



DermoDeep-A classification of melanoma-nevus skin lesions using multi-feature fusion of visual features and deep neural network

Qaisar Abbas¹  · M. Emre Celebi²

Received: 16 August 2018 / Revised: 6 March 2019 / Accepted: 15 April 2019 /

Published online: 7 May 2019

© Springer Science+Business Media, LLC, part of Springer Nature 2019

Abstract

The Scientific community has been developing computer-aided detection systems (CADs) for automatic diagnosis of pigmented skin lesions (PSLs) for nearly 30 years. Several works have addressed this issue and obtained encouraging results, however, there has not been much focus on the pre-processing step, determining the relevance of the features considered and how they may be important indicators of a lesion's malignancy. To differentiate between nevus and melanoma skin lesions, the development of CAD system is a challenging task due to the use of inaccurate image processing techniques. In this paper, a new classification system is developed for PSLs known as DermoDeep through a fusion of multiple visual features and deep-neural-network approach. A new aggregation of visual features along with descriptors are extracted in a perceptual-oriented color space. Moreover, a new five-layer architecture based DermoDeep system is proposed. This DermoDeep system applied on 2800 region-of-interest (ROI) PSLs including 1400 nevus and 1400 malignant lesions. The classification accuracy of DermoDeep system is compared with the state-of-the-art methods and evaluated by the sensitivity (SE), specificity (SP) and area under the receiver operating characteristics (AUC) curve. The difference between AUC of DermoDeep is statistically significant compared to other techniques with AUC: 0.96 ($p < 0.001$), SE of 93% and SP of 95%. The obtained results demonstrate that the DermatDeep can be used to assist dermatologists during a screening process.

Keywords Skin cancer · Digital dermoscopy · Computer-aided diagnosis system · Perceptual-oriented color space · Multi-feature fusion · Color features · Deep learning · Deep neural networks · Recurrent neural networks

✉ Qaisar Abbas
qaisarabbasphd@gmail.com; qaabbas@imamu.edu.sa

M. Emre Celebi
ecelebi@uca.edu

¹ FR-85, College of Computer and Information Sciences, Al Imam Mohammad Ibn Saud Islamic University (IMSIU), P.O. 5701, Riyadh 11432, Saudi Arabia

² Department of Computer Science, University of Central Arkansas, Conway, AR, USA

1 Introduction

The recognition of pigmented skin lesions (PSLs) using dermoscopy images [53] has significant improvements during last decades. For early detection of skin cancer, the clinical experts like dermatologists are widely using computer-aided diagnosis (CAD) [57] to classify the type of PSLs for nearly 30 years. In practice, the CAD systems are providing second opinion [20] to less experienced dermatologists for better diagnosis of skin cancer. To help practitioners, the CAD systems are developed by extraction of visual features in the past to classify melanoma skin lesions. According to review, the CAD systems are extensively studied but they recognized only certain types of pigmented skin [34] lesions. These systems are mainly developed based on the extraction of color and texture features [64] and then classified these features through machine learning algorithms.

For the past 10 years, the CAD system [51] is considered an active and challenging research area in medicine. There are two main types of pigmented skin lesions, namely melanocytic and non-melanocytic [59]. At the first stage, the dermatologists should differentiate between melanocytic and non-melanocytic lesions. Afterwards, the melanocytic lesions should be categorized as benign, suspect, or malignant. However, the automatic extraction of visual features is one of the difficult steps for the development of CAD systems because it may require domain-specific knowledge of pattern recognition techniques and segmentation of lesion area.

The development of such CAD systems will be truly beneficial in the routine diagnosis of skin lesions that can increase the accuracy and reproducibility of the results. In clinical practice, the classification of pigmented skin lesions (PSLs) is a difficult task due to limited size data sets and it is difficult to reproduce the wide variety of lesions. Furthermore, those CAD systems required lots of domain expert knowledge about pre- and post-processing on skin lesions to extract and classify skin lesions. Also, the state-of-the-art CAD system relies on correct segmentation of lesion area, which is often difficult to achieve, and computationally expensive due to use of conventional image processing algorithms.

In the literature, many CAD systems [8, 38, 51, 59, 60, 67] were proposed but a few systems were found, which focused on CADs in the field of automatic classification of skin cancer. Many of those CAD systems employed non-uniform color spaces such as RGB, CIELUV, and HSV. In general, if a computerized system uses a non-uniform color space, then it is not easy to extract informative features. In the field of skin cancer, the extraction of effective texture and color features [44] is a challenging task for differentiating between melanomas and nevi. In previous studies, a large feature vector is utilized for recognition of melanoma-nevus lesions. In fact, the previous recognition system is strictly based on the direct matching of malignant melanomas or benign lesions through feature extraction and selection steps.

Instead of CAD systems, the dermatologists have widely utilized the clinical rules such as ABCD, CASH or pattern analysis to categorize pigmented skin lesions (PSLs) when diagnosing through digital dermoscopy. Clinical diagnosis systems such as the seven-point checklist, the three-point checklist, the CASH (color, architectural order, asymmetry of pattern and homogeneity) algorithm and ABCD (asymmetry, borders, colors, diameter) [1]. Those clinical rules are subjective measured and therefore, there were some computerized systems develop to make this process automatic. Those previous CAD systems are explained in the following paragraphs.

Computer-aided diagnosis systems (CAD) [1–3, 6–9, 12, 14, 15, 28–31, 37, 38, 44, 50, 55, 56, 58, 60, 63, 67, 72] were developed in the past studies by using automatic complex image

processing techniques. In the literature studies, it concluded that those CAD systems were focused more on extracting color and texture features and emphasize on feature's selection methods. By segmenting the area of skin lesions, those techniques extracted color and texture features in non-uniform color space. Although, a significant research was devoted to the past studies to develop computer-aided diagnosis (CAD) systems for classification of pigmented skin lesions (PSLs). Those systems needed domain knowledge to develop CAD system in terms of image enhancement, segmentation, and feature selection [45] techniques. We have also observed that the classification accuracy of a limited category of PSLs is less than 90%, which indicates the need for developing a new algorithm. In addition to these issues, the computational time is also high on those kind of systems.

Instead of those classification systems, there are also CAD systems developed in the past studies based on deep learning algorithms [17, 18, 21, 32, 54, 71]. The authors reported that they have obtained significantly higher classification accuracies compared to previous systems. The CAD systems based on deep learning are explained in the next section.

1.1 Literature review

In contrast to all above-mentioned CAD systems, a few deep learning algorithms [17, 18, 21, 32, 54, 71] were utilized in the papers suitable for clinical images [32, 54] when compared to dermoscopy images [17, 18, 21, 71]. In this paper, we focus on dermoscopy images as these digital images contain more visual features than clinical images. However, where appropriate, we discuss both types of images in this paper because there are published systems for clinical images based on deep learning.

Deep learning is considered as the best object classification algorithm in many computer vision applications. In fact, it has many applications in medicine. Algorithms based on deep learning proved to be the best due to unsupervised learning, and feature transforms to eliminate the need for annotated data in the target task to learn good features. The deep-learning methods do not require pre- and post-processing steps and, when implemented carefully, they are computationally inexpensive.

In [54], the authors developed a melanoma recognition system based on clinical images. Image enhancement, segmentation, and extraction of features were integrated as main steps. To classify melanoma skin lesions, the authors used fifteen features that were fed into the deep learning and hybrid AdaBoost-Support Vector Machine (SVM) algorithms. This system was tested on 992 images and obtained a high 93% classification accuracy. However, the authors implemented the contrast enhancement solution and segmentation methods as pre-processing steps. Whereas in [32], a linear classifier was trained on clinical images to extract features through a convolutional neural network (CNN) model. The authors argued that this system did not require lesion segmentation or complex image preprocessing steps. A CNN model is used to extract multi-scale features from each image and then utilized pooling step to select most discriminative features. On a 5-class problem, the authors reported an accuracy of 85.8%.

A melanoma recognition system for dermoscopy images is developed in [17] by integrating deep learning, sparse coding, and support vector machine (SVM) learning algorithms. A dataset of 2624 images was obtained from the International Skin Imaging Collaboration (ISIC) archive. The authors achieved 93.1% accuracy, 94.9% sensitivity and 92.8% specificity for the first experiment to differentiate between melanoma and all non-melanoma skin lesions through two-fold cross-validation. Instead of recognition of melanoma, a pattern detection approach was

presented in [18] to detect a typical network, regular and globules. The authors discussed that the ABCD-score was computed by using these patterns from dermoscopic images. For these limited number of patterns, they used deep convolutional neural networks (DC-NN) machine learning algorithm with 8 layers. The authors reported an accuracy below 85% on 211 images.

Moreover, in [21], a deep neural network (DNN) machine learning algorithm was used to recognize melanoma skin lesions without conventional computer vision models. The authors reported that the best results were obtained using DNNs model compared to the bag of visual words with AUC of 89.3%. In [71], a CNN classifier with five-fold cross-validation test was utilized to test the melanoma recognition system on a set of 1760 dermoscopy images including 329 of melanomas and 1431 of nevi skin lesions.

From Table 1, it was also noticed that the previous computerized systems mostly utilized the convolutional neural networks (CNN) to extract features and then use linear softmax classification for prediction tasks. The CNN's model is one of the famous variants of deep learning algorithms. However, it is used to select features for multiple objects present in an image and it utilized grayscale features by ignoring color features that are most important for skin cancer images. Moreover, the CNN model also ignores spatial correlation among pixels. As there are six possible distinct colors present in skin lesion images that are crucial for differentiating between benign and malignant lesions. Accordingly, this CNN model is not ideal for skin lesion classification tasks.

Deep learning [35] is a machine learning framework that has best-in-class performance on problems and significantly outperforms other solutions in various domains. This includes speech, language, vision, and games [19]. The automatic encoding of features which previously had to be hand engineered. The exploitation of structurally/spatially associated features. In practice, deep learning algorithms were designed with input, multiple visible and output layers to extract and classify the features. Therefore, these algorithms intelligently eliminate the feature engineering step that is often the most time-consuming part [65]. Moreover, feature engineering needs extensive domain knowledge. These architectures of deep learning algorithms [24] are used in many computer vision and bioinformatics applications. We noticed that the deep learning algorithms take advantage over the neural network in terms of a limited set of features defined in the training dataset. In deep learning algorithm, it is better if you have thousands of training examples to outperform other approaches.

Instead of using deep features, there are many other developed past studies [25, 26, 36, 39–42, 52, 68–70] that focused on multi-feature fusion from digital images. Those algorithms were, however, not tested on dermoscopy images. Those studies were dedicated towards extracting different features and then classify them through a machine learning algorithm. To extract and define those multi-feature fusion, domain knowledge is required. However, multi-feature fusion algorithms are considered the best way to classify objects along with deep learning algorithms. As a result, in this paper, the multi-feature fusion based approach is developed in a perceptual-oriented color space.

Actually, the deep learning algorithm is set of machine learning algorithms that share the weights among different layers. In many different studies, the researchers are adding more layers to recognize objects. The deep learning algorithms have many variants to represent visual features such as the convolutional neural network (CNN), recurrent neural network (RNN), deep belief networks (DBN), restricted Boltzmann machine (RBM) and Autoencoders. In this paper, the color visual features are extracted in perceptual-oriented color space from image compared to grayscale features as utilized in the past studies. Afterwards, the

Table 1 Current state-of-the-art detection and classification of computer-aided diagnosis (CAD) systems for recognition of pigmented skin lesions (PSLs)

Cited	Purpose	Methodology	Results	Compare
[8]	A CAD System is proposed to diagnose dermoscopy images.	Color features and LDA algorithm	Precision: 83.8% and Recall: 89.8%	Yes
[60]	Classification of global patterns in dermoscopy images	a Gaussian model, a Gaussian mixture model, and a bag-of-features histogram model on CIEL*a*b* color space	Acc: 78.44%	No
[58]	To identify the absence or presence of streaks to detect melanoma	Color and texture features	Acc: 76.1% and AUC: 85%	No
[37]	shades of blue (lavender blue, light blue, and dark blue)	Color features with Fuzzy logic	Acc: 82.7%	Yes
[7]	systems for the detection of melanomas in dermoscopy images	local color and texture features and the bag-of-features classifier	SE: 100% and SP: 75%	No
[56]	Classification accuracy of melanocytic and non-melanocytic pigmented lesions	Color and texture features	SE: 79.3%, FPR: 90.5%	No
[12]	To develop a computer algorithm for the diagnosis of melanocytic lesions	Six parameters and logistic regression analysis	Acc: 82%	No
[29]	classification of acral melanoma skin lesions	Color and texture features with statistical analysis	AUC values: 0.985	No
[3]	Prediction of Dermoscopy Patterns for Recognition of both Melanocytic and Non-Melanocytic Skin Lesions	Color and texture features with majority-voting support vector machine (MV-SVM)	SE: 94%, SP: 84%, ACC: 93% and AUC: 0.94	Yes
[55]	A computational approach was developed to classify skin lesions.	After segmentation, a set of different features such as based on asymmetry, border, color and texture analysis are extracted.	NA	No
[31]	a system to classify different melanocytic lesions	image enhancement, lesion segmentation, feature extraction, and selection as well as classification	Acc: 92%	No
[14]	A classification system for two-class problem such as benign versus malignant.	A machine learning approach to the automated quantification of clinically significant colors in dermoscopy images.	SE: 62% and SP: 76%	No
[45]	A search for effective features was carried out using a support vector machine.	Shape and color features were utilized.	Acc: 83.0%.	No
[54]	developed melanoma recognition system through image enhancement, segmentation, and 15 features were extracted	Deep learning and hybrid AdaBoost-Support Vector Machine (SVM)	Acc: 93%	Yes
[32]			Acc: 81.8%	No

Table 1 (continued)

Cited	Purpose	Methodology	Results	Compare
	a fully convolutional network to extract multi-scale features, and by pooling over an augmented feature space	CNN and pooling techniques applied on a 10-class dataset of 1300 images		
[17]	A differentiation system for (1) melanoma vs. non-melanoma lesions and 2) melanoma vs. atypical lesions only.	A system developed to combine the deep learning, sparse coding, and support vector machine (SVM) learning algorithms	Acc: 93.1% SE: 94.9% SP: 92.8% ACC: < 86%	No
[18]	A pattern detection approach was developed.	8-layer model of deep convolutional neural networks (DC-NN)		No
[21]	Recognize melanoma skin lesions	deep neural networks (DNNs)	AUC of 89.3%	Yes
[71]	Detecting malignant melanomas and carcinomas matched.	convolutional neural networks (CNNs)	NA	No

^a AUC = Area under the receiver operating curve (AUC)

^b Acc = Accuracy

^c SE = Sensitivity

^d SP = Specificity

^e FPR = False positive rate

DBN model is adopted as the deep learning algorithm to describe features of an image containing a single object. Also, the RNN is used to further optimize features.

A review of the current literature suggests that there are few research studies that focused on deep learning algorithm especially they applied a CNN model to classify skin lesions based on dermoscopy images. In those studies, the authors claimed that there is no need to use pre- and or post-processing or domain knowledge. Those systems are briefly outlined in Table 1. According to this table, the CAD systems developed through deep-learning algorithms achieved better classification accuracy. However those CAD systems were developed in non-uniform color spaces and grayscale features extracted through CNN model that required lots of training examples. As a result, there is a dire need to develop a CAD system for classification of skin lesions in perceptual-oriented color space that includes the fusion of visual features. Hence, we have developed a DermoDeep system to solve all above-mentioned problems and to assist dermatologists for better diagnosis skin cancer. The main contributions to this article are briefly described in the next subsection.

1.2 Major contributions

We have highlighted the major contributions of this research paper as given below.

- (1) A comprehensive review of pre- and post- processing computer-aided diagnosis (CAD) and deep learning algorithm-based CAD systems is presented.
- (2) A new aggregation of visual features along with descriptors are extracted in a perceptual-oriented color space. We utilize color visual features compared to the grayscale-level features employed by the past studies and avoid the ABCD or CASH rule defined by clinical experts.
- (3) A novel DermoDeep system is developed based on multilayer architecture that contains five main steps such as the construction of visual features layer (VF-L), deep features layer (DF-L), features fusion layer (FF-L), optimization of features layer (OF-L) and prediction layer (FF-PL) to recognize melanoma-nevus skin lesions.

2 Methodology

2.1 System overview

The main objective of this paper is to consider the latest deep learning algorithms to make differentiate between melanomas and nevi in dermoscopy images without focusing on pre- or post-processing steps. In this study, the combination of visual features and multilayer architecture of deep learning algorithms are integrated together for better classification results. The fusion of multiple features and best model of deep-neural-network (DNN) techniques are integrated to develop the DermoDeep system. The DermoDeep contains five main steps: construction of visual features layer (VF-L), deep features layer (DF-L), features fusion layer (FF-L), optimization of features layer (OF-L) and prediction layer (FF-PL) to discriminate between melanomas and nevi in dermoscopic images. These steps are detailed explained in the methodology section. The systematic diagram of these five steps is given in Fig. 1. The development of DermoDeep is explained in the subsequent paragraphs.

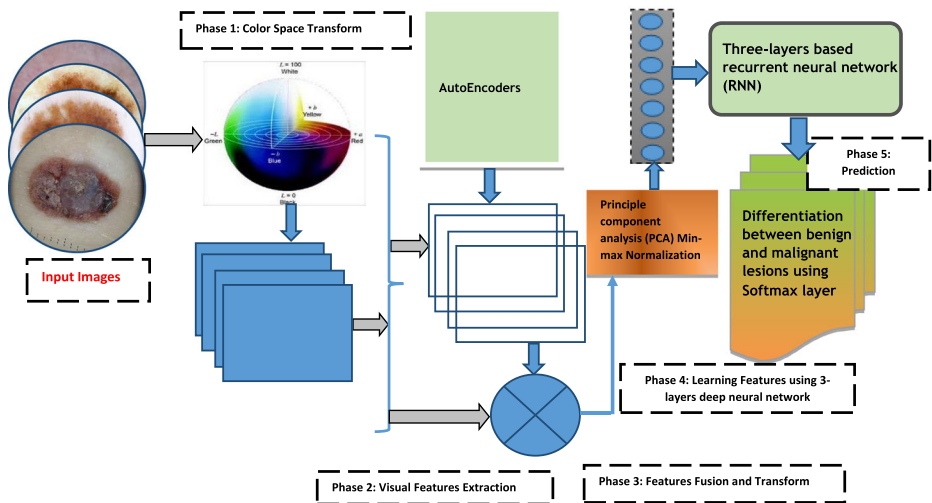


Fig. 1 Systematic diagram of the proposed system to classify nevi versus malignant melanoma pigmented skin lesions (PSLs) when diagnosis through dermoscopy

2.2 Construction of visual features layer (VF-L)

A digital dermoscope is used to acquire the skin lesion images and these images are often represented in the RGB color space. An example dermoscopy image is shown in Fig. 3. It is well known that the RGB color space is not perceptually uniform [1]. In many studies, these RGB images are converted to grayscale, which leads to a poor representation of dermoscopic structures. More details about this phenomenon can be found in [1]. State-of-the-art studies mostly use the CIECAM02, HSV, HSI, CIEL*a*b* [46] non-uniform color spaces. In some studies, the CIEL*a*b* color space is converted into perceptual-oriented color space using advanced distance metrics such as CIE94 or CIEDE2000 [43, 62].

In this study, the CIEL*a*b* color space is utilized with an advanced distance metric, namely CAT2000 [43] to account for all color appearance transformations. This color appearance model defines three dimensions of color appearance such as lightness L^* , Chroma (a^* and b^*) [47, 48, 61, 62] channels. As a result, this CIEL*a*b* with CAT2000 uniform color space [47] is adopted in this paper. All hand-crafted color and texture features are extracted in this paper from this perceptually-oriented CIEL*a*b* color space. Instead of using grayscale features through deep-learning CNN model, the hand-crafted visual features are also extracted to get better classification accuracy. The color coherence and texture features are derived in this space.

To define color coherence, the frequency of six colors such as black, red, light brown, dark brown, blue-gray and white are extracted through the Hill climbing algorithm (HCA) [4] technique. In this paper, the HCA algorithm technique is updated in terms of K-means clustering method [16] in the uniform color space CIE $L^*a^*b^*$ with the advanced color difference equation CIE2000. Moreover, the texture features are defined using the speed-up robust features (SURF) [10] technique from fixed point-of-interest (POIs). The goal of a SURF descriptor is to provide a unique and robust description of an image feature, around the POIs. Afterwards, a bag-of-words (BoW) representation is constructed using the k-means clustering algorithm based on SURF features that also produce a feature or visual word.

The bag-of-Features object defines the features, or visual words, by using the k-means clustering (implemented in MATLAB's Statistics and Machine Learning Toolbox) algorithm on the feature descriptors extracted from training sets. The algorithm iteratively groups the descriptors into k mutually exclusive clusters. The resulting clusters are compact and well separated. Each cluster center represents a feature or visual word.

2.3 Construction of deep features layer (DF-L)

Color and texture features defined from the previous VF-L layer are hand-crafted features. These hand-crafted features can lose important information whereas the features are extracted from stack-based-AutoEncoders (SAE) are designed to minimize information loss. Therefore in this paper, the hand-crafted and SAE features are both utilized to provide better local and abstract representation of features without losing color spatial information. Moreover, the SAE method is able to design features by discounting most of the illumination variance and artifacts presented in different dermoscopic datasets.

SAE autoencoders are a type of unsupervised neural network [11, 35] contained in multiple layers organized in as a stack-based. Each layer represents features of the new abstract feature along with minimal information loss. In practice, the deep neural networks are composed of many layers of neural units, and in SAE autoencoders, every pair of adjacent layers forms a full bipartite graph of connectivity. The features are an abstract representation of the whole image. An image is a collection of pixels, which are collectively summarized as edges that form objects. At the center-most layer, the dimensionality is at a minimum. From there, the network reconstructs the original data from the abstract features and compares the reconstruction result against the original data. We noticed that a similar reconstruction may not be feasible with hand-selected features compared to features defined by SAE autoencoders.

Stack-based auto-encoder (SAE) is trained on the dataset to encode the x (features or pixels) into some representation $R(x)$ so that the input can be reconstructed from that representation. Therefore, the target output of the SAE is the auto-encoder input itself. In SAE architecture of unsupervised learning, the hidden layer behaves like non-linear to capture multi-modal aspects of the input distribution. The formulation that we prefer generalizes the mean squared error criterion to the minimization of the negative log-likelihood of the reconstruction, given the encoding $R(x)$:

$$Error = -\log P(x|R(x)) \quad (1)$$

And the loss function would be calculated as

$$-\log P(x|R(x)) = -x_i \log f_i(R(c)) + (1-x_i) \log(1-f_i(R(c))) \quad (2)$$

Where $f_i(\cdot)$ is called the i^{th} decoder, and $f(R(x))$ is the reconstruction produced by the network. The code $c(x)$ is a distributed representation that captures the main factors of variation in the data.

2.4 Construction of features fusion layer

The features fusion layer (FF-L) combines together the features from visual descriptor layer (VD-L) and deep features layer (DF-L) through a transform technique that is known as a principal component analysis (PCA) process. In this features fusion layer, we have

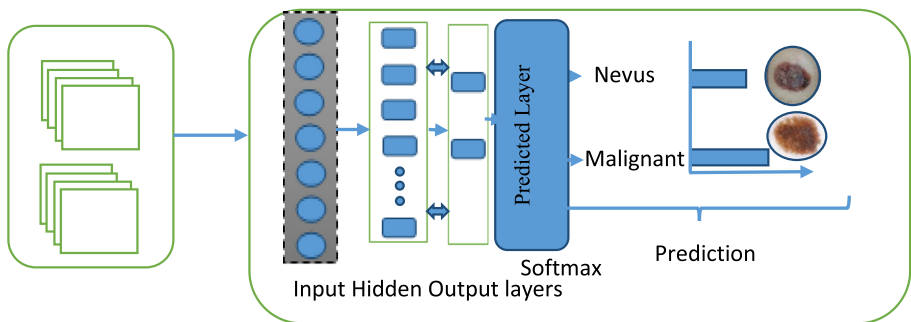
combined hand-crafted visual features along with deep-features to solve the problem of multi-components architecture. Those multi-component architectures are presented in many dermoscopy images. Therefore, we need a system to extract distinct effective features. Through the PCA process, the distinct features are selected to determine the identity of features. In this paper, the PCA is used to reduce the dimensionality of feature space by calculating the eigenvectors of the covariance matrix. The PCA algorithm is sensitive to noise but it provides the best discriminative power of features that are further optimized using recurrent neural network (RNN) model. This model is explained in the next sub-section.

2.5 Optimization of features layer

The recurrent neural networks (RNNs) model has the capability to optimize features from the features fusion (FF-L) layer. In the past studies, RNNs were used to achieve good results in sequence modeling and optimization tasks. Therefore in this paper, the RNN model is used to optimize the features that can provide better classification results. In the past studies, the CNN model was applied to extract and select features from input images. However, the RNN model [33] enhances this feature extraction process and using the outputs or hidden states of the recurrent units to compute the extracted features. The RNNs model is known as recurrent [23, 65] because they perform the same task for every feature of a sequence, with the output being dependent on the previous computations. The visual description of RNNs model is shown in Fig. 2.

The architecture of RNNs model is different in terms of neurons compared to feed-forward neural network approach. In the feed-forward neural network, the network is organized via layers and information flows uni-directionally from input pixels to the output. However, in RNN architecture, the flow of information is undirected cycles in the connectivity of like some patterns. The RNNs do not have to be organized in layers and directed cycles are allowed. In fact, neurons are actually allowed to be connected to themselves.

The selected features (f_1, f_2, \dots, f_T) from OF-L layer are provided to RNN model as an input vector sequence that pass through weighted connections to a stack of N recurrently connected hidden (h_1, h_2, \dots, h_T) layers and then the output layer (O_1, O_2, \dots, O_T) is defined. Each output vector (O_t) is utilized to parameterize a predictive probability of distribution $P(f_{t+1} | O_t)$ over



Input Features map

RNN model

Fig. 2 Visual features recognition using recurrent neural network (RNN) and softmax linear classifier to differentiate nevus and malignant melanoma pigmented skin lesions (PSLs) when diagnosis through dermoscopy

the possible next inputs f_{t+1} . The hidden layer activations are computed by iterating the following equations from $t = 1$ to T and from $n = 2$ to N and mathematically defined in Eq. (3).

$$h_t^1 = H(W_{ih^1} f_t + W_{h^1 h^1} h_{t-1}^1 + b_h^1) \quad (3)$$

and

$$h_t^n = H(W_{ih^n} f_t + W_{h^{n-1} h^n} h_{t-1}^{n-1} + W_{h^n h^n} h_t^1 + b_h^n) \quad (4)$$

where W_{ih^n} term represents weighted matrices connected with input features to n th hidden layer and recurrent hidden layers is denoted as $h^n h^n$. Also, the b terms shown as bias vector and H is hidden layer function.

The output vectors (O_t) are utilized to parameterize the predictive $P(f_{t+1} | O_t)$ for the next input layer. The backpropagation step is used with partial derivative to apply to RNN graph for computing the loss sequence and to train the network.

2.6 Construction of features classification layer

The Softmax linear classifier is used to discriminate between melanomas and nevi in a supervised fashion with known class labels (Y). It is a statistical model that attempts to learn all weight and bias parameters by using the learned features of the last hidden layer. In the case of binary classification ($\text{class} = 2$), the softmax regression hypothesis outputs h . If there are K text classes in total, and class k is fed in at time t , then O_t is a length K vector whose entries are all zero except for the k th, which is one. The predictive $P(f_{t+1} | O_t)$ is therefore a multinomial distribution, which can be naturally parameterized by a softmax function at the output layer:

$$P(O_{t+1} = K | y_t) = y_t^k \left(\exp(y_t^k) / \sum_{k=1}^n \exp(y_t^k) \right) \quad (5)$$

In general, the experiments in this paper aim to predict at the finest granularity found in the data, so as to maximize the generative flexibility of the network.

3 Experimental results

3.1 Acquisition of Datasets

To test and compare the DermoDeep, 1400 melanomas and 1400 nevi pigmented skin lesions (PSLs) were acquired from various public and private sources. The short description of these data sources is presented in Table 2. As shown in this table, the first dataset (Skin-EDRA) was collected as a CD resource from the two European university hospitals as part of the EDRA-CDROM [5]. There were 1064 dermoscopy images in this dataset but 500 PSLs including (Melanoma of 250 and benign of 250). In Skin-EDRA dataset, the size of images is (768×512) pixels. The second dataset (ISIC) was collected from the International Skin Imaging Collaboration (ISIC) [27]. In this ISIC dataset, there are more than 10,000 dermoscopy images. However, 800 melanoma and 800 benign skin lesions were selected in this research study. Moreover, the 300 melanomas and 300 benign skin lesions (DermNet) [66] were acquired from the Department of Dermatology, University of Auckland. Whereas the Ph2-dataset [49] contains a total of 200 dermoscopy images including 80 common nevi, 80 atypical nevi, and

Table 2 Acquisition of 2800 dermoscopy images dataset to test the performance of DermoDeep system

Cited	Data set	Class	Size	Dermoscopy Images
[5]	Skin-EDRA	250 of Nevus and 250 of Malignant skin lesions	(768 × 512) pixels	500
[27]	ISIC	800 of Nevus and 800 of Malignant skin lesions	(768 × 512) pixels	1600
[66]	DermNet	300 of Nevus and 300 of Malignant skin lesions	(768 × 512) pixels	600
[49]	Ph2-dataset	30 melanomas and 70 nevus lesions	(768 × 560) pixels	100
			Total Images:	2800

Skin-EDRA. Argenziano G, Soyer HP, De Giorgi V, Piccolo D, Carli P, Delfino M et al. Interactive atlas of dermoscopy CD. Milan: EDRA Medical Publishing and New Media, 2000

ISIC. University of Auckland, New Zealand. Dermatologic image, database. <http://dermnetnz.org/doctors/dermoscopy-course>

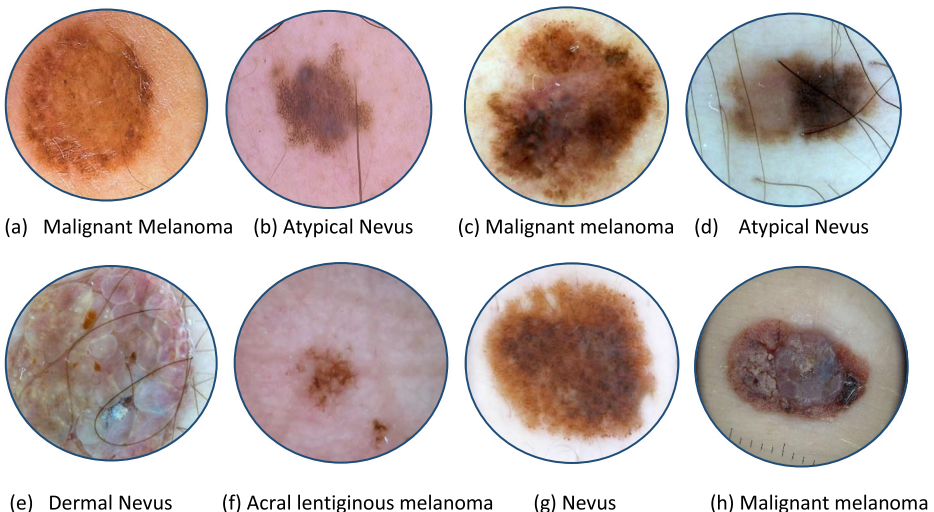
DermNet. Matt Berseth, ISIC 2017: Skin Lesion Analysis towards Melanoma Detection, Computer Vision and Pattern Recognition, <http://isdis.net/isic-project>

Ph2-dataset. Mendonca T, Ferreira PM, Marques JS, Marcal AR, Rozeira J. PH2- a dermoscopic image database for research and benchmarking. Conf Proc IEEE Eng Med Biol Soc. 2013; 2013:5437–40. doi: 10.1109/EMBC.2013.6610779. <http://www.fc.up.pt/addi/ph2%20database.html>

40 melanomas with a resolution of 768 × 560 pixels. From Ph2-dataset, the 30 melanoma and 70 nevus skin lesions are selected to test the performance.

The main objective of this paper is to develop a new recognition system to discriminate between melanomas and nevi without doing pre- or post- processing steps on dermoscopy images. The images in this dataset contain hair and artifacts such as dermoscopy gel. However, such artifacts should have negligible impact on classification accuracy thanks to the use of robust deep-learning algorithms.

The 2800 images were automatically resized to (600 × 600) pixels. Afterwards, the circular region-of-interest (ROIs) images of size (400 × 400) pixels were automatically extracted from the center of the image without losing an area of the tumor. Example ROI images are given in Fig. 3.

**Fig. 3** An example of skin lesions dataset utilized in the development of DermoDeep system

3.2 Experimental setup

The proposed DermoDeep system is implemented in MATLAB 2016. In total, 2800 dermoscopic images (1400 melanomas and 1400 nevi) were used to test the system. Features of these 2800 images are extracted and stored in a file to test the performance of the proposed DermoDeep system. The classification decision is performed by combining the fusion of features and multilayer deep learning algorithms. 50% of the images are used to train the proposed deep learning classifier and 50% are used to test the classifier (Fig. 4).

Parameters setup To minimize information loss, Stack-based-Autoencoders (SAE) are utilized to extract features from PSLs images. An encoder network is set as a fully connected multilayer perceptron (MLP) with dimensions (600×600) and the decoder network is a mirror of the encoder. Except for input, output and embedding layers, all internal layers are activated by the nonlinear ReLU function using 450 SGD epochs. After this, principle component analysis (PCA) is used to select features that have most discriminative strength. In this paper, 750-element eigenvalues are selected. Then, the RNN model is a statistical model applied that attempts to learn all of weight and bias parameters by using the learned features of the last hidden layer. We used $(\text{폭}=2)$ parameter value in the case of binary classification. To do comparisons, the basic CNN model was used to extract multi-scale features and then use pooling concept to select important features. We used network structure that is conv5 32 \rightarrow conv5 64 \rightarrow conv3 128 \rightarrow Pooling 2 where convk n denotes a convolutional layer with n filters, kernel size of $(k \times k)$ and stride length 2 as default value.

A statistical analysis was also performed to evaluate the suitability of the proposed DermoDeep system for skin cancer screening. This statistical analysis was also used to

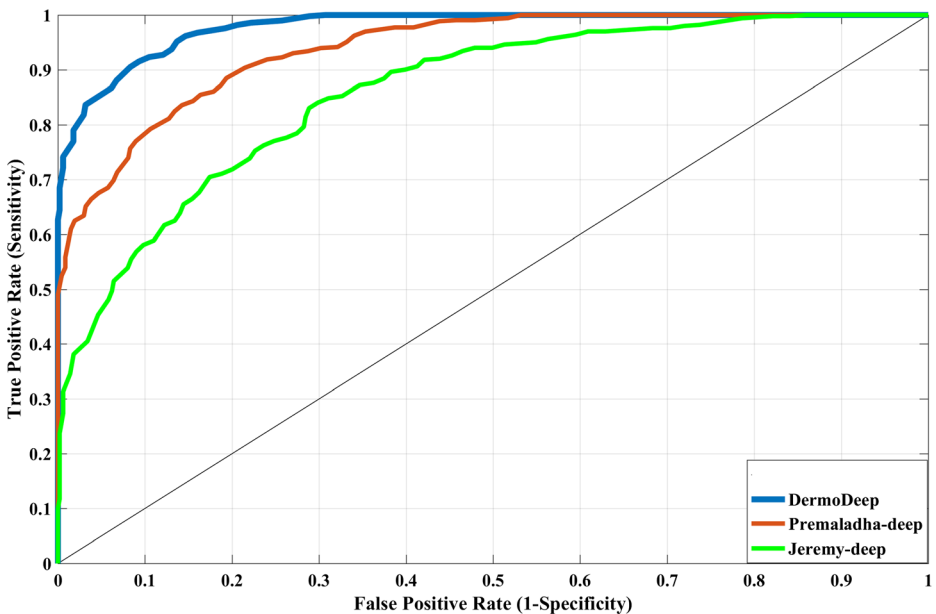


Fig. 4 Performance comparisons of proposed DermoDeep system with state-of-the-art classification systems in terms of Area under the Receiver operating curve

compare the DermoDeep algorithm with two state-of-the-art methods. The area under the receiver operating curve (AUC) [13] measure is used along with three other measures such as sensitivity (SE), specificity (SP) and training errors (E). AUC is a commonly used performance index for quantifying the overall discriminative capability of classifiers. AUC values range between 0.5 and 1.0 and the higher the value, the greater the classification accuracy. The receiver operating characteristic (ROC) curve is basically defined as a graphical plot to illustrate the performance via a fixed threshold value of 0.5. The SE and SP statistical measures are calculated by deriving the true positive rate (TPR) against the false positive rate (FPR), respectively.

The TPR metric is known as a sensitivity measure, while FPR is known as $(1 - \text{specificity})$. If the classifier is predicted positive (P) case and the actual (T) is also the same then this known as TP. However, if the actual class label is negative (N) then it is known as false positive (FP). On the other hand, a true negative (TN) has followed when both the estimated class label and the actual value are n, and false negative (FN) is when the prediction outcome is n while the actual value is p. Therefore, the ROC curve is plotting the Sensitivity in the y-axis versus the Specificity of the false-alarm in the x-axis.

$$\text{Sensitivity (SE)} = \text{True Positive Rate (TPR)} = TP / (TP + FN) \quad (6)$$

$$\text{Specificity (SP)} = 1 - (FP / (FP + TN)) \quad (7)$$

$$\text{False Positive Rate (FPR)} = FP / (FP + TN) \quad (8)$$

The comparisons are performed with Premaladha_deep [54] and Jeremy_deep [32] systems by using ROC curve analysis and 10-fold cross-validation test. We compare against these two systems because they too are based on deep learning algorithms. In Premaladha_deep [54] system, first, the dermoscopy images are enhanced using Contrast Limited Adaptive Histogram Equalization technique (CLAHE) and median filter. After that, the tumor region is separated through a Normalized Otsu's Segmentation (NOS) algorithm. As mentioned in [30], fifteen features are extracted and fed into the Deep learning based Neural Networks and Hybrid Adaboost-Support Vector Machine (SVM) algorithms. On the other hand, in Jeremy_deep [32] system, a convolutional neural network (CNN) model is used to extract features and then classified using a linear softmax classifier. A CNN is used as in [32] to extract multi-scale features and then the pooling concept is used to select important features for classification.

3.3 Computational cost

The DermoDeep took on average of 6.46 s to extract all of the features based on hand-crafted and deep features in perceptual-oriented color space i.e., CIEL*a*b*. The features transform step took around 2.5 s to normalize all features. The training of the RNN optimization of the features based on deep learning took 13.45 s on average while the final out layer took 1.78 s on average to be created. Therefore, a total time of 5.20 s was consumed for melanoma-versus-nevus class, on a fixed number of 300 iterations, for training the deep learning neural network classifier. However, when the training has been done and the test is performed, an average time

of 8.12 s is only needed to classify the image. The computation time spent on feature extraction and creation of the dictionary of visual features could be further reduced by using an optimized C/C++ implementation. However, the computational time of DermoDeep was 2.1 s high compared to both systems developed by Premaladha_deep and Jeremy_deep using convolutional neural networks (CNN) model. It is due to the fact that we have introduced hand-crafted visual features in a perceptual-oriented color space.

3.4 Evaluation of results

To evaluate the effectiveness of DermoDeep system, a comparison with the state-of-the-art system is also done. The comparisons are performed with Premaladha_deep [54] and other presented by Jeremy_deep [32] using ROC curve analysis and 10-fold cross validation test. The results of the DermoDeep system on this dermoscopy data set containing 2800 dermoscopic images are shown in Table 3. This Table 3 demonstrates that sensitivity: SE, specificity: SP, Accuracy: ACC, training errors: E and area under the ROC curve (AUC) statistical analysis results. Moreover, the Fig. 5 has shown the corresponding receiving operating characteristic curve (ROC). The area under the curve (AUC) shows the significant result of this DermoDeep system, which is greater than 0.5. From this table, it is cleared that DermoDeep system has provided greatly improved results by extracting fusion of informative visual features and deep features in a perceptual-oriented color space.

The average value of AUC of 0.96, SE of 93%, SP of 95% and ACC of 95% are measured to recognize melanoma versus nevus skin lesions. The average low training rate (E) of 0.60 is also observed. These results are represented in Table 3. For differentiation, the results have been improved as compared to state-of-the-art systems. It is because of designing an effective feature fusion and multilayer deep classification solution along with optimal color-texture features extraction in a perceptual uniform color space. Therefore, this proposed system is giving a valuable and intuitive aid to the clinician in the decision-making process. In fact, this system can be useful as a training tool for medical students and inexperienced practitioners given its ability to recognize large collections of PSLs images using their visual attributes.

These results are also compared with the state-of-the-art systems in terms of SE, SP, ACC and AUC statistical measures. These comparisons results are displayed in terms of training errors in Table 4. A low SE, SP, ACC and AUC measures are obtained in the case of Premaladha_deep [54] and Jeremy_deep [32] compared to systems. As shown in this table, the SE of 78%, SP of 80%, ACC of 79% and AUC of 0.82 average values are obtained in the case of Jeremy_deep classification system. Whereas in the case of Premaladha_deep [54] system, the SE of 80%, SP of 83%, ACC of 82% and AUC of 0.84 average values are measured. The Premaladha_deep system is somehow obtained higher classification results

Table 3 The average values of Sensitivity (SE), Specificity (SP), ACC:Accuracy, E: training errors and area under the receiver operating curve (AUC) of the DermoDeep system on 2800 dermoscopy images

No.	Data set	SE	SP	ACC	AUC	E
1	Malignant Melanomas	92%	95%	93%	0.94	0.76
2	Nevus	94%	95%	94%	0.95	0.67
	Average Result	93%	95%	95%	0.96	0.60

SE: Sensitivity, SP: Specificity, ACC: Accuracy, E: training errors and AUC: area under the receiver operating curve

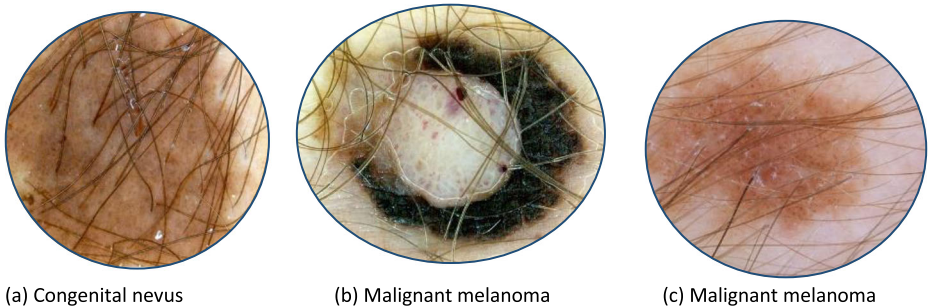


Fig. 5 An example of misclassified where figure (a) represents nevus and other are malignant melanomas (b, c)

compared to the Jeremy_deep [32] system but overall less significant compared to the DermoDeep system. These results indicate that this new DermoDeep system can be used to effectively provide the second opinion to fewer experienced dermatologists for diagnosis of pigmented skin lesions.

The dermoscopy images are used in this paper to classify skin cancer into benign and malignant lesions. As mentioned-above, the CNN model is difficult to train due to multi-layer architecture and this model is only used to extract features from raw pixels of images. The features selection is also hard in this CNN model compared to Derma-Deep system.

4 Discussions

The proposed DermoDeep system has tested on 2800 dermoscopic images and the results demonstrate the viability of this automatic classification system for diagnosis of PSLs lesion. A major advantage of this algorithm is that the accuracy achieved for detection is very high which implies greater accuracy of PSLs classification. Therefore, computational modeling and analyses done here can serve as foundation chunks for the advanced research in skin cancer and eventually help in their better management. Regarding other a comparison with state-of-the-art methods for recognition of melanoma versus nevus types, the Premaladha_deep [54] and Jeremy_deep [32] systems were used but the DermoDeep obtained significance higher classification rate. It is due to the fact that the DermoDeep system used perceptual-oriented color space, the fusion of visual features and three multilayer deep neural network concepts. As it was already discussed in the related work section, only a few efforts have been devoted to the recognition task [1–3, 6–9, 12, 14, 15, 28–31, 37, 38, 44, 45, 50, 55, 56, 58–60, 63, 67, 72] on a limited number of types of PSLs lesions. Especially in [22], the authors utilized pre-processing and segmentation steps to detect region-of-interest for extraction of features.

Table 4 The average values of Sensitivity (SE), Specificity (SP), ACC:Accuracy and area under the receiver operating curve (AUC) of the DermoDeep system compare to other two systems on 2800 dermoscopy images

No.	Data set	SE	SP	ACC	AUC
1	Jeremy_deep [54]	78%	80%	79%	0.82
2	Premaladha_deep [32]	80%	83%	82%	0.84
3	DermoDeep	93%	95%	95%	0.96

SE: Sensitivity, SP: Specificity, ACC: Accuracy and AUC: area under the receiver operating curve

Afterward, they used the traditional old classifiers such as the neural network to classify the features. However, it was very difficult to select the prominent features to perform recognition tasks. Also in the past studies, the majority of the methods rely on the pre-processing to enhance the skin lesion, precise segmentation of some elements of the images and post-processing step such as selection of features. Those studies utilized the machine learning techniques that are not recognized the wide variety of different types of pigmented skin lesions (PSLs).

In particular, the Neural network or SVM algorithms are designed to learn features based on feed forward or backward step using input layer - hidden layer - output layer fashion. Whereas deep learning algorithms are set of machine learning algorithms that share the weights among different layers. In the past studies, the researchers are adding more layers to use it in different applications. In deep learning algorithms, the convolutional neural network (CNN), deep belief network (DBN), stack-based AutoEncoders (SAE), Restricted Boltzmann Machine (RBM) and recurrent neural network (RNN) are the different variants. These variants are used in different situation. For example, the CNN used when an image contains multiple objects. It is used to select features from direct images and represented them using feature map. The DBN is used to best describe features of the image means single object. The SAE model is used to direct learn features from the image and try to automatically select most distinguishable features from an image. Whereas the RNN deep learning variant is used to optimize features. After describing the features, the supervised Softmax linear classifier is used to recognize them according to class. For classification of nevus-malignant skin lesions, the combination of SAE and RNN variant of deep learning model is performed to effectively define features and recognize using Softmax linear classifier. The overall system is represented as DermoDeep based on the fusion of features and deep learning techniques for classification of skin lesions.

The DermoDeep system is developed based on five main layers to classify melanoma-nevus skin lesions. These layers are visual features layer (VF-L), deep features layer (DF-L), features fusion layer (FF-L), optimization of features layer (OF-L). In the visual features layer (VF-L), the RGB dermoscopic image is transformed to perceptual-oriented CIEL*a*b* color space. Then color coherence features are derived from each image in CIEL*a*b* color space and the bag-of-words (BoW) is then constructed using the simple k-means clustering algorithm based on SURF features to best describe the visual features. In the VF-L, the texture features are also extracted using stack-based-AutoEncoders (SAE). Afterward, these color and texture features are combined and normalized in the visual descriptor layer (VD-L) by using PCA technique. In the fusion of features layer (FF-L), the recurrent neural networks (RNNs) model is used to select most discriminative features that can provide better classification results. Finally, the Softmax linear classifier is used in a supervised manner to classify malignant versus nevus skin lesions in a supervised fashion with already known class labels (Y).

Despite the popularity of the deep learning algorithm, it requires a large amount of data set for training dataset. If you do not have thousands of example the deep learning is unlikely to outperform other approaches. It is extremely computationally expensive to train. The most complex models take weeks to train using hundreds of machines equipped with expensive GPUs. Deciding the positive/negative labels for automated melanoma research is far from obvious, and at the very least, requires careful documentation. It also observed that the past models (global descriptors, BoVWs) can no longer compete with deep learning, even in the context of small-dataset medical applications. If some model can beat deep learning algorithms, it is a model for the future, not from the past. In this paper, the current research efforts are focused on using the variant of deep learning algorithms for automatic diagnosis based on dermoscopic images.

The results of DermoDeep compared to other systems indicate that many melanomas versus nevus skin lesions are correctly classified. In practice, the dermoscopic diagnosis is very complex and tends to be subjective. Automatic diagnosis of pigmented skin lesion based on the clinical history context and visual structures can give some quantitative measure of images, be more accurate and objective. This paper sums up the results obtained with a use of proposed DermoDeep and compare with other deep learning classifiers. The experiments confirmed the correct selection of classifiers which was based on the input data characteristics. Taking into account an insufficient number of the positive cases (with melanoma), applying the deep neural network raised up its sensitivity to 93% and gave AUC of 0.95 ($P < 0.001$).

Compare to these results, there are some dermoscopic images that are not correctly classified by the proposed DermoDeep system. For example, it is difficult to classify PSLs when the skin lesions contained hairs or dermoscopic gel pixels. If the lesion contained these artifacts then it would definitely effect on classification accuracy. To overcome this issue, a research will be required to conduct. Also, there are other color spaces such as CIECAM02 color appearance models that better represent the color content of the images. As future work, the extraction of visual features and the multilayer deep learning algorithms will be implemented in C++ and further, the speed of the system can be enhanced by introducing parallelism through GPU-based and parallel-based libraries such that it can take advantage of hardware with the massive level of parallelism.

Moreover, the future works intend to explore more in-depth the issue of dataset organization for experimental validity. To get more comprehensive training data set for deep learning algorithm, the communication between the Computer Science and the Medical Science College are already started in order to better understanding computerized screening system. Many challenges lie ahead of automated melanoma screening as a practical tool, beyond the issues of machine learning: for example, usability and privacy. As computer vision models improve fast, the community expects to move to those (harder) issues.

Regarding the issue of image databases availability, the private and public data sources have utilized for training and testing the proposed automatic DermoDeep system. In future works, a free access of PSLs database will be addressed so that other researchers used it to their automatic systems.

The achieved results confirm that dermoscopic structures can enhance accurate melanoma recognition. It would be of great assistance if some global characteristics for such a lesions were automatically recognized. The image processing and feature extraction procedures could be applied to identify visual patterns perceived by experienced clinicians. The visual similarity among skin disorders has to be also investigated.

5 Conclusion

In this paper, a novel system DermoDeep system is presented to discriminate between melanomas and nevi in a perceptual-oriented color space. In the proposed system, the visual features and five layers including multilayers of deep learning neural network model are integrated to obtain significant results. The primary aim of this paper is to focus skin lesion classification without performing pre- and post-processing steps. To the best knowledge, although there are a number of studies devoted towards classification of pigmented skin lesions (PSLs), none of them focused on effectively defining the multi-feature fusion approach and importance of deep learning algorithms. Moreover, existing systems were dedicated to the detection of skin lesions, feature extraction and classification based on the so obtained regions' parameters. However, it is very

difficult and time-consuming to accurately extract all of the features with conventional image processing techniques due to the diversity of possible appearances.

In contrast to state-of-the-art approaches, a use of a powerful classifier is focused along with visual features extracted at several points of the images. Therefore, there is no need to identify characteristic elements, such as lesion patterns on the images saving time and reducing error propagation. In this DermoDeep system, adaptability and accuracy are achieved with an implementation based on visual features and a tri-level optimization technique on the deep learning neural network method.

The overall performance of the proposed algorithm was measured in terms of effectiveness and time. A sensitivity (SN) of 93%, a specificity (SP) of 95% and an area under the receiver operating curve (AUC) of 0.96 were obtained on a large and diverse set of images. The proposed DermoDeep system has achieved significantly higher classification rate with less computational time than other state-of-the-art methods. Hence, it could be used to provide second opinion to a dermatologist. This work can be extended to easily use in any domain of images, like Satellite images, Industrial images, CT images, MRI images, etc. and these predictions which are estimated only on the features of the images.

Acknowledgements This study was funded by Al Imam Mohammad Ibn Saud Islamic University (IMSIU) (grant number 360905).

Compliance with ethical standards

Conflict of interest Authors declare that they do not have conflict of interest.

References

1. Abbas Q et al (2013) Pattern classification of dermoscopy images: a perceptually uniform model. *Pattern Recogn* 46(1):86–97
2. Abbas Q et al (2013) Melanoma recognition framework based on expert definition of ABCD for dermoscopic images. *Skin Res Technol* 19:93–102
3. Abbas Q, Sadaf M, Akram A (2016) Prediction of dermoscopy patterns for recognition of both melanocytic and non-melanocytic skin lesions. *Computers* 5(13):1–16
4. Aghbari ZA, Al-Haj R (2006) Hill-manipulation: an effective algorithm for color image segmentation. *Image Vis Comput* 24:894–903
5. Argenziano G et al (2000) Interactive atlas of dermoscopy CD. EDRA medical publishing and new media. Milan, Italy
6. Barata C, Marques J, Rozeira J (2012) A system for the detection of pigment network in Dermoscopy images using directional filters. *IEEE Transactions on Biomedical Imaging* 59(10):2744–2754
7. Barata C et al (2014) Two systems for the detection of melanomas in dermoscopy images using texture and color features. *IEEE Syst J* 8:965–979
8. Barata C et al (2016) Clinically inspired analysis of dermoscopy images using a generative model. *Computer Vision and Image Understanding* 151: 124–137.
9. Barata C, Celebi ME, Marques JS (2017) Development of a clinically oriented system for melanoma diagnosis. *Pattern Recogn* 69:270–285
10. Bay H et al (2006) SURF: speeded up robust features. *Computer Vision, Lecture Notes in Computer Science* 3951:404–417
11. Bengio Y (2009) Learning deep architectures for AI found. *Trends Mach Learn*:1–127
12. Blum A et al (2004) Digital image analysis for diagnosis of cutaneous melanoma. Development of a highly effective computer algorithm based on analysis of 837 melanocytic lesions. *Br J Dermatol* 151:1029–1038

13. Bradley AP (1997) The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recogn* 30(7):1145–1159
14. Celebi ME, Zornberg A (2014) Automated quantification of clinically significant colors in Dermoscopy images and its application to skin lesion classification. *IEEE Syst J* 8(3):980–984
15. Celebi ME et al (2007) A methodological approach to the classification of dermoscopy images. *Comput Med Imaging Graph* 31:362–373
16. Celebi ME, Kingravi H, Vela PA (2013) A comparative study of efficient initialization methods for the K-means clustering algorithm. *Expert Syst Appl* 40(1):200–210
17. Codella N et al (2015) Deep learning, sparse coding, and SVM for melanoma recognition in dermoscopy images. Machine learning in medical imaging. In: *Proceeding of the 6th international workshop, MLMI 2015, Munich, Germany, lecture notes in computer science, springer, Berlin/Heidelberg, Germany, vol 9352, pp 118–126*
18. Demyanov S et al (2016) Classification of dermoscopy patterns using deep convolutional neural networks. *Proceeding of 13th IEEE International Symposium on Biomedical Imaging (ISBI), San Francisco, Canada:364–366*
19. Deng L (2014) A tutorial survey of architectures, algorithms, and applications for deep learning. *APSIPA Trans Signal Inf Process*:1–29
20. Dreiseitl S, Binder M (2005) Do physicians value decision support? A look at the effect of decision support systems on physician opinion. *Artif Intell Med* 33(1):25–30
21. Fornaciari M et al (2016) Towards automated melanoma screening: proper computer vision & reliable results. <https://arxiv.org/abs/1604.04024>. Accessed 5 April 2017
22. Gamavi R, Aldeen M, Bailey J (2012) Computer-aided diagnosis of melanoma using border- and wavelet-based texture analysis. *IEEE Trans Inf Technol Biomed* 16(6):1239–1252
23. GravesA MAR, Hinton G (2013) Speech recognition with deep recurrent neural networks. *Proceeding of IEEE International Conference on Acoustics, Speech and Signal Processing, Vancouver, BC:6645–6649*
24. Hinton G et al (2012) Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups. *IEEE Signal Process Mag* 29(6):82–97
25. Hu H, Li Y, Liu M, Liang W (2014) Classification of defects in steel strip surface based on multiclass support vector machine. *Multimed Tools Appl* 69:199–216
26. Hu H, Liu Y, Liu M, Nie L (2016) Surface defect classification in large-scale strip steel image collection via hybrid chromosome genetic algorithm. *Neurocomputing* 181:86–95
27. International Skin Imaging Collaboration (2017) ISIC 2017: skin lesion analysis towards melanoma detection. *Computer vision and pattern recognition*, <http://isdis.net/isic-project/>. Accessed 1 January 2017
28. Ishihara Y et al (2006) Early acral melanoma in situ: correlation between the parallel ridge pattern on dermoscopy and microscopic features. *Am J Dermatopathol* 28:21–27
29. Iyatomi H et al (2008) Computer-based classification of dermoscopy images of melanocytic lesions on acral volar skin. *J Investig Dermatol* 128: 2049–2054.
30. Iyatomi H et al (2008) An improved internet-based melanoma screening system with dermatologist-like tumor area extraction algorithm. *Comput Med Imaging Graph* 32(7):566–579
31. Jaworek-Korjakowska J, Kleczek P (2016) Automatic classification of specific melanocytic lesions. *Biomed Res Int* 2016:1–17
32. Kawahara J, BenTaieb A, Hamarneh G (2016) Deep features to classify skin lesions. *Proceeding of 13th IEEE International Symposium on Biomedical Imaging (ISBI), San Francisco, Canada:1397–1400*
33. Keren G, Schuller B (2016) Convolutional RNN: an enhanced model for extracting features from sequential data. In *Neural Networks (IJCNN), the IEEE 2016 International Joint Conference:3412–3419*
34. Laura K et al (2015) Computer-aided classification of melanocytic lesions using dermoscopic images. *J Am Acad Dermatol* 73(5):769–776
35. LeCun Y, Bengio Y, Hinton G (2015) Deep learning. *Nature* 521:436–444
36. Lie M, Zhang L, Hu H, Nie Liqiang DJ (2016) A classification model for semantic entailment recognition with feature combination. *Neurocomputing* 208:127–135
37. Lingala M et al (2014) Fuzzy logic color detection: blue areas in melanoma dermoscopy images. *Comput Med Imaging Graph* 38:403–410
38. Liu Z et al (2012) Distribution quantification on dermoscopy images for computer-assisted diagnosis of cutaneous melanomas. *Medical & Biological Engineering & Computing* 50(5):503–513
39. Liu W, Mei T, Zhang Y, Che C, Luo J (2015) Multi-task deep visual-semantic embedding for video thumbnail selection. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA 2015:3707–3715*
40. Liu L, Wiliem A, Chen S, Lovell BC (2017) What is the best way for extracting meaningful attributes from pictures? *Pattern Recogn* 64:314–326
41. Liu L, Nie F, Wiliem A, Li Z, Zhang T, Lovell BC (2018) Multi-modal joint clustering with application for unsupervised attribute discovery. *IEEE Trans Image Process* 27(9):4345–4356

42. Liu D, Liu L, Tie Y, Qi L (2018) Multi-task image set classification via joint representation with class-level sparsity and intra-task low rankness. *Pattern Recognition Letter* in press
43. Luo MR, Cui G, Rigg B (2001) The development of the CIE 2000 colour-difference formula: CIEDE2000. *Color Research & Application* 26(5):340–350
44. Ma Z, Tavares JMRS (2017) Effective features to classify skin lesions in dermoscopic images. *Expert Syst Appl* 84:92–101
45. Ma Z, Tavares JMRS (2017) Effective features to classify skin lesions in dermoscopic images. *Expert Syst Appl* 84:92–101
46. McDonald R, Smith KJ (1995) CIE94 - anew colour-difference formula. *J Soc Dyers* 111:376–379
47. Md F (2005) Color appearance models, vol 82. Wiley-IS&T
48. Melgosa M, Huertas R, Berns RS (2004) Relative significance of the terms in the CIEDE2000 and CIE94 color difference formulas. *J OpImage Sci Vis* 21(12):2269–2275
49. Mendonca T et al (2013) PH² - a dermoscopic image database for research and benchmarking. *Proceeding of IEEE conference on Eng Med BiolSoc, Piscataway NJ USA*, In, pp 5437–5440
50. Mirzaalian H, Lee TK, Hamarneh G (2016) Skin lesion tracking using structured graphical models. *Med Image Anal* 27:84–92
51. Mishra NK, Celebi ME (2016) An Overview of Melanoma Detection in Dermoscopy Images Using Image Processing and Machine Learning. <https://arxiv.org/abs/1601.07843>. Accessed 6 August 2018
52. Nie L, Zhang L, Yan Y et al (2017) Multiview physician-specific attributes fusion for health seeking. *IEEE Transactions on cybernetics* 47(11):3680–3691
53. Pathan S, Prabhu KG, Siddalingaswamy PC (2018) Techniques and algorithms for computer aided diagnosis of pigmented skin lesions—a review. *Biomedical Signal Processing and Control* 39:237–262
54. Premaladha J, Ravichandran KS (2016) Novel approaches for diagnosing melanoma skin lesions through supervised and deep learning algorithms. *J Med Syst* 40(96):1–12
55. Roberta B et al (2016) A computational approach for detecting pigmented skin lesions in macroscopic images. *Expert Syst Appl* 61:53–63
56. Rosendahl C et al (2011) Diagnostic accuracy of dermatoscopy for melanocytic and nonmelanocytic pigmented lesions. *J Am Acad Dermatol* 64:1068–1073
57. Ruiz D et al (2011) A decision support system for the diagnosis of melanoma: a comparative approach. *Expert Syst Appl* 38(12):15217–15223
58. Sadeghi M et al (2013) Detection and analysis of irregular streaks in dermoscopic images of skin lesions. *IEEE Trans Med Imaging* 32:849–861
59. Sadri AR et al (2017) WN based approach to melanoma diagnosis from Dermoscopy images. *IET Image Process* 11(7):475–482
60. Sáez A, Serrano C, Acha B (2014) Model-based classification methods of global patterns in dermoscopic images. *IEEE Trans Med Imaging* 33(5):1137–1147
61. Schanda J (2007) International commission on illumination. *Colorimetry: understanding the CIE system*. John Wiley Sons, New Jersey
62. Sharma G, Wu W, Dalal EN (2005) The CIEDE2000 color-difference formula: implementation notes, supplementary test data, and mathematical observations. *Color Res Appl* 30(1):21–30
63. Shimizu K (2015) Four-class classification of skin lesions with task decomposition strategy. *IEEE Trans Biomed Eng* 62(1):274–283
64. Stoecker WV et al (2011) Detection of granularity in dermoscopy images of malignant melanoma using color and texture features. *Comput Med Imaging Graph* 35:144–147
65. Thomas S et al (2013) Deep neural network features and semi-supervised training for low resource speech recognition. In: *proceeding of IEEE International Conference on Acoustics, Speech and Signal Processing, Vancouver, BC*, pp. 6704–6708.
66. University of Auckland, (2016) New Zealand. *Dermatologic image, database*. <http://dermnetnz.org/doctors/dermoscopy-course/>. Accessed 15 March 2016
67. Unlu E, Akay BN, Erdem C (2014) Comparison of dermatoscopic diagnostic algorithms based on calculation: the ABCD rule of dermatoscopy, the seven-point checklist, the three-point checklist and the CASH algorithm in dermatoscopic evaluation of melanocytic lesions. *J Dermatol* 41(7):598–603
68. Wang B, Qi G, Tang S, Zhang L, Deng L, Zhang Y (2018) Automated pulmonary nodule detection: high sensitivity with few candidates. *MICCAI* 2:759–767
69. Weifeng L, Xinghao Y, Dapeng T, Jun C, Yuanyan T (2018) Multiview dimension reduction via hessian multitset canonical correlations. *Information Fusion* 41:119–128
70. Ye L, Nie L, Han L, Zhang L, Rosenblum DS (2015) Action2Activity: Recognizing Complex Activities from Sensor Data. *IJCAI'15 Proceedings of the 24th International Conference on Artificial Intelligence Pages* 1617–1623, Buenos Aires, Argentina — July 25–31.

71. Yoshida T et al (2016) Simple and effective pre-processing for automated melanoma discrimination based on cytological findings. In: Proceedings of the 2016 IEEE international conference on big data, pp 3439–3442
72. Zortea M et al (2014) Performance of a dermoscopy-based computer vision system for the diagnosis of pigmented skin lesions compared with visual evaluation by experienced dermatologists. *Artif Intell Med* 60:13–26

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



Abbas Qaisar received Doctor of engineering degree (PhD) from the University of HUST at (Wuhan, China) in 2011. He is currently working as an Assistant Professor in the Department of Computer Science at the Al-Imam Ibn Saud Islamic University. He has published more than 40 research papers in both reputable journals and conferences. His research interests include: image processing, computer vision and artificial intelligence.



M. Emre Celebi received his BSc degree in computer engineering from Middle East Technical University (Ankara, Turkey) in 2002. He received his MSc and PhD degrees in computer science and engineering from the University of Texas at Arlington (Arlington, TX, USA) in 2003 and 2006, respectively. He is currently an associate professor in the Department of Computer Science at Louisiana State University, Shreveport, LA, USA). His research interests include medical image analysis, color image processing, content-based image retrieval, and open-source software development.