trial and error method. C is a constant and its value is in the range of 0 to 1.

Similarly mean deviation based segmentation is also calculated on enhanced image by following an activation function, having σ_{MD} calculated to be 0.7979 by following trial and error method.

$$I(\kappa) = \frac{1}{(1 + (\frac{\sigma_{MD}}{I_c}))^\zeta} + \frac{1}{2\sigma_{MD}} + C \quad (11)$$

$$I_{(\kappa, \sigma^2)}^{MD} = \begin{cases} 1 & \text{if } I(\kappa) \geq \varphi_{thresh}; \\ 0 & \text{otherwise.} \end{cases} \quad (12)$$

Segmented image from both distributions are later fused to get the resultant image.

$$I_{seg} = I_{(\kappa, \sigma^2)}^{MD} \cap I_\mu^{MS} \quad (13)$$

Sample segmentation results are provided in Fig. 5, where it can be observed that they are visually similar when compared with the available ground truths. In some

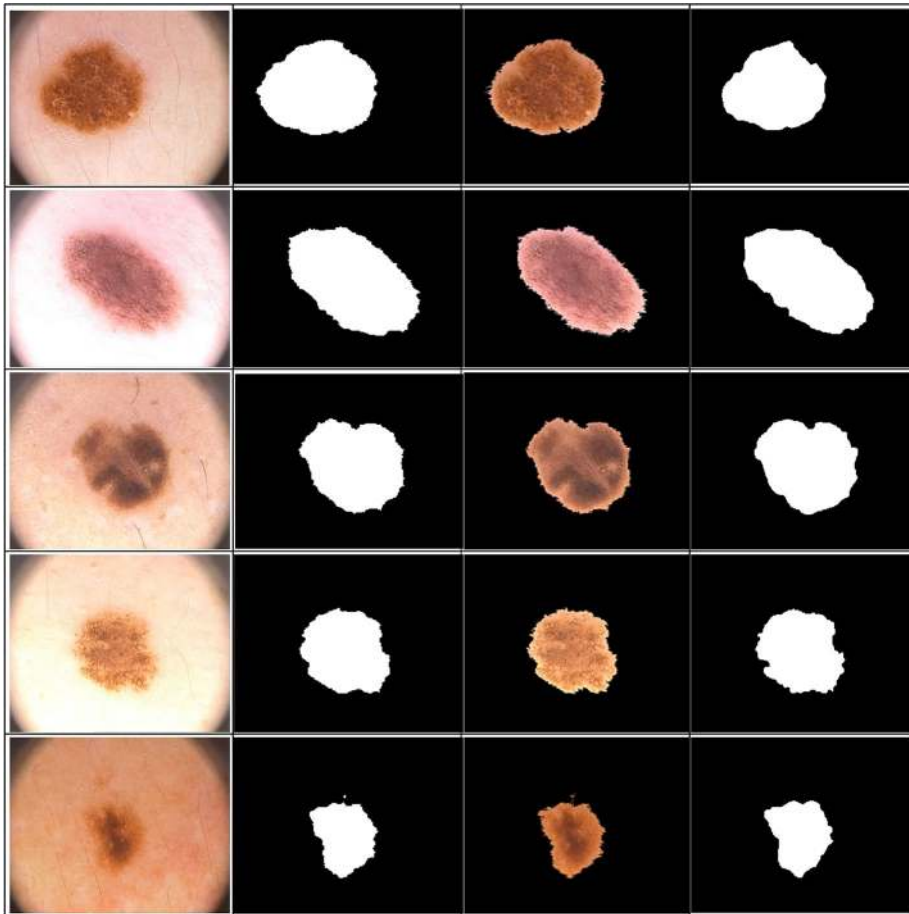


Fig. 5 Segmentation results: (1) original image; (2) fused segmented image; (3) mapped RGB image; (4) ground truth

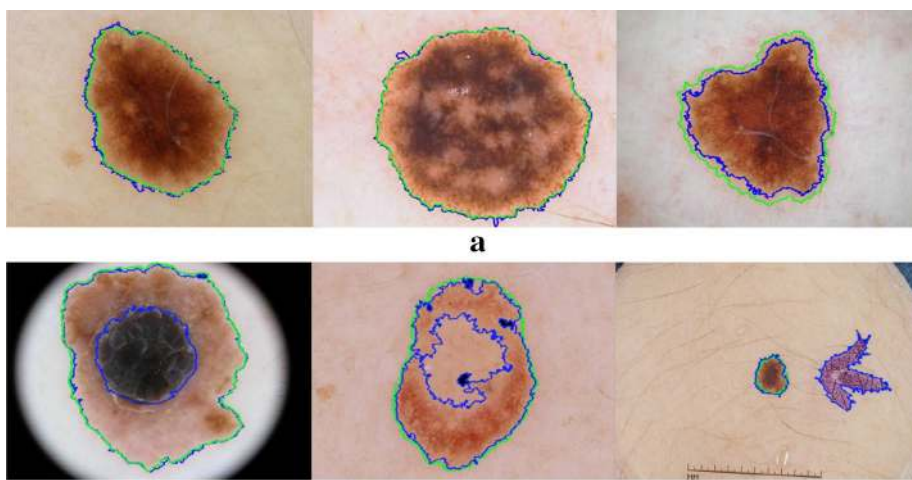


Fig. 6 Segmented skin lesions: **a** correctly segmented, **b** incorrectly segmented

cases, the foreground and background are not distinct enough; the segmentation, in such cases, does not pan out sufficiently acceptable. This may be correlated with the images given in Fig. 6.

Deep features extraction

The proposed framework can be observed in Fig. 4, showing various stages from extraction to the final classification. Following the segmentation step, the proposed hierarchical design is applied on the extracted set of features to conserve the salient deep features.

Feature layers

It has been observed that the systems relying on deep features extracted from a single layer and utilizing a single pre-trained model, are not robust enough [60]. Therefore, alternative strategies are opted—multiple models and even multiple layers are utilized. The most discriminant features from all the three re-trained (transfer learning) models are selected by exploiting three fundamental output layers, fc1000 and predictions. During the training phase, transferred weights are kept frozen on their initial values to extract off-the-shelf deep features. A complete information regarding the selected deep layers, along with their notations, is provided in Table 3. The fully connected layers of Densenet-201, Inception-Resnet-V2, and Inception-V3 are selected as $FV0$, $FV1$, and $FV2$ respectively.

Fusion mechanism

Rather than utilizing independent features from the selected pre-trained models, we adopted a feature fusion strategy. Feature sets originating from different re-trained models are consolidated to generate a fused feature set to retain most discriminant features. Our objective here is to explore the classifier's behavior upon fusing multiple ConvNet fetures. A rudimentary strategy of feature fusion is opted by serially concatenating them to construct a resultant feature vector, which takes advantage of all feature spaces. Let us consider a joint vector $FV \in \mathbb{R}^{1 \times 3} = \{FV_k^i\}$, where $i \in \{1, 2, 3\}$ —representing selected pre-trained architecture, and $k \in \{1, 2, 3\}$ be a selected layer.

The fused feature vector $FV^\kappa = FV_k^i || FV_l^j$, exhibits set of two or three pre-trained models, having $\kappa = \{1, \dots, 4\}$ combinations. Its not imperative for the systems that adopt feature fusion strategy to perform better than those which are using single layer. Fusion strategy increases features redundancy, which makes the classifier behave inefficiently. Therefore, an addition of feature selection and dimensionality reduction steps not only decrease the redundancy but also computation time—leads to an improved classification accuracy. On contrary, overall classification accuracy increases.

Table 3 Fully connected layers of different pre-trained models and their notations

| Pre-trained model | FC-layer | Feature-vector notation |
|---------------------|-------------|-------------------------|
| Densenet201 | fc1000 | $FV0$ |
| Inception-Resnet V2 | Predictions | $FV1$ |
| Inception V3 | Predictions | $FV2$ |

Entropy-controlled NCA

Our proposed strategy revolves around the core concept—achieve best classification accuracy by exploiting minimum number of features. In this regard, a hierarchical framework is implemented, which consolidates both feature selection and dimensionality reduction—so as to avoid the problem of *curse of dimensionality*.

Feature selection

The resultant fused vector FV^k , may include redundant or irrelevant features which are formally passed through an attribute or variable selection procedure. This complete process of selecting a subset of most discriminant variables is termed as feature selection [60]. In the proposed work, the concept of entropy [61] is utilized, which has a capacity to analyze uncertain data and unveil the signal's randomness by exhibiting the system's disorder.

Let $FV^k = \{(x_1, t_1), \dots, (x_k, t_k), \dots, (x_N, t_N)\}$ be a set of training matrix containing N labels, where $X \in \{x_j\}_{j=1}^N \in \mathbb{R}^v$ is a v -dimension feature vector, and $T = \{t_j\}_{j=1}^N$ are the class labels with $t_j \in [0, 1]$ to be a binary class. This feature space has ϕ measure with the probability $\phi(X) = 1$, then the entropy is calculated as:

$$\mathbb{E}(X) = - \sum_{j=1}^N (x_j) \log \phi(x_j) \quad (14)$$

where $\phi(x_j)$ is an observation probability for a particular features $x_i \in X$. The basic purpose of applying entropy is to identify a set of unique features having natural variability, whilst entropy value tends towards 0 with minimum feature variability. The concept of entropy has been adopted in one of the recent works [18], where the authors proposed to apply entropy on a distance matrix generated from feature space—yielding restricted OA. On the other hand, in the proposed approach, we assign ranks to the features, $FV^{\mathbb{E}}$, having $(R < N)$ dimensions. The top 80% features with maximum entropy value are included to generate the resultant set. This rank based selection criteria at this stage only down-samples the original feature space, while keeping the original information conserved for the next level, dimensionality reduction.

Dimensionality reduction

Classifiers behave ineptly when there exists too many variables or these variables are highly correlated. At this stage, dimensionality reduction techniques play their vital role by reducing the number of random variables and retain the resultant vectors in the lower dimensions, $FV^{\mathbb{S}}$, where $(S \ll R)$. For this application, we are implementing NCA as a dimensionality reduction technique, on contrary, it is mostly used as a feature selection method. NCA, originally introduced by Goldberger et al. [62], is a distance metric learning algorithm which selects the projection in the projected space by optimizing the performance of nearest neighbor classifier. NCA learns projections from both features and their associated labels that will be cogent enough at partitioning classes in the projected space. For the function, NCA optimizes the criterion related to leave-one-out (LOO) accuracy of a stochastic NN classifier in the projection of space induced

by the training set. Selected entropy-controlled fused training vector, $FV^{\mathbb{E}}$, consists of $\{(x_1, t_1), \dots, (x_R, t_R)\}$, where $\{x_j, y_j\} \in \mathbb{R}^m$. NCA learns a projection matrix $\mathbb{Q} \in \mathbb{R}^{s \times m}$, representing transformation that projects x_j into s dimensional space, $\varpi_j = \mathbb{Q}x_j \in \mathbb{R}^s$, and $s \leq m$. The projection matrix \mathbb{Q} construe a Mahalanobis distance metric, calculated between two samples x_j and x_k in the projected space.

$$\mathfrak{D}(x_j, x_k) = (\mathbb{Q}x_j - \mathbb{Q}x_k)^T (\mathbb{Q}x_j - \mathbb{Q}x_k) \quad (15)$$

The primary objective of this method is to learn a projection \mathbb{Q} that maximizes the separation of a labeled data by construing the cost function, in the transformed space, based on soft-neighbor assignments. Stating a rationale that every sample x_j keeps the neighboring sample x_k as a reference with some associated probability, p_{jk} .

$$p_{jk} = \begin{cases} \frac{\Upsilon(-\mathfrak{D}(x_j, x_k))}{\sum_{j \neq k} \Upsilon(-\mathfrak{D}(x_j, x_k))} & \text{if } j \neq k, \\ 0 & \text{otherwise.} \end{cases} \quad (16)$$

where $\Upsilon(\psi) = \exp(-\psi/\zeta)$ represents a kernel function having kernel width ζ to an input argument that has a clear influence on the data samples probability—this additional step makes the model more robust and influential. Under the power of stochastic selection rule, the optimization criterion comfortably be defined by utilizing soft-neighbor assignments. The probability that the quantity x_j will be assigned a correct class label.

$$p_j = \sum_{k \in C_j} \omega_{jk} p_{jk} \quad (17)$$

$$\omega_{jk} = \begin{cases} 1 & \text{iff } \omega_j = \omega_k, \\ 0 & \text{otherwise.} \end{cases} \quad (18)$$

The optimization criterion searches to maximize the correct labels under leave-one-out policy:

$$\mathfrak{E}(\mathbb{Q}) = \sum_j \sum_k \omega_j p_{ij} = \sum_j p_j \quad (19)$$

To perform a featured reduction, as well to avoid the problem of overfitting, a regularization term $\hbar > 0$ is introduced as a standard weight in the cost function which can be tuned via cross validation [63], given as:

$$\mathfrak{E}(\mathbb{Q}) = \sum_j \sum_k \omega_j p_{ij} - \hbar \sum_{k=1}^d q_k^2 \quad (20)$$

This complete criterion gives rise to a gradient rule, used to maximize the projection matrix \mathbb{Q} and solve by differentiating $\mathfrak{E}(\mathbb{Q})$ with respect to q_k as follow:

$$\frac{\partial \mathfrak{E}(\mathbb{Q})}{\partial q(k)} = 2q_k \left(\frac{1}{\tau} \sum_j \left(p_i \sum_{j \neq k} p_{jk} |x_{ik} - x_{jk}| \right) - \sum_j \omega_{jk} p_{jk} |x_{ik} - x_{jk}| - \hbar \right) \quad (21)$$

To maximize the objective function, several gradient optimizers can be employed. However, in this article, we employed conjugate gradient method. Algorithm 1 explains the proposed approach from feature extraction (after transfer learning) to final classification.

| Algorithm 1: Proposed Algorithm | |
|--|--|
| 1 | Input: Set training data $\delta_{tr} = \{(x_1, t_1), (x_2, t_2), \dots, (x_N, y_N)\} \subseteq \delta_D \in \mathbb{R}^V$ |
| | Parameters $FV \leftarrow$ feature vector, $\Gamma_{hr} \leftarrow$, high ranked features |
| | $FV^E \leftarrow$ entropy-controlled features, $\mathbb{Q} \leftarrow$ projection matrix, $\mathbb{D} \leftarrow$ distance matrix, |
| | $p_{jk} \leftarrow$ mutual probability, $FV_{NCA} \leftarrow$ dimensionality reduced feature vector, |
| | $N \leftarrow$ total number of samples |
| 2 | Extract: $FV_i, i \in \{1, 2, 3\}$, compute FV^κ , where $\kappa \in \{1, \dots, 4\}$ |
| 3 | for $i = \{1 \text{ to } L\}$ do |
| 4 | Compute entropy $\mathbb{E}(X)$ |
| 5 | Identify high ranked feature vectors Γ_{HR} using Eqn. 13 to construct a new subspace $FV^E \in \mathbb{R}^{1 \times R}$, where $(R < N)$ |
| 6 | for $i = \{1, \dots, R\}$ do |
| 7 | Initialize projection matrix $\mathbb{Q}_{\varphi=0} = [1]^{m \times 1}$ |
| 8 | Compute mahalanobis distance matrix, $\mathbb{D}(x_j, x_k)$ using Eqn.14 |
| 9 | Compute assorted & labeling probability, P_{jk}, P_k using Eqn. 15 & 16 |
| 10 | for $k = \{1, \dots, d\}$ do |
| 11 | Compute the cost function using Eqn. 20 & update projection matrix \mathbb{Q} using conjugate gradient optimizer |
| 12 | $\varphi = \varphi + 1$ |
| 13 | Repeat step 9 & 10 k times until convergence to find \mathbb{Q} |
| 14 | Select subspace features FV_{NCA} based on projection matrix \mathbb{Q} |
| 15 | Train: selected classifier, ES-KNN, F-KNN |
| 16 | Output: Class Labels |

Results and discussion

Simulations are performed on four publicly available datasets, Table 2. Three families of state-of-the-art classifiers are utilized for classification including KNN, SVM, and Ensemble (ES). The evaluation of the proposed framework is carried out using three simulation setups: in the first, the classification results are obtained from a few selected individual layers of the pre-trained models. The second simulation setup incorporates two cases: while in the first, we simply fuse the selected layers; in the second, we combine NCA technique with the proposed feature reduction approach. We have also tested the proposed technique with other state-of-the-art classifiers. All the base parameters for the selected classifiers are given in Table 4. Additionally, a fair comparison with recent methods is also provided with remarks on the effectiveness of the proposed technique, in comparison to the state-of-the-art approaches.

Evaluation of the single layer features

Figure 7 presents classification results of each of the different layers used on the four datasets discussed in "Dataset" section. It has been observed that the models that were pre-trained by CNN architectures are powerful features representatives. From the selected pre-trained models, it has been observed that DenseNet-201 and Inception-ResNet-V2 show almost similar performance on all datasets. For example, in ISIC-UDA dataset, OA of FV0 is found to be 80.5%, whereas, OA of FV1 is 81.6%. It has also been observed that Inception-V3 shows decline in performance; hence, it is not a suitable candidate for skin cancer detection.

Table 4 Selected classifiers and their base parameters

| Classifier | Base parameters |
|---------------------|---|
| Fine tree | Maximum splits: 100 split criterion: Gini's Diversity Index |
| Medium tree | Maximum splits: 20 split criterion: Gini's Diversity Index |
| Coarse tree | Maximum splits: 4 split criterion: Gini's Diversity Index |
| Linear SVM | Kernel function: linear multi-class method: one-vs-one |
| Q-SVM | Kernel function: quadratic multi-class method: one-vs-one |
| Cubic SVM | Kernel function: cubic multi-class method: one-vs-one |
| Fine KNN | Number of neighbors: 1 distance metric: Euclidean distance weight: equal |
| Medium KNN | Number of neighbors: 10 distance metric: Euclidean distance weight: equal |
| W-KNN | Number of neighbors: 10 distance metric: Euclidean distance weight: squared inverse |
| Ensemble-BT | Ensemble method: AdaBoost learner type: decision tree maximum splits: 20 number of learners: 30 |
| Ensemble subset KNN | Ensemble method: subspace learner type: nearest neighbor number of learners: 30 |
| Ensemble RUSB | Ensemble method: RUSBoost learner type: decision tree number of learners: 30 maximum splits: 20 |

Evaluation of the proposed technique

Prior to the feature selection and dimensionality reduction step, the extracted features from various architectural layers are concatenated. Table 5 shows reduction percentage of fused feature vectors achieved after applying a hierarchical framework of entropy and NCA, before the classification phase. It is evident from the figures that maximum reduction percentage achieved is 98.50% on PH^2 dataset, whilst, average reduction on all dataset is 95.17%. We create four combinational feature vectors from each dataset.

Table 6 presents a comparison of classification results, in terms of OA, for two different cases: (1) simple fusion approach, (2) entropy-controlled NCA (proposed). The two cases are implemented on fused feature vector, and on four different datasets, using the selected classifiers. Discussion for the two cases are given below:

- Case 1: on PH^2 dataset, the best classification accuracy achieved is 83.2% using Fine KNN (F-KNN), 82.2% using SVM and 82.4% using ES-KNN classifier, when FV0–FV1–FV2 are fused. Similarly, on ISIC-MSK dataset, by using the same fusion, F-KNN outperforms SVM and ES-KNN by achieving 76.4%. In case of ISIC-UDA, F-KNN yields 76.5% classification accuracy, which is greater than SVM (73.5%) and ES-KNN (76.0%). On ISBI-2017 dataset ES-KNN gives 76.1% accuracy, which is greater than both SVM and F-KNN. It has been observed, and hence concluded, that irrespective of the given dataset, the best classification results are obtained with the fusion of FV0–FV1–FV2, thereby validating the strength of the feature fusion approach.
- Case 2: using entropy-controlled feature fusion approach, on PH^2 , ISIC-MSK, and ISIC-UDA datasets, F-KNN yields the best accuracy of 98.8%, 99.2%, and 97.1% respectively, courtesy the feature fusion approach. In case of ISBI-2017 dataset, however, ES-KNN gives maximum accuracy of 95.9%. Note that the number of image samples in ISBI-2017 is larger as compared to other datasets; it may be concluded that ES-KNN gives classification results better as compared to other classifiers for datasets having greater number of samples.

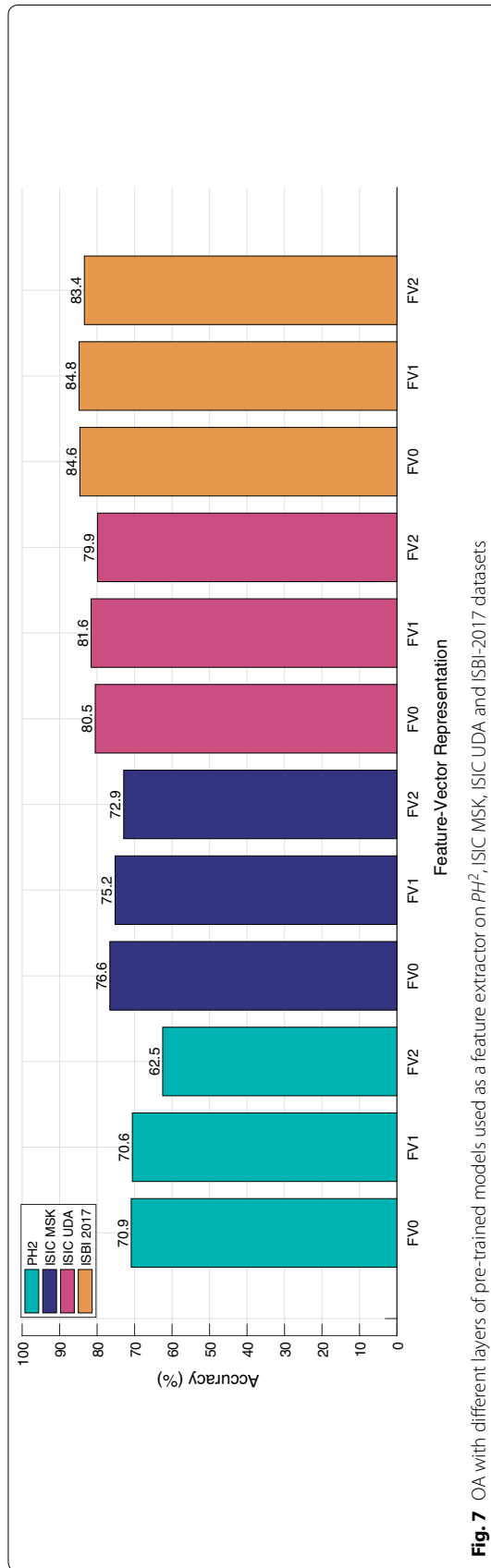


Fig. 7 OA with different layers of pre-trained models used as a feature extractor on PH², ISIC MSK, ISIC UDA and ISBI-2017 datasets

Table 5 Features fusion and reduction percentage

| Vector fusion | Input dimension | Output dimension | Percentage reduction (%) |
|-----------------------|-----------------|------------------|--------------------------|
| <i>PH²</i> | | | |
| FV0–FV1 | 160 × 2000 | 160 × 50 | 97.50 |
| FV0–FV2 | 160 × 2000 | 160 × 53 | 97.35 |
| FV1–FV2 | 160 × 2000 | 160 × 55 | 97.25 |
| FV0–FV1–FV2 | 160 × 3000 | 160 × 45 | 98.50* |
| ISIC-MSK | | | |
| FV0–FV1 | 180 × 2000 | 180 × 99 | 95.05 |
| FV0–FV2 | 180 × 2000 | 180 × 125 | 93.75 |
| FV1–FV2 | 180 × 2000 | 180 × 167 | 91.65 |
| FV0–FV1–FV2 | 180 × 3000 | 180 × 95 | 96.83 |
| ISIC-UDA | | | |
| FV0–FV1 | 446 × 2000 | 446 × 227 | 88.65 |
| FV0–FV2 | 446 × 2000 | 446 × 197 | 90.15 |
| FV1–FV2 | 446 × 2000 | 446 × 161 | 91.95 |
| FV0–FV1–FV2 | 446 × 3000 | 446 × 104 | 96.53 |
| ISBI-2017 | | | |
| FV0–FV1 | 2200 × 2000 | 2200 × 107 | 95.13 |
| FV0–FV2 | 2200 × 2000 | 2200 × 65 | 97.05 |
| FV1–FV2 | 2200 × 2000 | 2200 × 53 | 97.59 |
| FV0–FV1–FV2 | 2200 × 3000 | 2200 × 49 | 98.37 |

* Shows the highest value among all datasets

In Table 7, the average classification time and average accuracy of all datasets are shown. From this table it is evident that the proposed technique outperforms simple fusion approach with substantial time margin and with maximum classification accuracy. Additionally, a confidence interval is plotted in Fig. 8 against all selected datasets and using two different classifiers (F-KNN, ES-KNN), which works best compared to others. Moreover, to provide a better insight and to facilitate researchers working in this domain, a comprehensive comparison of set of classifiers is also provided, Table 8. From the stats, its quite clear that the classifiers belong to the family of KNN performs best both in terms of average classification accuracy (94.73%) and average computational time (1.30 s). The second best family in this domain is SVM—showing average classification accuracy of 93.83% and average computational time of 1.96 s. Ensemble and Tree family is not showing improved results in terms of average classification accuracy (89.87%, 84.91%), whilst, average computational time of ensemble family is 6.05 sec, but tree family is time efficient by taking only 1.57 s. Same trend is being followed in calculating AUC.

Comparison with state of the art techniques

A comprehensive comparison with existing techniques utilizing *PH²*, ISBI-2017 and ISIC-MSK datasets is given in Table 9. It can be clearly observed that our proposed methodology achieves best classification accuracy on all the given datasets. The maximum classification accuracy achieved by the previous works on *PH²* dataset is 96.00% using color and texture features, while using the proposed methodology, it is 98.80%.

Table 6 Classification results of the proposed technique compared to simple feature fusion with four datasets using F-KNN, ES-KNN and SVM

| Vector Fusion | OA (%) | | | | | |
|-----------------|-------------------------|------|--------|------------------|------|--------|
| | Feature Fusion Approach | | | Proposed (ECNCA) | | |
| | F-KNN | SVM | ES-KNN | F-KNN | SVM | ES-KNN |
| PH ² | | | | | | |
| FV0-FV1 | 82.8 | 80.0 | 80.1 | 96.9 | 93.7 | 95.1 |
| FV0-FV2 | 82.1 | 81.7 | 81.7 | 95.1 | 94.8 | 93.2 |
| FV1-FV2 | 82.9 | 82.0 | 82.1 | 97.4 | 95.0 | 97.1 |
| FV0-FV1-FV2 | 83.2 | 82.2 | 82.4 | 98.8* | 95.1 | 98.1 |
| ISIC-MSK | | | | | | |
| FV0-FV1 | 74.2 | 71.7 | 74.6 | 93.7 | 87.2 | 88.8 |
| FV0-FV2 | 73.9 | 71.0 | 73.0 | 89.1 | 89.0 | 87.4 |
| FV1-FV2 | 73.1 | 72.5 | 75.1 | 86.5 | 91.0 | 89.7 |
| FV0-FV1-FV2 | 76.4 | 74.8 | 74.9 | 99.2* | 95.1 | 96.9 |
| ISIC-UDA | | | | | | |
| FV0-FV1 | 71.9 | 70.0 | 75.9 | 88.8 | 80.1 | 84.7 |
| FV0-FV2 | 73.3 | 71.2 | 74.1 | 90.7 | 84.5 | 88.0 |
| FV1-FV2 | 74.1 | 75.9 | 75.8 | 92.8 | 82.7 | 94.2 |
| FV0-FV1-FV2 | 76.5 | 73.5 | 76.0 | 97.1* | 93.3 | 95.7 |
| ISBI-2017 | | | | | | |
| FV0-FV1 | 73.2 | 70.9 | 71.1 | 88.5 | 88.0 | 88.8 |
| FV0-FV2 | 74.7 | 70.7 | 72.8 | 89.7 | 87.3 | 89.3 |
| FV1-FV2 | 72.1 | 70.5 | 73.3 | 90.0 | 88.9 | 90.7 |
| FV0-FV1-FV2 | 75.3 | 75.1 | 76.1 | 94.1 | 93.4 | 95.9* |

* Shows the highest value in each dataset

Table 7 Average classification time and accuracy on all datasets

| Dataset | Average classification time (s) | | Average accuracy | |
|-----------------|---------------------------------|---------------|--------------------|---------------|
| | Entropy-controlled | Simple fusion | Entropy-controlled | Simple fusion |
| PH ² | 0.60 | 36 | 98.80 | 83.20 |
| ISIC-MSK | 0.73 | 43 | 99.20 | 74.40 |
| ISIC-UDA | 1.62 | 96 | 97.10 | 76.50 |
| ISBI-2017 | 7.59 | 455 | 95.90 | 76.10 |

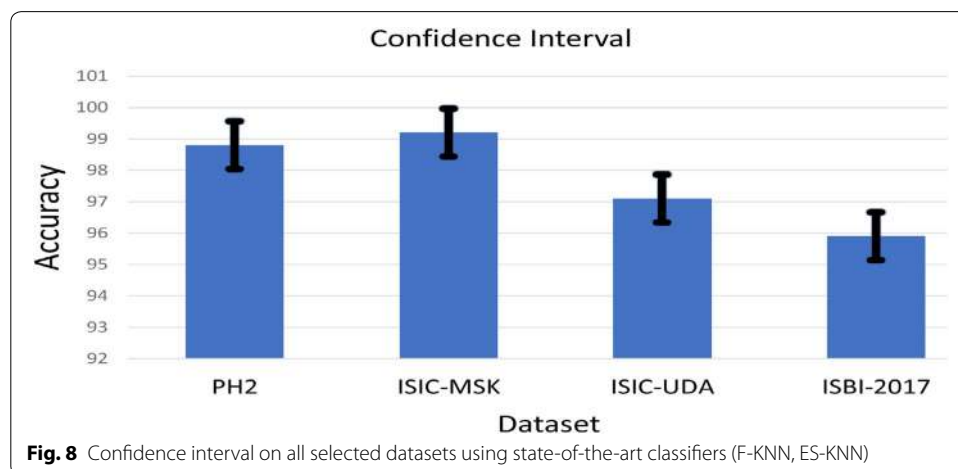


Fig. 8 Confidence interval on all selected datasets using state-of-the-art classifiers (F-KNN, ES-KNN)

Table 8 Performance comparison of state-of-the-art classifiers on selected datasets using set of performance measures

| Classifier | Dataset | | | | Performance Measures | | | | | | |
|----------------|---------|----|-----|----|----------------------|--------|-----------|------|------|------|------------|
| | I | II | III | IV | OA (%) | Recall | Precision | FNR | FPR | AUC | Time (sec) |
| Fine Tree | ✓ | | | | 87.5 | 83.0 | 80.5 | 12.5 | 0.17 | 0.84 | 0.76 |
| | | ✓ | | | 79.1 | 78.5 | 79.0 | 20.9 | 0.22 | 0.73 | 0.93 |
| | | | ✓ | | 82.6 | 72.9 | 72.9 | 17.4 | 0.34 | 0.69 | 1.87 |
| | | | | ✓ | 89.3 | 77.9 | 78.9 | 10.2 | 0.20 | 0.73 | 5.04 |
| Medium Tree | ✓ | | | | 87.5 | 83.0 | 80.5 | 12.5 | 0.17 | 0.84 | 0.58 |
| | | ✓ | | | 79.1 | 78.5 | 79.0 | 20.9 | 0.22 | 0.73 | 0.76 |
| | | | ✓ | | 84.2 | 74.9 | 74.4 | 15.8 | 0.32 | 0.72 | 1.70 |
| Coarse Tree | | | | ✓ | 87.8 | 69.4 | 75.4 | 11.7 | 0.29 | 0.73 | 2.16 |
| | ✓ | | | | 86.9 | 78.0 | 80.0 | 13.1 | 0.22 | 0.77 | 0.59 |
| | | ✓ | | | 80.4 | 79.0 | 80.5 | 19.6 | 0.22 | 0.74 | 0.73 |
| Linear SVM | | | | ✓ | 87.1 | 70.9 | 71.4 | 12.9 | 0.36 | 0.67 | 1.60 |
| | ✓ | | | | 87.5 | 68.4 | 74.4 | 12 | 0.30 | 0.67 | 2.06 |
| | | ✓ | | | 93.1 | 87.5 | 90.5 | 6.9 | 0.12 | 0.97 | 0.60 |
| Quadratic SVM | | | | ✓ | 88.7 | 88.0 | 88.5 | 11.3 | 0.13 | 0.91 | 0.78 |
| | ✓ | | | | 95.2 | 80.9 | 90.9 | 4.8 | 0.26 | 0.91 | 1.78 |
| | | ✓ | | | 91.3 | 72.9 | 84.9 | 8.2 | 0.25 | 0.84 | 4.3 |
| Cubic SVM | | | | ✓ | 95.6 | 93.0 | 93.5 | 4.4 | 0.07 | 0.99 | 0.61 |
| | ✓ | | | | 90.9 | 90.5 | 91.0 | 9.1 | 0.10 | 0.93 | 0.78 |
| | | ✓ | | | 96.2 | 87.9 | 91.9 | 3.8 | 0.19 | 0.94 | 1.71 |
| F-KNN | | | | ✓ | 94.0 | 82.4 | 88.4 | 5.5 | 0.20 | 0.88 | 4.5 |
| | ✓ | | | | 96.9 | 93.5 | 97.0 | 3.1 | 0.07 | 1.00 | 0.61 |
| | | ✓ | | | 91.8 | 91.5 | 91.5 | 8.2 | 0.09 | 0.96 | 0.78 |
| Medium KNN | | | | ✓ | 96.6 | 89.4 | 92.9 | 3.4 | 0.18 | 0.95 | 1.69 |
| | ✓ | | | | 95.6 | 87.4 | 89.4 | 3.9 | 0.21 | 0.90 | 5.42 |
| | | ✓ | | | 98.8* | 97.0 | 99.0 | 1.2 | 0.03 | 0.97 | 0.60 |
| Weighted KNN | | | | ✓ | 99.2* | 99.0 | 99.0 | 0.8 | 0.02 | 0.94 | 0.73 |
| | ✓ | | | | 97.1* | 93.9 | 93.4 | 2.9 | 0.13 | 0.88 | 1.62 |
| | | ✓ | | | 95.8 | 92.9 | 92.4 | 4.6 | 0.19 | 0.86 | 2.53 |
| Ensemble BT | | | | ✓ | 92.5 | 81.5 | 95.5 | 7.5 | 0.19 | 1.00 | 0.63 |
| | ✓ | | | | 91.8 | 91.0 | 91.5 | 8.2 | 0.10 | 0.96 | 0.72 |
| | | ✓ | | | 95.5 | 78.9 | 91.9 | 4.5 | 0.28 | 0.90 | 1.50 |
| Ensemble S-KNN | | | | ✓ | 89.3 | 62.4 | 91.4 | 10.2 | 0.35 | 0.86 | 2.15 |
| | ✓ | | | | 93.1 | 83.0 | 96.0 | 6.9 | 0.17 | 1.00 | 0.62 |
| | | ✓ | | | 94.4 | 93.5 | 95.0 | 5.6 | 0.07 | 0.98 | 0.72 |
| Ensemble BT | | | | ✓ | 95.2 | 80.9 | 90.9 | 4.8 | 0.26 | 0.92 | 1.60 |
| | ✓ | | | | 94.1 | 75.9 | 87.9 | 5.4 | 0.22 | 0.91 | 2.12 |
| | | ✓ | | | 80.0 | 50.0 | 40.0 | 20 | 0.20 | 0.90 | 0.78 |
| Ensemble S-KNN | | | | ✓ | 82.6 | 82.0 | 82.5 | 17.4 | 0.19 | 0.83 | 3.47 |
| | ✓ | | | | 93.2 | 81.4 | 86.4 | 6.8 | 0.26 | 0.85 | 7.87 |
| | | ✓ | | | 92.3 | 73.4 | 88.9 | 7.2 | 0.25 | 0.87 | 13.48 |
| Ensemble S-KNN | | | | ✓ | 92.3 | 73.4 | 88.9 | 7.2 | 0.25 | 0.87 | 13.48 |
| | ✓ | | | | 98.1 | 95.5 | 99.0 | 1.9 | 0.04 | 1.00 | 4.06 |
| | | ✓ | | | 96.2 | 96.0 | 96.0 | 3.8 | 0.05 | 0.99 | 3.49 |
| Ensemble S-KNN | | | | ✓ | 96.8 | 89.9 | 91.4 | 3.2 | 0.17 | 0.92 | 5.36 |
| | | | | ✓ | 95.9* | 93.4 | 95.4 | 3.6 | 0.17 | 0.92 | 7.59 |

Table 8 (continued)

| Classifier | Dataset | | | | Performance Measures | | | | | | |
|---------------|---------|----|-----|----|----------------------|--------|-----------|------|------|------|------------|
| | I | II | III | IV | OA (%) | Recall | Precision | FNR | FPR | AUC | Time (sec) |
| Ensemble RUSB | ✓ | | | | 88.8 | 86.0 | 81.5 | 11.2 | 0.14 | 0.93 | 4.74 |
| | | ✓ | | | 85.7 | 85.0 | 85.5 | 14.3 | 0.16 | 0.88 | 5.24 |
| | | | ✓ | | 85.5 | 87.9 | 83.4 | 14.5 | 0.21 | 0.88 | 7.09 |
| | | | | ✓ | 83.3 | 82.4 | 74.9 | 16.2 | 0.20 | 0.85 | 9.45 |

* Shows the highest value in each dataset

Table 9 Comparison with state-of-the-art methods

| Ref | Year | Dataset | Method | OA (%) |
|----------|------|-----------|---|--------|
| [65] | 2016 | PH^2 | ABCD rule | 90.00 |
| [66] | 2016 | PH^2 | wavelet transform with morphological operations | 93.87 |
| [15] | 2017 | PH^2 | multistage fully convolutional network | 94.24 |
| [67] | 2017 | PH^2 | color and texture features | 96.00 |
| [68] | 2018 | ISBI-2017 | regularised discriminant learning | 83.20 |
| [13] | 2018 | ISBI-2017 | fully convolutional residual networks & lesion index calculation unit | 85.70 |
| [69] | 2018 | ISBI-2017 | Ensemble Of Deep Neural Networks | 84.8% |
| [18] | 2018 | ISIC-MSK | probabilistic distribution and best features selection | 97.20 |
| Proposed | 2019 | ISBI-2017 | ECNCA | 95.90 |
| Proposed | 2019 | ISIC-UDA | ECNCA | 97.10 |
| Proposed | 2019 | ISIC-MSK | ECNCA | 99.20 |
| Proposed | 2019 | PH^2 | ECNCA | 98.80 |

Similarly on ISBI-2017 dataset, the maximum accuracy achieved by the proposed methodology is 95.90%, compared to other methods, e.g. [64] achieving 94.08% on the same dataset. Similarly on ISIC-MSK, the accuracy achieved by [18] is 97.20%, while the proposed methodology gives 99.20%.

Conclusion

Considering the recent success of deep architectures, we presented an effective approach for the classification of skin lesion. Comparing with conventional techniques, we introduced a hierarchical framework of discriminant features selection followed by a dimensionality reduction step. We exploited extracted information from the selected pre-trained models after fine tuning, which contributed significantly in the improvement of classification accuracy. With the proposed method, we utilized less than 3% of total features, which not only improves the classification accuracy by removing redundancy but also minimizes the computational time. After implementing this idea, we are in a position to put forth a few claims including: (a) fusion of extracted features from set of pre-trained models improves the overall accuracy, (b) an addition of feature selection and dimensionality reduction step significantly improve the classification results. As a future work, an improved segmentation criteria will be our primary focus along with the

extended feature selection criteria. Moreover, we will include a few more and challenging datasets in order to provide a comprehensive comparison.

Abbreviations

ECNCA: Entropy-controlled neighborhood component analysis; ISIC: International Skin Imaging Collaboration; ISBI: International Symposium on Biomedical Imaging; WHO: World Health Organization; CNMD: Consensus Net Meeting on Dermoscopy; CAD: Computer-aided diagnosis; CNN: Convolutional neural networks; TL: Transfer learning; OA: Overall accuracy; NCA: Neighborhood component analysis; SVM: Support vector machines; KNN: K Nearest Neighbors; LBP: Local binary patterns; ROI: Region of interest; MLR: Multi-scale lesion biased representation; FFT: Fast Fourier transform; BVLC: Berkeley Vision and Learning Center; LOO: Leave-one-out; ES: Ensemble; RUSB: Random UnderSampling Boost.

Acknowledgements

Research is funded by Deanship of Scientific Research at University of Ha'il.

Authors' contributions

Conceptualization: TA. Data curation: SRN, NNQ. Funding acquisition: MALhaisoni. Investigation: HMJL. Methodology: SRN, TA. Project administration: NNQ. Resources: SAH, MALi. Software: SN, MALi. Supervision: TA, SRN. Validation: TA, MALhaisoni. Visualization: HMJL, SAH. Writing original draft: SN, SRN, TA. All authors read and approved the final manuscript.

Funding

Research is funded by Deanship of Scientific Research at University of Ha'il.

Availability of data and materials

Four datasets are utilized in this research including PH2 (<https://www.fc.up.pt/addi/ph2%20database.html>), ISBI-2017 (<https://biomedicalimaging.org/2017/challenges/>), ISIC-UDA (<https://www.isic-archive.com/#!/topWithHeader/onlyHeaderTop/gallery>), and ISIC-MSK (<https://www.isic-archive.com/#!/topWithHeader/onlyHeaderTop/gallery>). These datasets are publicly available.

Competing interests

Authors declare that they have no competing interests.

Author details

¹ Department of Electrical and Computer Engineering, COMSATS University Islamabad, Wah Campus, Wah Cantonment, Pakistan. ² College of Computer Science and Engineering, University of Ha'il, Ha'il, Saudi Arabia.

Received: 15 April 2019 Accepted: 5 March 2020

Published online: 31 March 2020

References

1. Skin cancer facts, 2017. URL <https://seer.cancer.gov/statfacts/html/melan.html>
2. Barata C, Ruela M, Francisco M, Mendonca T, Marques J (2014) Two systems for the detection of melanomas in dermoscopy images using texture and color features. *Syst J* 8:965–979
3. Hoshyar AN, Al-Jumaily A (2014) The beneficial techniques in preprocessing step of skin cancer detection system comparing. *Procedia Comput Sci* 42:25–31
4. Nachbar F, Stolz W, Merkle T, Cognetta AB, Vogt T, Landthaler M, Bilek P, Braunfalco O, Plewig G (1994) The ABCD rule of dermatoscopy. *J Am Acad Dermatol* 4:521–527
5. Delfino M, Argenziano G, Fabbrocini G, Carli P, Giorgi VD, Sammarco E (1998) Epiluminescence microscopy for the diagnosis of doubtful melanocytic skin lesions. Comparison of the ABCD rule. *Arch Dermatol* 134:1563–1570
6. Menzies SW, Ingvar C, Crotty KA, McCarthy WH (1996) Frequency and morphologic characteristics of invasive melanomas lacking specific surface microscopic features. *Arch Dermatol* 132:1178–1182
7. Argenziano G, Soyer HP, Chimenti S, Talamini R, Corona R, Binder M, Sera F, Cerroni L, De Rosa G, Ferrara G (2003) Dermoscopy of pigmented skin lesions: results of a consensus meeting via the internet. *J Am Acad Dermatol* 48:679–693
8. Ruela M, Barata C, Mendonca T, Marques J (2013) On the role of shape in the detection of melanomas. In: 8th international symposium on image and signal processing and analysis (ISPA 2013)
9. Nida N, Irtaza A, Javed A, Yousaf MH, Mahmood MT (2019) Melanoma lesion detection and segmentation using deep region based convolutional neural network and fuzzy C-means clustering. *Int J Med Inform* 124:37–48
10. Fernando B, Fromont E, Tuytelars T (2014) Mining mid-level features for image classification. *J Comput Vis* 108(3):186–203
11. Akram T, Khan MA, Sharif M, Yasmin M (2018) Skin lesion segmentation and recognition using multichannel saliency estimation and M-SVM on selected serially fused features. *J Ambient Intell Humaniz Comput*. <https://doi.org/10.1007/s12652-018-1051-5>
12. Khatami A, Nazari A, Khosravi A, Lim CP, Nahavandi S (2020) A weight perturbation-based regularisation technique for convolutional neural networks and the application in medical imaging. *Expert Syst Appl* 149:113196
13. Li Y, Shen L (2018) Skin lesion analysis towards melanoma detection using deep learning network. *Sensors* 18(2):556
14. Dolz J, Desrosiers C, Wang L, Yuan J, Shen D, Ayed IB (2020) Deep CNN ensembles and suggestive annotations for infant brain MRI segmentation. *Comput Med Imaging Gr* 79:101660

15. Bi L, Kim J, Ahn E, Kumar A, Fulham M, Feng D (2017) Dermoscopic image segmentation via multi-stage fully convolutional networks. *IEEE Trans Biomed Eng* 64(9):2065–2074
16. Marques JS, Barata C, Mendonca T (2012) On the role of texture and color in the classification of dermoscopy images. In: Annual international conference of the IEEE engineering in medicine and biology society (EMBC)
17. Ganster H, Pinz A, Rohrer R, Wildling E, Blinder M, Kittler H (2001) Automated melanoma recognition. *IEEE Trans Biom Eng* 20(3):233–239
18. Khan MA, Tallha A, Muhammad S, Aamir S, Khurshheed A, Musaed A, Syed IH, Abdualziz A (2018) An implementation of normal distribution based segmentation and entropy-controlled features selection for skin lesion detection and classification. *BMC Cancer* 18(1):638
19. Naeem S, Riaz F, Hassan A, Miguel Tavares C, Nisar R (2015) Description of visual content in dermoscopy images using joint histogram of multiresolution local binary patterns and local contrast. In: Proceedings of 16th international conference on intelligent data engineering and automated learning (IDEAL 2015), Poland
20. Khan MA, Sharif M, Akram T, Bukhari SA, Nayak RS (2020) Developed newton-raphson based deep features selection framework for skin lesion recognition. *Pattern Recognit Lett* 129:293–303
21. Sumithra R, Suhil M, Guru DS (2015) Segmentation and classification of skin lesions for disease diagnosis. *Procedia Comput Sci* 45:76–85
22. Attia M, Hossny M, Zhou H, Nahavandi S, Asadi H, Yazdabadi A (2019) Digital hair segmentation using hybrid convolutional and recurrent neural networks architecture. *Comput Methods Progr Biomed* 177:17–30
23. Joseph S, Panicker JR (2016) Skin lesion analysis system for melanoma detection with an effective hair segmentation method. In: International conference on information science (ICIS). IEEE, New York, pp 91–96
24. Cheerla N, Frazier D (2014) Automatic melanoma detection using multi-stage neural networks. *Int J Innov Res Sci Eng Technol* 3(2):9164–9183
25. Khan KA, Shanir PP, Khan YU, Farooq O (2020) A hybrid local binary pattern and wavelets based approach for EEG classification for diagnosing epilepsy. *Expert Syst Appl* 140:112895
26. Hawas AR, Guo Y, Du C, Polat K, Ashour AS (2020) OCE-NGC: a neutrosophic graph cut algorithm using optimized clustering estimation algorithm for dermoscopic skin lesion segmentation. *Appl Soft Comput* 86:105931
27. Hajiaghayi M, Kortsarz G, MacDavid R, Purohit M, Sarpatwar K (2020) Approximation algorithms for connected maximum cut and related problems. *Theor Comput Sci* 814:74–85
28. Pour MP, Seker H (2020) Transform domain representation-driven convolutional neural networks for skin lesion segmentation. *Expert Syst Appl* 144:113129
29. Ahn E, Bi L, Jung YH, Kim J, Li C, Fulham M, Feng DD (2015) Automated saliency-based lesion segmentation in dermoscopic images. In: 2015 37th annual international conference of the IEEE engineering in medicine and biology society (EMBC). IEEE, New York, pp 3009–3012
30. Khan MA, Akram T, Sharif M, Javed K, Rashid M, Bukhari SAC (2019) An integrated framework of skin lesion detection and recognition through saliency method and optimal deep neural network features selection. *Neural Comput Appl*. <https://doi.org/10.1007/s00521-019-04514-0>
31. Barata C, Ruela M, Francisco M, Mendona T, Marques JS (2014) Two systems for the detection of melanomas in dermoscopy images using texture and color features. *IEEE Syst J* 8(3):965–979
32. Qaisar A, Garcia IF, Emre Celebi M, Ahmad W, Mushtaq Q (2013) A perceptually oriented method for contrast enhancement and segmentation of dermoscopy images. *Skin Res Technol* 19(1):e490–e497
33. Nagarajan G, Babu LD (2019) A hybrid of whale optimization and late acceptance hill climbing based imputation to enhance classification performance in electronic health records. *J Biomed Inform* 94:103190
34. Chatterjee S, Dey D, Munshi S, Gorai S (2019) Extraction of features from cross correlation in space and frequency domains for classification of skin lesions. *Biomed Signal Process Control* 53:101581
35. Bi L, Kim J, Ahn E, Feng D, Fulham M (2016) Automatic melanoma detection via multi-scale lesion-biased representation and joint reverse classification. In: 2016 IEEE 13th international symposium on biomedical imaging (ISBI). IEEE, New York, pp 1055–1058
36. Abuzagheh O, Faezipour M, Barkana BD (2016) A portable real-time noninvasive skin lesion analysis system to assist in melanoma early detection and prevention
37. Mahbod A, Schaefer G, Ellinger I, Ecker R, Pitiot A, Wang C (2018) Fusing fine-tuned deep features for skin lesion classification. *Comput Med Imaging Gr* 71:19
38. Al-masni MA, Kim DH, Kim TS (2020) Multiple skin lesions diagnostics via integrated deep convolutional networks for segmentation and classification. *Comput Methods Progr Biomed* 190:105351
39. Ibtehaz N, Rahman MS (2020) MultiResuNet: rethinking the U-Net architecture for multimodal biomedical image segmentation. *Neural Netw* 121:74–87
40. Hajabdollahi M, Esfandiarpour R, Sabeti E, Karimi N, Soroushmehr SM, Samavi S (2020) Multiple abnormality detection for automatic medical image diagnosis using bifurcated convolutional neural network. *Biomed Signal Process Control* 57:101792
41. Kadampur MA, Al Riyae S (2020) Skin cancer detection: applying a deep learning based model driven architecture in the cloud for classifying dermal cell images. *Inform Med Unlocked* 18:100282
42. Xie D, Lei Z, Li B (2017) Deep learning in visual computing and signal processing. *Appl Comput Intell Soft Comput*. <https://doi.org/10.1155/2017/1320780>
43. Karl W, Khoshgoftaar TM, Wang DD (2016) A survey of transfer learning. *J Big Data* 3(1):9
44. Yosinski J, Clune J, Bengio Y, Lipson H (2014) How transferable are features in deep neural networks? *Advances in neural information processing systems*. Springer, Singapore, pp 3320–3328
45. Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*. MIT Press, Cambridge, pp 1097–1105
46. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*
47. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A (2015) Going deeper with convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1–9

48. Iandola FN, Han S, Moskewicz MW, Ashraf K, Dally WJ, Keutzer K (2016) Squeezenet: alexnet-level accuracy with 50x fewer parameters and < 0.5 mb model size. arXiv preprint [arXiv:1602.07360](https://arxiv.org/abs/1602.07360)
49. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 770–778
50. Huang G, Liu Z, Weinberger KQ (2016) Densely connected convolutional networks. CoRR, abs/1608.06993, [arXiv:1608.06993](https://arxiv.org/abs/1608.06993)
51. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z (2016) Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 2818–2826
52. Szegedy C, Ioffe S, Vanhoucke V, Alemi AA (2017) Inception-v4, inception-resnet and the impact of residual connections on learning. In: AAAI, vol 4, p 12
53. Duan Y, Fang L, Licheng J, Peng Z, Zhang L (2017) SAR image segmentation based on convolutional-wavelet neural network and markov random field. *Pattern Recognit* 64:255–267
54. Tang P, Hanli W (2017) Richer feature for image classification with super and sub kernels based on deep convolutional neural network. *Comput Electr Eng* 62:499–510
55. Mendoncy T, Ferreira PM, Marques J, Marcyal ARS, Rozeira J (2013) A dermoscopic image database for research and benchmarking. Presentation in proceedings of PH2 IEEE EMBC
56. Gutman D, Codella NCF, Celebi E, Helba B, Marchetti M, Mishra N, Halpern A (2016) Skin lesion analysis toward melanoma detection: a challenge. In: The international symposium on biomedical imaging (ISBI), hosted by the international skin imaging collaboration (ISIC). arXiv preprint [arXiv:1605.01397](https://arxiv.org/abs/1605.01397). 2016
57. Codella NCF, Gutman D, Celebi ME, Helba B, Marchetti MA, Dusza SW, et al (2017) Skin lesion analysis toward melanoma detection: a challenge at the 2017 int. symp. biomed. imaging. arXiv preprint [arXiv:1710.05006](https://arxiv.org/abs/1710.05006)
58. Tim L, Vincent N, Richard G, Andrew C, David M (1997) Dullrazor: a software approach to hair removal from images. *Comput Biol Med* 27(533–43):12. [https://doi.org/10.1016/S0010-4825\(97\)00020-6](https://doi.org/10.1016/S0010-4825(97)00020-6)
59. Duan Q, Akram T, Duan P, Wang X (2016) Visual saliency detection using information contents weighting. *Optik* 127(19):7418–7430
60. Akram T, Laurent B, Naqvi SR, Alex MM, Muhammad N (2018) A deep heterogeneous feature fusion approach for automatic land-use classification. *Inf Sci* 467:199–218
61. Sankar AS, Nair SS, Dharan VS, Sankaran P (2015) Wavelet sub band entropy based feature extraction method for BCI. *Procedia Comput Sci* 46:1476–1482
62. Goldberger J, Hinton GE, Roweis ST, Salakhutdinov RR (2005) Neighbourhood components analysis. *Advances in neural information processing systems*. MIT Press, Cambridge, pp 513–520
63. Wei Y, Kuanquan W, Wangmeng Z (2012) Neighborhood component feature selection for high-dimensional data. *JCP* 7(1):161–168
64. Bi L, Kim J, Ahn E, Kumar A, Feng D, Fulham M (2019) Step-wise integration of deep class-specific learning for dermoscopic image segmentation. *Pattern Recognit* 85:78–89
65. Zaqout I (2016) Diagnosis of skin lesions based on dermoscopic images using image processing techniques. *Int J Signal Process Image Process Pattern Recognit* 9(9):189–204
66. Shehzad K, Uzma J, Kashif S, Usman Akram M, Manzoor W, Ahmed W, Sohail A (2016) Segmentation of skin lesion using Cohen-Daubechies-Feauveau biorthogonal wavelet. *SpringerPlus* 5(1):1603
67. Waheed Z, Waheed A, Zafar M, Raiz F (2017) An efficient machine learning approach for the detection of melanoma using dermoscopic images. In: International conference on communication, computing and digital systems (C-CODE). IEEE, New York, pp 316–319
68. Sultana NN, Mandal B, Puhon NB (2018) Deep residual network with regularised fisher framework for detection of melanoma. *IET Comput Vis* 12(8):1096–1104
69. Harangi B, Baran A, Hajdu A (2018) Classification of skin lesions using an ensemble of deep neural networks. In: 2018 40th annual international conference of the IEEE engineering in medicine and biology society (EMBC), pp 2575–2578

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)
