Multimedia Tools and Applications (2020) 79:20483–20518
https://doi.org/10.1007/s11042-019-07988-1

Computer-aided diagnosis (CAD) system based on multi-layer feature fusion network for skin lesion recognition in dermoscopy images

Ibtissam Bakkouri[1] · Karim Afdel[1]

Received: 29 November 2018 / Revised: 21 June 2019 / Accepted: 10 July 2019 /
Published online: 26 July 2019

Abstract

Skin lesion recognition is one of the most important tasks in dermoscopic image analysis. Current Convolutional Neural Network (CNN) algorithms based recognition methods tend to become a standard methodology to fix a large array of Computer-Aided Diagnosis (CAD) and interpretation problems. Besides significant practical and theoretical improvements in their architecture, their effectiveness is built on the existence of the flexible pre-trained models which generalize well to novel tasks and handle the problem of having small set of dermoscopic data. However, existing works pay little attention to exploring the benefits of hierarchical multi-feature fusion for classifying the skin lesions in digital dermoscopic images. Practically, it has been found that integrating multi-layer features has significant potential for improving performance of any pattern recognition task. In this paper, we developed a robust CAD system based on transfer learning and multi-layer feature fusion network to diagnose complex skin diseases. It is a convenient approach in terms of overfitting prevention, convergence speed and high morphological feature similarity processing. Our research focuses exclusively on obtaining optimal performance with addressing the various gaps in the skin pattern recognition area. For validation and comparison purposes, the proposed approach was evaluated on publicly dermoscopic dataset, and achieved the high recognition precision compared with fully trained CNN models, fine-tuning process, single CNN model and other related works. Therefore, the study demonstrates that our proposed approach can dramatically improve the performance of CAD systems which are based on the conventional recognition and classification algorithms for skin lesion recognition in dermoscopic data.

Keywords Skin diseases · Dermoscopic pattern recognition · Transfer learning · Multi-layer feature fusion · Multi-class classification

✉ Ibtissam Bakkouri
ibtissam.bakkouri@gmail.com

[1] LabSIV, Department of Computer Science, Faculty of Science, Ibn Zohr University,
BP 8106, 80000 Agadir, Morocco

1 Introduction

Skin cancer is the most common and prevalent type of cancer over the world [49]. It has risen dramatically during the last four decades due to weakened immune system, accumulated exposure to ultraviolet radiation, harmful chemical elements and sunlight during the winter months. More than two million cases of melanoma and basal cell carcinoma are diagnosed every year [4]. This is over the combined counts of breast cancer, lung cancer and colon cancers [90]. Owing to increase of incidence of the disease steadily in around the world, it has been one of the main health problems. Therefore, early automated diagnosis of skin lesions using dermoscopic digital images is a challenging multi-class classification task because of complex morphological structures of some skin patterns. It is very important in terms of reducing the number of patient deaths. Nevertheless, early diagnosis requires an accurate and a reliable diagnosis procedure that allows dermatologist to distinguish benign skin lesions from malignant ones. The suspected skin tissues usually have morphology characteristics such as asymmetrical characteristic, irregular edge of the lesion, different color composition and a large diameter [61]. It has been observed that geometrical features (shape and margin) and heterogeneous texture can be effectively used for classifying the skin structures as Melanoma (MEL), Melanocytic Nevus (NV), Basal Cell Carcinoma (BCC), Actinic Keratosis (AKIEC), Benign Keratosis (BKL), Dermatofibroma (DF), or Vascular Lesion (VASC). Visual screening during the diagnosis and analysis of skin patterns suffers from the resemblance among the abnormal and normal skin tissues, which can result in incorrect interpretations. Hence, one of the most significant challenges related to biomedical engineering is to recognize and classify a skin lesion efficiently using dermoscopic images. In the past decade, dermoscopy, which is a non-invasive imaging device that evaluates pigmented skin lesions by acquiring the magnified images of these skin lesions using polarized light waves, has been used to assist dermatologists in boosting the examination for suspected skin lesions. It displays deeper details of the skin tissue by removing the reflection of the normal skin surface [80]. Although dermoscopy enhances the diagnostic performance of skin lesion recognition compared to that carried out by visual screening, examination through dermoscopy images by dermatologists is subjective and requires additional time and manual effort which can lead to inaccurate decision-making [51]. Therefore, automated computerized diagnostic systems for skin lesion recognition and analysis are highly requisite to aid and assist dermatologists in providing effective diagnosis and fast interpretation.

Finding a robust and efficient feature representation space without risking to lose useful information is one of the key aspect in inter-class separation, and it heavily brought significant improvement in several domains [24, 85–87]. Recently, CAD system, which is developed using deep CNN algorithms and data representation techniques [6], has become one of the major research subjects in medical imaging and diagnostic radiology. It is widely considered as one of the most effective process to assist the physicians in order to recognize the suspected tissues at early stages. In fact, it is greatly used as a part of the routine clinical work for recognition of different types of cancers on medical images at many screening sites and hospitals. These successes of CAD system are benefited from using CNN architectures which are becoming the state-of-the-art method for several computer vision problems including image classification [1], object localization [55] and image segmentation [38]. They are able of generating discriminative features from raw data to form gradually abstract descriptors, replacing the traditional approach of carefully hand-crafting features and algorithms. Despite their high performance achievement, deep CNN methods are still limited. They usually necessitate datasets comprising tens of thousands of images. However, when dealing with dermoscopic images, we notice that they suffer from small number of available samples to build robust deep CNN models. Thus, a common problem that can arise during the training of a deep CNN is overfitting, which occurs when a network customizes itself too much to describe the relation between training data and the labels, and as a consequence, it fits too well to this training set. In this case, the model is able to memorize the training examples, but leads to poor performance on new ones that it has not already observed. Current deep CNN based data augmentation methods [30] for skin patterns recognition need some adjustment and tuning a large number of different parameters to achieve satisfactory performance. As a consequence, the main drawback of data augmentation approach is that it requires extremely long training time and high computational costs [66, 78].

To overcome these limitations pointed out above, we took advantage of transfer learning where the network weights are initialized by training with non-medical images and adapted on dermoscopic dataset. It is the common solution for using deep CNN on small dataset which helps the network optimization process, resulting in faster convergence of the training [72]. In general, CNN provides three levels of abstraction in the feature hierarchies [88]: Low, medium and high levels. It was noticed that the earlier convolutional layers design low-level features which are more generic features [67]. They try to visualize the difference between the colors in foreground and background, including boundary of objects [73], textual difference [71] and structure descriptor [36], which can be useful to many tasks and domains [2]. By contrast, the medium and high layers [62, 63] become gradually more specific to the details of the patterns contained in the target dataset [34], and they generate the high-level features. These high-level features are able to capture the structured information, less fine-grained spatial details and semantic context in the image [14]. Therefore, inspired by fine-tuning technique, the significant advantages have become apparent when copying the first convolutional layers from a pre-trained network to the target network, while other layers are fine-tuned towards the dermoscopic dataset with completely keeping a balance between high performance achievement level and low computational times, efforts and complexities [76]. However, this immediately raises some interesting questions. Which layers are helpful to design low-level features of skin lesions? How many layers to copy and to fine-tune?

In this paper, we present an empirical investigation of these questions. One of the contributions of our work is to explore the application of CNN based transfer learning technique in multi-class classification task for skin abnormality recognition. In fact, the abnormal skin tissues characterized by complex morphological structures like shape with convex outside borders and undefined edges, margin irregularity and texture heterogeneity. In addition, the most of skin patterns suffer from semantic feature ambiguity and high inter-class visual similarities problems. To enable the feasibility of training deeper network with multi-class classification problem, we adopt convolutional multi-layer feature fusion network which fuses the features extracted from different pre-trained models. This hierarchical fusion structure tends to perform well in maximally exploiting the input data and improving classification accuracy. Despite the popularity of CNNs, there has only been preliminary research on hierarchical multi-feature fusion network for medical image analysis. Thus, the using of CNN architectures to this issue for different computer vision techniques such as classification, detection and segmentation has yet to be fully investigated. Moreover, running multi-layer features of CNN on dermoscopic images requires more model build time and memory consuming. Then, how can we alleviate such weaknesses?

To explore gaps in the topics of fine-tuning method and hierarchical multi-feature fusion network, this paper makes the following major contributions for skin abnormality recognition:

1. It designs efficient low-level features which can strongly capture contrast and spatial information in the dermoscopic image by exploring state-of-the-art of three pre-trained CNN models: VGG-16 [74], ResNet-18 [39] and DenseNet-121 [41], and proposing different fine-tuning configurations. The best configurations have been selected and used as low-level feature extractors.
2. It introduces a new CNN paradigm to generate high-level features, which was called Convolutional Fusion Unit (CFU), based on merging of convolutional layers in order to extract the significant discriminative features to make our CAD for classifying dermoscopic skin abnormalities into MEL, NV, BCC, AKIEC, BKL, DF and VASC more robust and efficient.
3. It presents a simple but effective CFU configuration for designing three feature descriptors which denoted by VGG-CFU, ResNet-CFU and DenseNet-CFU in order to address the problem of high inter-class visual similarities between the classes with avoiding the high cost of memory accessibility, vanishing gradient, decreasing the bias shift, convergence to local minima and overfitting issue.
4. It proposes a discriminative global dermoscopic feature vector that represents semantic information benefited from multi-layer feature fusion network to combine the features from VGG-CFU, ResNet-CFU and DenseNet-CFU feature maps together, and as a consequence, to formulate a roust CAD system for skin abnormality recognition.
5. It suggests a network that effectively exploits fine-tuning technique and depth features simultaneously to obtain more discriminative global representation and estimate the probability of each skin pattern class. This network is trained to get optimal fusion of three complementary discriminative features through VGG-16, ResNet-18 and DenseNet-121 for low level features generation combined with CFU algorithms to extract high level features in order to ensure maximum information flow between layers in the network.

The proposed framework has been validated on the public Human Against Machine with 10000 training images (HAM10000) dataset [82], which is designed for skin lesion diagnosis, and has been compared to state-of-the-art skin pattern classification approaches. The performance evaluation analysis shows the potential clinical value of the proposed framework. To the best of our knowledge, no prior work exists on skin lesion characterization based on convolutional multi-layer feature fusion network and transfer learning that is investigated here.

The remainder of this paper is organized as follows: In Section 2, we describe the most important methods applied in dermoscopic images. In Section 3, we detail the proposed approach. Then, we present the results of experiments realized on HAM10000 dataset in Section 4 and we discuss in detail the results of our proposed method in Section 5. Finally, we conclude this paper in Section 6.

2 Related works

In related works, some approaches have been proposed using CNN algorithm, continuous fine-tuning technique and semantic multi-layer features which are directly related with our proposed method. The followings are the relative literature review in these three areas, and they are summarized in Tables 1 and 2.

2.1 CNN for skin disease recognition

Over the last few years, deep learning techniques have appeared as a strong alternative for supervised learning with high model performance and the ability to learn hierarchically discriminative features for the task at hand [94]. These features often outperform classical hand-crafted features and conventional descriptors [50, 81]. In the context of the dermoscopic images, deep CNN has begun to achieve excellent machine learning performance on patterns recognition, diagnosis and interpretation tasks. Currently, it attracts significant interest from many researchers in the field, and many studies have been developed to

Table 1 Overview of the related works focusing on CNN for skin disease recognition, continuous fine-tuning CNN and hierarchical multi-feature fusion

Main axes	Authors	Year	Methods
CNN for skin disease recognition	Thao et al. [79]	2017	VGG-16
	Menegola et al. [68]	2017	VGG-16
	Yu et al. [100]	2018	ResNet-50+SVM
	Zhang et al. [102]	2018	Inception-V3
	Dorj et al. [25]	2018	AlexNet+ECOC SVM
	Han et al. [37]	2018	ResNet-152
Continuous fine-tuning CNN	Yosinski et al. [96]	2014	AlexNet
	Azizpour et al. [8]	2015	AlexNet
	Chu et al. [19]	2016	AlexNet
	Zheng et al. [105]	2016	VGG-19+AlexNet
	Castrejon et al. [15]	2016	AlexNet
	Wei et al. [89]	2016	DeCAF
	Akcay et al. [3]	2016	AlexNet
	Chougrad et al. [18]	2018	VGG-16+ResNet-50+Inception-V3
	Li et al. [56]	2018	AlexNet+VGG-16+Inception-V3
	Bakkouri et al. [10]	2018	AlexNet
Hierarchical multi-feature fusion	Dai et al. [23]	2017	FASON
	Li et al. [53]	2017	MIFK+PCA+SRKDA
	Yu et al. [99]	2017	MLFF
	Gogate et al. [31]	2017	DLDMF
	Wu et al. [91]	2017	MFE
	Ye et al. [95]	2017	3DCNN+GMU
	Golrizkhatami et al. [32]	2018	TLF
	Banerjee et al. [13]	2018	TL+MF
	Huo et al. [42]	2018	FFE
	Vu et al. [84]	2018	3DVGG-16
	Kuai et al. [48]	2018	SiamFC
	Liu et al. [59]	2018	MMC-3DCNN

Table 2 Summary of related works classified by application domain

Main axes	Application area
CNN for skin disease recognition	Skin lesion recognition [25, 37, 102]
	Melanoma recognition [68, 79, 100]
Continuous fine-tuning CNN	Visual recognition [8, 19, 96, 105]
	Image retrieval [15, 89, 105]
	X-ray baggage classification [3]
	Mammographic lesion classification [18]
	Endomicroscopic image classification [56]
	Melanoma recognition [10]
Hierarchical multi-feature fusion	Visual recognition [23, 42]
	Visual tracking [48]
	Remote sensing classification [53]
	Image retrieval [99]
	Deception detection [31, 91]
	Electrocardiogram classification [32]
	Rhabdomysarcoma classification [13]
	Glioma classification [95]
	Alzheimer's disease diagnosis [59, 84]

perform skin lesion analysis and help dermatologists give an accurate diagnosis. Inspired by the layers in the deep networks [39], authors of [100] applied very deep residual neural network (ResNet-50) to extract discriminative features and support vector machine (SVM) with a Chi-squared kernel to classify melanoma images. This method was evaluated on the International Skin Imaging Collaboration (ISIC) 2016 Skin lesion [35] challenge dataset, and it reached the accuracy of 86.81% and the Mean Average Precision (mAP) of 68.49%. Authors of [102] presented a computer aided diagnosis system to classify abnormal skin lesions in dermoscopic image into melanocytic nevus, seborrheic keratosis, basal cell carcinoma or psoriasis. The proposed method was evaluated on clinical database originated from the dermatology department of Peking Union Medical College Hospital, and the feature extraction and classification were carried out by GoogleNet Inception-V3 based on domain expert knowledge, achieving an accuracy of 87.25 ± 2.24% in the test dataset with 1067 images. In [25], an error-correcting output coding (ECOC) classifier combined with SVM for the multi-class classification scenario was introduced. The authors are proposed pre-trained AlexNet convolutional neural network model [29, 47] as feature extractor, and they obtained a classification rate of 92.3% for Actinic Keratoses, 91.8% for Basal cell carcinoma, 95.1% for Squamous cell carcinoma and 94.2% for Melanoma. In this context, the similar approach based on the transfer learning technique using ResNet-152 architecture was suggested by authors of [37] in order to classify 12 skin diseases: basal cell carcinoma, squamous cell carcinoma, intraepithelial carcinoma, actinic keratosis, seborrheic keratosis, malignant melanoma, melanocytic nevus, lentigo, pyogenic granuloma, hemangioma, dermatofibroma and wart. The proposed method was evaluated on the testing subset of the Asan, Hallym and Edinburgh datasets, and it reached promising results. VGG-16 architecture with fine-tuning technique was suggested to recognize melanoma abnormality in [79]. The proposed method was evaluated on ISIC 2017 challenge dataset [22, 83], yielding an average Receiver Operating Characteristic Area Under Curve (ROC-AUC) rate of 81.6%. A knowledge transfer based on VGG-16 model for melanoma recognition with deep learning was proposed by authors of [68]. The sources datasets employed for the transfer were selected from the retinopathy Kaggle challenge for diabetic retinopathy detection dataset and the ImageNet large scale visual recognition challenge 2012 dataset. The target images were obtained from the interactive atlas of dermoscopy and ISIC challenge 2016 dataset, and the proposed system provided promising results, achieving a ROC-AUC of 80.7% and 84.5% for the two skin-lesion datasets evaluated.

Despite the high achievement of deep CNN based on transfer learning for skin lesion diagnosis, it needs some improvement concerning the systematically analysis the factors that affect the transferability of CNN representation for skin lesion recognition tasks, and the evaluation the transferability of pre-trained weights from source domain to target one which reveals their generality or specificity without precisely harming the classification accuracy. The main weakness with all these previous studies is that they only fine-tuned the final fully connected layer and froze the previous deep layers. Current paradigms for recognition in the computer vision community involve fine-tuning a CNN by iteratively freezing the first layers and adjusting the last ones. However, research on this topic is still relatively scarce in dermoscopic domain.

2.2 Continuous fine-tuning CNN

Currently, deep CNNs have shown remarkable abilities in transferring knowledge between apparently different image classification tasks or even between imaging modalities for the same task. In most cases, this is done by fine-tuning technique which has been extensively studied over the past few years, especially in the field of computer vision, with several interesting findings. The authors of [96] investigated the effects of different fine-tuning procedures of AlexNet model on the transferability of knowledge, while a procedure is proposed to quantify the generality or specificity of a particular layer. These authors found that fine-tuning was the optimal technique, with performance slightly improving as more layers were copied. When layers were frozen, they saw performance degrade as more layers were copied. However, this work studied a target dataset that was virtually similar to the source dataset and had a considerably large number of images. The factors that influence the transferability of knowledge in the fine-tuning of AlexNet are presented by authors of [8]. These factors include the network's architecture, the resemblance between source and target tasks and the training framework. In a similar study, the authors of [19] evaluated the performance of fine-tuning methods by systematically explore the design-space for fine-tuning of AlexNet and give recommendations based on two key characteristics of the target dataset include visual distance from source dataset and the amount of available training data. This approach was validated on seven target datasets. Through their analysis, the authors found that as distance increases, fine-tuning improves relative to freezing, supporting the notion that learned features are less transferable to distant datasets. At one end, in the low data and low distance setting, freezing outperforms fine-tuning. The authors of [105] proposed pre-trained VGG-19 and AlexNet models to extract off-the-shelf CNN features for image search and classification. In this work, the authors evaluated the performance of features from different CNN layers. This idea was motivated by observing that features vary in their receptive field sizes and semantic levels. Their impact on recognition and search accuracy was measured by testing the seven layers on six datasets. The experimental results showed that features from the bottom layers are generally inferior to those from the top layers. This is expected because bottom-layer filters are sensitive to low-level visual patterns which are generally considered not discriminative enough. A cross-modal transfer learning was proposed by authors of [15, 89]. The first study is based on AlexNet pre-trained model while the second one is focused on the pre-trained Deep Convolutional Activation Features (DeCAF) architecture. They estimate that finding which and how many layers to transfer depends on the proximity of the two tasks and their modalities. It has been demonstrated that the earlier layers of the CNN model are specified for modality while the last layers of the CNN model are fixed for task. The authors of [3] performed fine-tuning approach based on AlexNet architecture to train over the X-ray baggage dataset. The results showed that true positives and true negatives have a general trend to decrease as the number of fine-tuned layers reduces. Likewise, freezing more layers lowers the accuracy of the models. A result can be reached from these practical experiments that freezing lower layers and fine-tuning higher achieve promising performance. A number of studies have also utilized transfer learning techniques in order to adapt well-known networks and classify medical images. In most of the cases, the network used is the AlexNet, VGG, ResNet or Inception pretrained on ImageNet. In [18], a CAD system for mammography mass lesion classification was proposed. The authors employed three pre-trained models: VGG-16, ResNet-50 and Inception-V3 with fine-tuning of only the last two convolutional blocks. The performance was validated on Breast Cancer Data Repository (BCDR), Mammographic Image Analysis Society (MIAS), Digital Database for Screening Mammography (DDSM) and INbreast databases. The authors achieved 97.35% accuracy and 0.98 ROC-AUC on the DDSM database, 95.50% accuracy and 0.97 ROC-AUC on the INbreast database, 96.67% accuracy and 0.96 ROC-AUC on the BCDR database, and 98.23% accuracy and 0.99 ROC-AUC on the MIAS dataset. In a similar study carried out by the authors of [56], a fine-tuned technique based on three pre-trained models, AlexNet, VGG-16 and Inception-V3, was investigated for brain tissue characterization for context aware diagnosis support in neurosurgical oncology. The authors observed that as they increase the depth of tuning layers, the features tend to be more low-level. The experimental results indicate that model tuning of medium depth can generate discriminative features and provide an efficient representation of Probe-based confocal laser endomicroscopy (pCLE) data. A convolutional neural-adaptive network for melanoma recognition was proposed by authors [10]. The authors tried to adapt the pre-trained low-level weights on dermoscopic dataset using AlexNet architecture and fine-tuning technique. The images were obtained from ISIC dataset, and the proposed system provided efficient results, achieving an average ROC-AUC rate of 96.66%.

Broadly, all these previous studies support the idea that the CNN models are based on a representation learning perspective. This view assumes that a CNN is built from layers that learn increasing levels of abstraction which form a hierarchy of concepts. They showed that the last layers of the network are discriminative feature specific while the earlier layers of the network are spatial information specific.

2.3 Hierarchical multi-feature fusion

In the past few years, a diversity of studies emphasizes the benefits which arise from multi-layer feature fusion of CNN. As an example, supplementary advantages have been reported for content-based image retrieval [98], pixel-level semantic image segmentation [26], pattern recognition [65] and real time object detection [20]. In computer vision and pattern recognition field, there are several researches that exploite CNN based multi-feature fusion. The authors of [48] exploited the combination of different layers of fully convolutional Siamese network, named SiamFC, to constitute the hyper-feature representations of the target and train the network in an end-to-end manner to implement efficient and effective tracking. The evaluation was performed on the OTB100 and TC128 benchmarks, achieving promising results compared with the state-of-the-art trackers. Deep Effective Fusion Architecture (FASON) for texture recognition was designed by authors of [23]. They effectively combined second order information, calculated from a bilinear model, and first order information, preserved through their leaking shortcut, in an end-to-end deep network in order to take advantage of the multiple features from different convolution layers. Additionally, the full advantage of both the global and local features in CNN was taken in [42]. The authors combined different levels of features extracted from different convolutional layers of the CNN architecture using Fused Feature Encoding (FFE) technique to explore richer semantic information within the image. The performance was evaluated on CIFAR-10, Caltech101 and NUS-WIDE databases for optimizing the efficiency of retrieving in high dimensional features. In [53], the authors presented a feature fusion framework based on Multiscale Improved Fisher Kernel (MIFK) to integrate multilayer features for remote sensing scene classification. The classification was carried out by a step-by-step dimensionality reduction based on Principal Component Analysis (PCA) and Spectral Regression Kernel Discriminant Analysis (SRKDA) strategies. The complementary strengths of Multi-Layer Feature Fusion (MLFF) for image retrieval were exploited by authors of [99]. They demonstrated their methods on three well-known benchmarks of image retrieval: Oxford 5K, Holiday and UKB, yielding competitive performance compared with the state-of-the-art methods. Not far away, deception detection task has been approached by Deep Learning Driven Multimodal Fusion (DLDMF) in [31]. The authors integrated audio, visual and textual cues to develop a fully automated multimodal deception detection approach. In a similar study carried out by authors of [91], three modalities was proposed like vision, audio and text to extract motion, Mel-Frequency Cepstral Coefficients (MFCCs) and transcript features, respectively, and improve performance of deception detection system by using Multimodal Feature Encoding (MFE) approach. As a key enabler to improve diagnosis, multi-layer feature fusion of CNN has revolutionized the medical imaging research area. Authors of [32] introduced the combination of hand-crafted and CNN-based features for ECG signal classification. They employed multi-stage learned features from different layers of a CNN for Three-Level Fusion (TLF) and they achieved superior classification results compared to other state-of-the-art methods. A radiomics framework based on Transfer Learning (TL) and Multimodal Fusion (MF) techniques that classifies alveolar from embryonal subtypes of rhabdomysarcoma was proposed in [13]. It was developed by exploiting a tight integration of advanced image processing methods and cutting-edge deep learning techniques. The classification rate was improved by performing a fusion of Diffusion-Weighted MR scans (DWI) and gadolinium chelate-enhanced T1-Weighted MR scans (MRI). The authors of [95] employed 3-Dimensional Convolutional Neural Network (3DCNN) combined with Gated Multimodal Unit (GMU) for glioma grading. This multimodal MRI images fusion framework was evaluated on BRATS datasets and achieved efficient results to distinguish benign gliomas and malignant ones. Moreover, diagnosis of Alzheimer's disease has been also approached in [84] where the combination of MRI and Fluorodeoxyglucose Positron-Emission Tomography (FDG-PET) scans has the potential to capture crucial local 3D amyloid uptake patterns without redundant, noisy information of white matter. The classification of Alzheimer's disease was carried out by an adaptive 3-Dimensional VGG-16 (3DVGG-16) architecture, providing satisfactory performance. Another recent issue to fence this section has been suggested by authors of [59] where Multi-Modality Cascaded 3-Dimensional Convolutional Neural Networks (MMC-3DCNN) for Alzheimer's disease diagnosis was presented. The authors presented a multi-modality classification algorithm based on hierarchical CNNs model to learn and combine the multi-level and multi-modality features of MRI and PET images for Alzheimer's disease diagnosis.

As a matter of fact, multi-layer feature fusion is a crucial factor to effectively explore the feature domain and enhance performance on a visual recognition task. However, existing works do not pay attention to exploring the benefits of multi-layer features for improving the pattern recognition in dermoscopic image. In view of the scarce existing studies on this domain, a detailed plan to overcome the problem of semantic feature ambiguity and high inter-class visual similarities between these seven classes: MEL, NV, BCC, AKIEC, BKL, DF and VASC will be described in a fair amount of detail.

3 Proposed method

Pattern recognition in dermoscopic images based CAD system is a challenging task in skin lesion diagnosis which plays a vital role to assist dermatologist for efficient diagnosis and fast interpretation. Until now, the existing CAD methods used to benefit a lot from a single feature map representation of dermoscopy image which always fails due to semantic feature ambiguity and high level similarities between classes for multi-class classification task. For this reason, they are not that good at the simulation of real skin abnormality diagnostic and interpretation processes in a large number of categories, adding difficulties in enhancing the recognition and classification performances. In this paper, we propose a novel CAD system based on multi-layer feature fusion of CNN. The suggested approach is efficient for recognition dermoscopic patterns in multi-class classification problem. Our proposed methodology based CAD system was developed in several stages. First, the representative square regions with size of 224 × 224 pixels were selected from the digitized dermoscopic images and enhanced by the pre-processing techniques. Inspired by transfer learning, to resolve the challenge of dealing with limited availability of the training dermoscopic samples, improve model building time and avoid overfitting, we designed efficient low-level features which can strongly capture contrast and spatial information in the dermoscopic image by exploring state-of-the-art of three pre-trained CNN models: VGG-16, ResNet-18 and DenseNet-121. To generate high-level features, we introduced a new CFU paradigm based on merging of convolutional layers in order to extract the significant discriminative features. Then, we built three feature descriptors which denoted by VGG-CFU, ResNet-CFU and DenseNet-CFU. The most significant extracted features benefited from convolutional multi-layer feature combination through fully connected fusion (FCF1, FCF2 and FCF3) layers. Finally, the dermoscopic patterns were efficiently recognized and classified by applying the multi-level descriptors on the fused feature maps. An overview of the proposed approach is given in Fig. 1.

3.1 Pre-processing of dermoscopic data

Pre-processing of the digital dermoscopic data is performed in four stages, including data size normalization technique, class balancing process, skin hair removal and intensity correction to facilitate application of the convolutional multi-layer feature fusion algorithm.

3.1.1 Image size normalization

As it can be seen in Fig. 2, the dermoscopic HAM10000 dataset contains images of high resolutions. The resolutions of all dermoscopic images are 600 × 450 pixels, which require

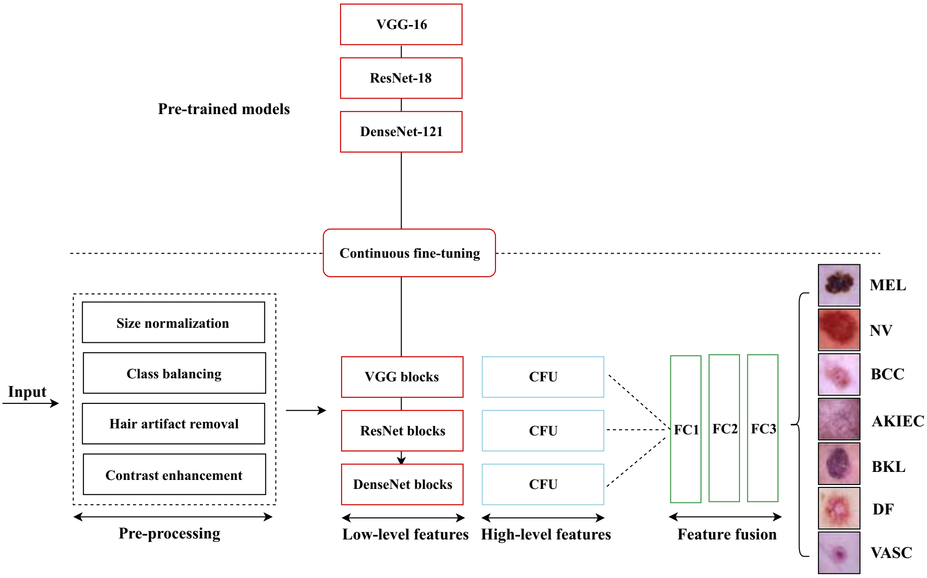

Fig. 1 Overview of the integrated workflow. The outline of our CAD system is composed of four stages: pre-processing, fine-tuning, CFU building, and hierarchical multi-feature fusion. The goal of the pre-processing stage is to facilitate the training of deep neural networks on dermoscopic data. The fine-tuning stage aims to select the first n frozen blocks to design low-level features with keeping the balance of high performance level and low computational cost. CFU is implemented to capture high-level features. The hierarchical multi-feature fusion stage is designed to handle the problem of high morphological feature similarity between classes

high cost of computation. It is necessary to rescale the input images for deep learning network. As directly resizing image may distort the morphological features such as the shapes and forms of dermoscopic patterns, we first extracted the region of interest using square cropping from the center of original image and proportionally resized the area to an appropriate resolution. The square cropping is carried out using the determinant of the Hessian for multi-scale blob-like structure detection as [11] pointed out. The down-sampling operation is performed using pyramidal REDUCE decomposition as mentioned in [9]. This pyramid data structure operation involves filtering and down-sampling an image to get a new image at to lower resolution. Therefore, it produces a sequence of copies an original image in which both sample density and resolution are decreased in regular steps. Then, each level in

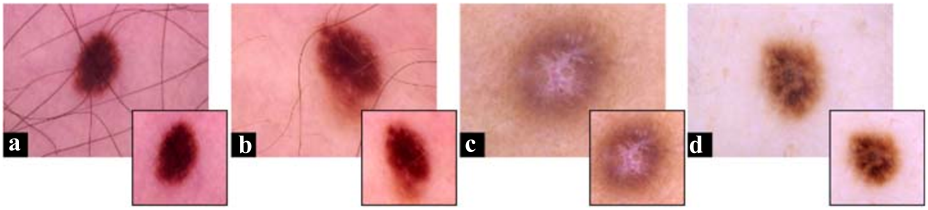

Fig. 2 Examples of four dermoscopic cases enhanced by the pre-processing algorithms. **a**-**b**-**d** represent square region extraction with size of 224 × 224 and hair artifact removal from melanocytic nevus structures with intensity correction. **c** represents constract enhancement of dermatofibroma tissue and pattern extraction centered in square window with down-sampling to 224 × 224 pixels

the pyramid has a smaller size than the previous level. The final image is down-sampled to 224 × 224 pixels.

3.1.2 Class balancing

As the dermoscopic image volumes of different categories vary widely, data augmentation processing is needed to balance the image volumes of different classes [93]. In this paper, we balance the training set by augmenting the data for the minority classes, applying geometric transformation techniques. A geometric operation transforms pixel (x_1,y_1) value in an input image to another location (x_2,y_2) in an output image. Three processing techniques were adopted: rotation, translation and reflection [11]. These transformations can be described by the first order polynomial which is defined as follows:

$$\begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = R \times \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} + T \tag{1}$$

Translation can be calculated by specifying values for the T matrix, while rotation and reflection are defined by assigning values to variables in the R matrix. Each training image is randomly rotated by 10 selected angles θ in the interval $0 \leq \theta \leq 360$. For each rotation, it is mirrored and translated by four values of horizontal and vertical pixel displacements. Finally, after data preparation process, the number of dermoscopic images is expanded at least by 20 times. The experiments show that our proposed framework is invariant to these transformations.

3.1.3 Hair artifact removal and contrast enhancement

Currently, the CAD system, which is developed using image processing and deep CNN algorithms, is widely considered as one of the most effective process to assist the dermatologists in order to recognize the skin lesion at early stages. It plays an important role in the improvement of visual quality for computer vision and medical pattern recognition. However, dermoscopic image is usually characterized by hair-like regions within skin lesions. In addition of noising by hair artifact, the most of these images are suffer from low tissue brightness and background luminance which can result from many situations, including incompetence of operator expertise and inefficiency of the image capture device. The enhancing methodology, which is used, should have the efficiencies in both skin hair inpainting and contrast enhancement. To reduce the effects of hair artifact, we employed a computerized method described in [52]. It consists of three parts; the first deals with the identification of hair location using morphological closing operation, while the second part concerns the replacement of identified hair pixels with neighboring non-hair pixels and finally, the smoothing of the final result with an adaptive median filter. To boost image quality for better human visual perception, a technique must be applied which creates a balance between high levels of visual quality and low computational complexity. It is based on local adaptive bi-histogram equalization [77] to preserve and enhance dermoscopic image brightness.

3.2 Low-level feature designing

As it was previously mentioned, feature construction is carried out in two main stages: low and high level feature extraction. In this section, we are interested in designing a novel architecture that uses an existing pre-trained CNN models to generate the low-level features.

Inspired by transfer learning, we are concerned with evaluating our continuous fine-tuning system which aims to resolve the challenge of dealing with limited data and reduce the time required for training a CNN by gradually adapting pre-trained weights and excluding them from the back-propagation process. This proposition is motivated by the observation that the earlier convolutional layers, which are the low layers, design more generic features such as the orientation of edges, the color blotches and the simple blob-like image structures, which can be useful to many tasks and domains. By contrast, the deeper layers, which include the medium and high layers [62, 63], become gradually more specific to the details of the patterns contained in the dermoscopic dataset which tend to be readily recognizable [43]. Besides being able to capture universal features, some researchers noticed the fact that in most of the deep architectures, the first few convolutional layers require most of the efforts, time and a very large amount of the training data to start build a model [40, 104]. To support this explanation, three models trained on ImageNet dataset using VGG-16 [74], ResNet-18 [39] and DenseNet-121 [41] architectures have been proposed as the source of low-level feature extractors. We chose these models due to their performance on a difficult 1000 non medical classes for classification task, considering that the activities of the neurons in their hidden earlier layers could work as low-level feature extractor for a variety of dermoscopic pattern recognition task. As [39, 41, 74] pointed out, the VGG-16 architecture consists of 16 learnable weight layers: 13 convolutional layers and 3 fully connected layers. The second architecture is ResNet-18 which is composed of 18 weights layers: 17 convolutional layers and one fully connected layer. And finally, the third architecture is DenseNet-121 includes 120 convolutional layers and only one fully connected layer. All the networks take 224×224 pixels RGB image as the input, use Linear Rectified Units (ReLU) as activation function, and are trained using stochastic gradient descent (SGD).

The objective of this continuous fine-tuning system is not only to search the accurate earlier convolutional layers which are capable to design the low-level features, but also to study the effectiveness of these three pre-trained models (VGG-16, ResNet-18 and DenseNet-121) and adjust the hyper-parameters for all these pre-trained networks to obtain the optimal performance in dermoscopic pattern recognition. More formally, we cluster the convolutional layers into 5 groups which are separated by pooling layers for VGG-16 and DenseNet-121, and residual block for ResNet-18. In the rest of our analysis, we denote the convolutional layer groups as block1, block2, block3, block4 and block5. Our proposed continuous fine-tuning method was evaluated in six iterations which denoted by $\mathrm{Level_n}$, where $\mathrm{n} \in \{0,1,2,3,4,5\}$. With fine-tuning, the first n blocks from a pre-trained network are copied to the target network and we leave them frozen, while other 5-n blocks are updated and trained towards the dermoscopic dataset. More specifically, as shown in Fig. 3, firstly, the entire model is trainable ($\mathrm{Level_0}$). At $\mathrm{Level_1}$, the first block is frozen, and the rest of the model is continued to train. At $\mathrm{Level_2}$, the first and second blocks are frozen, and so on. In other words, we transmit the frozen weights to lower blocks of target domain model and update the weights of the higher blocks of the network by continuing the back-propagation process. This process of continuous fine-tuning system was applied on each pre-trained model: VGG-16, ResNet-18 and DenseNet-121 which denoted by $\mathrm{VGG_n}$, $\mathrm{ResNet_n}$, $\mathrm{DenseNet_n}$, respectively.

We begin building our models to learn weights for dermoscopic domain with a classification function which is the softmax classifier addressed for minimizing error in this domain. Therefore, we replace the classification function and optimize the network again to minimize error in dermoscopic domain. Under this setting, we are adapting the weights of the network from the broad domain to the particular one. The softmax function in the original pre-trained models extracts the probability of 1000 categories of the ImageNet dataset. To

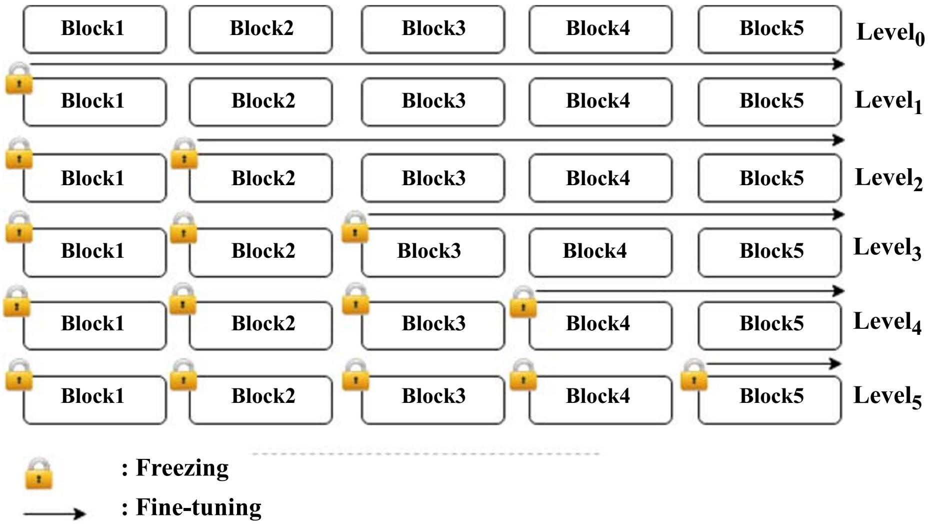

Fig. 3 The Flow chart of the frozen block searching strategy from three pre-trained models: VGG-16, ResNet-18 and DenseNet-121 to extract low-level features. This is motivated by the observation that the first blocks of pre-trained networks tends to have general features comprised of edges and blobs, while the last blocks are thought to be specific to the training data. Thus, at each Level$_n$, three pre-trained models (VGG$_n$, ResNet$_n$ and DenseNet$_n$) were evaluated using the most reliable evaluation metrics, where n is the number of first frozen blocks

start the continuous fine-tuning process, we replace this softmax function by a new one with random initialization values to compute the probability of 7 classes of the dermoscopic data. Precisely, the new replaced softmax layer employed the back-propagation algorithm to train their weights generated from scratch with data from the skin lesion recognition task, which has seven classes: MEL, NV, BCC, AKIEC, BKL, DF and VASC. In order to start the back-propagation algorithm for fine-tuning, we should bear in mind that choosing a convenient learning rate for SGD optimization is essential for effective learning. If the learning rate is too small, the optimization process through SGD is too far to sufficiently detect the parameter space and can highly diverge from the original values. On the other hand, a large value for learning rates can lead to exploding gradients which prevents the convergence during learning due to numerical issues. Based on this evaluation, it is mainly significant to set the learning rates of each step of Level$_n$ (freezing or fine-tuning) appropriately. The freezing action needs a very small constant learning rate because we want to keep the parameters of the pre-trained model to transfer that knowledge into the dermoscopic domain. However, notice that the learning rate is not set to zero in the freezing step; they will be optimized again at a slower pace. The fine-tuning stage needs a relatively variable small learning rate because during a weights updating point, the learning rate can be efficiently reduced after a few epochs without hinder convergence.

3.3 High-level feature designing

As was mentioned earlier, in this section, our study focuses exclusively on high-level feature generation. We are concerned with learning more complex discriminative representations from dermoscopic data using a novel CNN architecture called CFU. The proposed CFU gives the efficient performance in dermoscopic abnormality recognition task. In addition, the experiments, which will be shown and explained in a fair amount of detail in Section 4, show that the high-level discriminative features can be learned from CFU than from VGG-16, ResNet-18 and DenseNet-121. The proposed CFU takes advantage of using the Leaky Linear Rectified Units (LReLU) instead of ReLU for non-linearity activation function [64], Adaptive Moment Estimation (Adam) instead of SGD for network optimization [45], and robust fusion layer instead of conventional layer for discriminative feature designing [17]. It is not only can extract hidden features, but forms a hierarchical representation and a high-level abstraction of the dermoscopic abnormality by using a multi-layer neural network. They perform dramatically better than traditional CNN. Our proposed CFU consists of convolution layers, non-linear, local response normalization, fusion layers, pooling layers and concatenation layer (See Fig. 4).

In each convolutional layer (C), first, the output of previous layers is convolved with multiple learned weight matrices. Then, the result is processed by a non-linear operation to generate the layer output. The linear operation in the $(i+1)^{th}$ convolution layer, whose input is denoted by C_i (output of the i^{th} layer) comprises of a two-dimensional convolution as shown in (2):

$$C_{i+1}(m,n) = LReLU(C_i(m,n) \times W_i + b_i) \tag{2}$$

Where W_i denotes the learned weight matrix and b_i denotes bias or offset of the i^{th} convolution layer, and LReLU is an activation function. The LReLU function is as shown below; it gives an output a if a is positive and $0.01 \times a$ otherwise:

$$LReLU(a) = \begin{cases} a & a \geq 0 \\ 0.01 \times a & \text{otherwise} \end{cases} \tag{3}$$

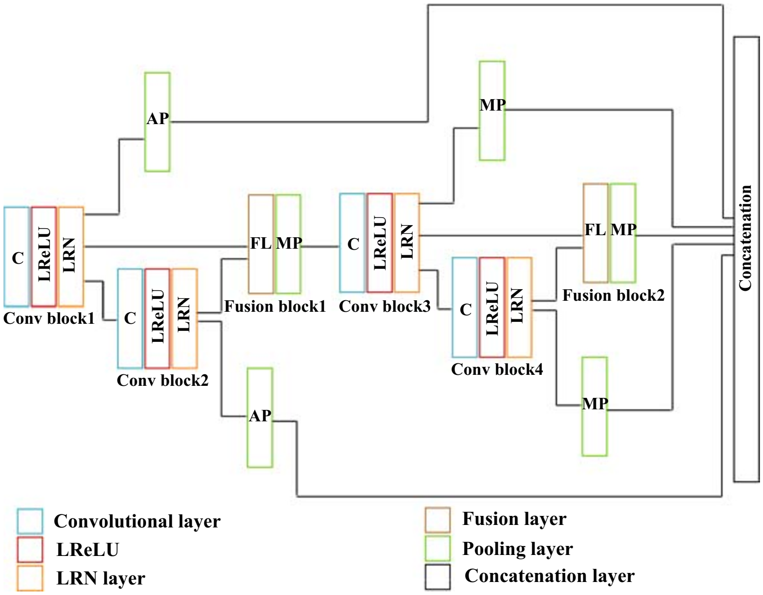

Fig. 4 The flow chart of the proposed CFU architecture for high-level feature designing task. Our architecture is built from four convolutional blocks (conv blocks), each group consists of convolutional layer (Blue rectangle), LReLU activation function (Red rectangle), and LRN layer (Orange rectangle). The diagram is composed of two fusion layers (FL) which are the element wise layers (Brown rectangle) and one concatenation layer (Black rectangle). A green rectangle represents downsampling layer (AP and MP). Note that the last layer concatenates the feature maps from four pooling paths with the same shape (width, height and depth)

Local Response Normalization (LRN) performs competition between neurons at the same location, but in different activation maps. It is useful when using neurons with unbounded activation function such as LReLU, because it allows the amplification of excited neuron with high-frequency and big response, while dampening the neurons that are quite large in a local neighborhood. It is a type of regularizer unit that encourages the competition for big activities among nearby groups of neurons. Denoting by $\mathrm{C_i(m,n)}$ the activity of a neuron computed by applying weight $\mathrm{W_i}$ at position (m,n) and the LReLU non-linearity, The LRN is expressed as follow:

$$LRN_{\mathrm{i}}(m,n) = C_{\mathrm{i}}(m,n) / \left(\lambda_1 + \lambda_2 \times \sum_{j=max(0,i-\frac{z}{2})}^{min(T-1,i+\frac{z}{2})} (C_{\mathrm{j}}(m,n))^2 \right)^{\lambda_3} \tag{4}$$

Where $\mathrm{C_i(m,n)}$ represents the i^{th} convolutional layer output after applying LReLU function at the position of (m,n) in the feature map, T is total number of weights, z represents the size of the normalization neighborhood and the constants (λ_1, λ_2 and λ_3) are hyper-parameters.

Our CFU is performed by inserting the fusion layers (FL) which enable automatic hierarchical feature fusing at every local response normalization layers. The FL is expressed as follow:

$$FL_{\mathrm{i}}(m,n) = \varphi_1 \times LRN_{\mathrm{i}}(m,n) \odot \varphi_2 \times LRN_{i+1}(m,n) \tag{5}$$

Where $\odot$ is element-wise product, φ_1 and φ_2 are two weights. Each element of $\mathrm{LRN_i}$ and $\mathrm{LRN_{i+1}}$ is multiplied by φ_1 and φ_2, respectively.

Pooling operation is important module and significant layer in the CNN architecture. It can be performed as small and square regions to reduce matrix size without losing important features and requiring large amount of resources. When the pooling operation works as the stacked layer in CNN, the most common advantage is that pooling operation decreases the spatial resolution of feature map and, as a result, it decreases the precise position of each of the irrelevant features that can not detect and identify the pattern. In other words, pooling layer forms the feature maps more relevant and achieves the translational invariance in CNN architecture. In this paper, we used two styles to do downsampling operation: Max-Pooling (MP) and Average Pooling (AP) [103]. More formally, the features of pooling region are arranged in chronological order and denoted as $X = [x_1, x_2, ...x_k]$, where x_1 is the first element extracted from the given pooling region and k is its size. The MP and AP are expressed as:

$$MP = max(X) \tag{6}$$

$$AP = \frac{1}{k} \times \sum_{i=1}^{k} x_{\mathrm{i}} \tag{7}$$

Our CFU configurations are quite different from the VGG-16, ResNet-18 and DenseNet-121. Rather than using a stack of consecutive convolutional layers in each block, ReLU as activation function, and SGD as algorithm optimizer, we use one convolutional layer per block with relatively deeper depth throughout the whole net to reduce the computational cost, introducing fusion blocks to maintain the information flow and ensure that no feature is lost along the network. Moreover, we used LReLU activation function to fix vanishing gradient problem [64, 101], and Adam optimization algorithm to improve the backpropagation algorithm with avoiding local minima and slow convergence speeds [45, 46]. Our CFU model consists of four convolutional blocks (conv block1, conv block2, conv block3 and conv block4). In each convolutional block, a convolutional layer, LReLU activation function, and LRN layer are involved. Besides of conv blocks, it is composed of two fusion layers (FL) which take as input previous LRN ones. Pooling layer is used after each conv block and fusion layer; the first and second conv blocks are interspersed with Average Pooling (AP) and the rest are followed by Max Pooling (MP) layers. The last layer concatenates and fuses the output from all previous layers. The arrangement and description of all CFU layers is shown in Fig. 4. All convolutional layers (C) built by same square filter size (f), depth (d), stride (s) and padding (d). Spatial pooling layers were carried out by same $m \times m$ pixel window and stride m for MP layers and $a \times a$ pixel window and stride a for AP ones. The CFU layers are randomly initialized using Xavier weight initialization algorithm [58] and trained towards the dermoscopic dataset. The Table 3 shows the parameters specification of the proposed CFU architecture.

3.4 Multi-layer feature fusion

Starting from the idea that different feature representations of the dermoscopic image could bring different information [5], we study the effects of the fusing dermoscopic features extracted from three CNN models. Unlike the previous methods that either simply trained features from a single feature vector using conventional CNN architecture [7], in this work, we propose a convolutional multi-layer fusion network to formulate a roust CAD system for skin abnormalities recognition. This technique is employed to select most discriminative feature descriptors from the outputs of the three VGG$_n$-CFU, ResNet$_n$-CFU and DenseNet$_n$-CFU models. As it can be seen in Fig. 1, these models produce three feature representations which are combined through the first fully connected fusion (FCF1) layer, generating a highly concentrated dermoscopic representation and resulting in an activation volume with size of 3 × 4096 features. This fusion layer is followed by second fully connected fusion (FCF2) layer. Then, the output features are fed into the third fully connected fusion (FCF3) layer, which outputs a 7-dimensional probability vector. Finally, a softmax layer is used to train all the parameters of convolutional multi-layer fusion network based on the back-propagation algorithm. It is a linear classifier that uses the cross-entropy function and considered as maximum likelihood estimation method which calculates the sample's log-likelihood (8) through training phase using Adam algorithm. Learning of the convolutional multi-layer fusion network is based on measuring a loss function that indicates the error of learned network parameters. The learning objective is to compute the parameters to minimize the loss function. It should minimize the negative log-likelihood of the correct class. The cross-entropy for a single data point is defined as below:

$$H = -\sum_{i=1}^{k} T_{\mathrm{i}} \times \log(S_{\mathrm{i}}) \tag{8}$$

Table 3 Parameters setting of the CFU layers

Layers	Specifications
C	f = 3, d = 512, s = 1, p = 1
LRN	z = 5, $\lambda_1 = 2$, $\lambda_2 = 10^{-5}$, $\lambda_3 = 0.75$
FL	$\varphi_1 = 0.5$, $\varphi_2 = 0.2$
MP	$m = 2$
AP	$a = 4$

Where k is the number of classes, $\mathrm{T_i}$ represents the target class probability and $\mathrm{S_i}$ is the softmax score for the predicted class probability. $\mathrm{S_i}$ for $i = [1, ..., k]$ and $Y = [y_1, ..., y_\mathrm{k}]$ is expressed as follow:

$$S_\mathrm{i} = e^{y_\mathrm{i}} / \sum_{j=1}^{k} e^{y_\mathrm{j}} \tag{9}$$

Where Y is the output of last fully connected layer before activation takes place, e is the exponential function, $\mathrm{y_i}$ refers to each element in the input vector Y. $\mathrm{e}^{y_\mathrm{i}}$ is normalized by dividing by the sum of all exponential elements of vector Y.

The loss function is defined by computing the cross-entropy over an entire dataset. Then, it is done by taking the average of $\mathrm{H_i}$ set:

$$L = \frac{1}{N} \times \sum_{i=1}^{N} H_\mathrm{i} \tag{10}$$

Where N is the number of training data.

To help our models generalize better in these circumstances, prevent overfitting, and provide a way of approximately combining exponentially many different neural network architectures efficiently, we introduce dropout layer after FCF1 and FCF2. It consists of randomly setting the output hidden neuron to zero with probability ρ=0.5, and then its weights will not get updated [12]. The neurons which are dropped out do not participate in the forward and back-propagation processes. At test stage, we multiply all the output neurons by ρ=0.5, which is a suitable probability value to obtain the wide vision of the predictive network outputs generated by the exponentially many different networks sharing the same parameter value. Moreover, we used L2 regularization [54] for driving the weights to decay towards smaller values and penalize large ones. It is the most common types of weight penalty which can prevent overfitting in our network by updating the loss function as follow:

$$L = \frac{1}{N} \times \sum_{i=1}^{N} H_\mathrm{i} + \epsilon \times \sum_{i} \sum_{j} W_\mathrm{ij}^2 \tag{11}$$

Where ϵ is the regularization hyper-parameter and W is the weight matrix. In this paper, we used ϵ=0.05 as it gave the best results and penalized 5% of the current weight value.

4 Experiment analysis and evaluation

At this stage, the experimental results and their interpretation for multi-layer feature fusion network based on continuous fine-tuning technique, which were pointed out in the Section 3 to this paper, are shown and explained precisely and in exquisite detail. To classify dermoscopic patterns into MEL, NV, BCC, AKIEC, BKL, DF and VASC, we fuse and concatenate three discriminative representations: $\mathrm{VGG_n}$-CFU, $\mathrm{ResNet_n}$-CFU and $\mathrm{DenseNet_n}$-CFU, where the low-level features were confirmed and validated according to the number of the frozen blocks n. We demonstrate the robustness of the methodology on HAM10000 dataset. In order to prove the efficiency of our proposed strategy, statistical measurements and qualitative evaluation are carried out. Our approach was compared with the most recent existing methods for skin lesion recognition. All these methods were implemented, tested on same dataset (HAM10000) and evaluated by same metrics (Accuracy, sensitivity, specificity, precision, F1-score and ROC-AUC).

4.1 Acquisition images

In our experiments, we evaluated the performance of our proposed system on HAM10000 dataset which is the official ISIC 2018 challenge dataset hosted by the annual MICCAI conference in Granada, Spain, but is also available to research groups who do not participate in the challenge [82]. This dataset was collected over a period of 20 years from two different sites, the department of dermatology at the medical university of Vienna, Austria and the skin cancer practice of Cliff Rosendahl in Queensland, Australia. It includes seven generic classes: MEL, NV, BCC, AKIEC, BKL, DF and VASC. The dataset consists of 10015 dermoscopic images with size of 600 × 450 showing skin lesions which have been diagnosed based on expert consensus, serial imaging and histopathology. The dermoscopic images from this database can be used for the research, development and comparison of various algorithms for identifying skin cancer. According of data percentage shown in Table 4, the distribution is completely imbalanced and most of the samples are melanocytic nevus, has 6705 images while the smallest category is dermatofibroma, it only has 115 images. To address this problem, we have selected 2000 images from NV class, and then we applied the geometric transformation techniques described in Section 3.1.2 that could augment the MEL by 44.35%, BCC by 74.3%, AKIEC by 83.65%, BKL by 45.05%, DF by 94.25% and VASC by 92.9%. The typical data set proportions are 60% for training, 20% for validation and 20% for testing [70, 92]. Dermoscopic data distribution for each class after class balancing is shown in Table 5.

4.2 Experimental environment

Training and testing of our proposed method is performed using GPU programming to achieve higher time efficiency over CPU. It has been developed using caffe deep learning framework [44] with python wrapper and Compute Unified Device Architecture (CUDA) enabled parallel computing platform to access the computational resources of Graphics Processing Unit (GPU) [21]. Conversion of image datasets into LMDB, computations of image means and learning pre-trained weights from VGG-16, ResNet-18 and DenseNet-121 models are performed using the modules in caffe framework. We used Deep Learning GPU Training System (DIGITS). It is a web server providing a suitable and robust web interface for training and visualizing the features in the convolutional layers based on caffe. The available hardware, used for training, is a PC with a Core i7 CPU, 8 GB RAM and a single NVIDIA GeForce GTX 1060 with 6 GB memory. Training dermoscopic images using the convolutional multi-layer feature fusion algorithm took about 9 hours and 27 minutes. For testing, on average, it takes 4.02 seconds in GPU and 7.45 seconds in CPU per image.

Table 4 Number and percentage of dermoscopic images per class of the HAM10000 dataset

Pattern	Count	Percentage
MEL	1113	11.11%
NV	6705	66.95%
BCC	514	05.13%
AKIEC	327	03.27%
BKL	1099	10.97%
DF	115	01.15%
VASC	142	01.42%

Table 5 Detailed description of the representative skin pattern categories divided into training, validation and testing subsets after class balancing process

	MEL	NV	BCC	AKIEC	BKL	DF	VASC	Total
Train	1200	1200	1200	1200	1200	1200	1200	8400
Validation	400	400	400	400	400	400	400	2800
Test	400	400	400	400	400	400	400	2800
Total	2000	2000	2000	2000	2000	2000	2000	14000

4.3 Network configuration

The performance of a deep CNN model depends critically on its structure and the network configuration [60]. In this paper, there still remain a number of network hyper-parameters to be defined. Many of these hyper-parameters were chosen by experimenting until the network began to train effectively. The low-level of our system, which consists of frozen layers, was built using small learning rate $\alpha = 10^{-5}$, while the high-level of our proposed architecture was trained using Adam algorithm [45] with learning rate $\alpha = 10^{-3}$, first moment-decay $\beta_1 = 0.9$ and second moment-decay $\beta_2 = 0.999$ using Xavier algorithm for weight initialization and LReLU activation function inserted after each convolutional layer, where the softmax activation function was combined with the cross-entropy for training the network. The optimization ran for 100 epochs with mini batch size 32. The network was regularized using the dropout technique with factor $\rho = 0.5$ and L2 regularization with $\epsilon = 0.05$ to prevent the overfitting issues.

4.4 Evaluation metrics

The evaluation of the performance of our proposed system was carried out using three quantitative methods including confusion matrix, Receiver Operating Characteristic (ROC) and Precision-Recall (PR) curves.

The confusion matrix is a statistical process that is employed to evaluate the effectiveness and robustness of our classification system. It contains information about actual and predicted classifications done by a learning model. Calculating a confusion matrix can give us a better idea of what our classifier is getting right and what types of errors it is making. Performance of such systems is commonly evaluated using the data in the matrix. It produces four outcomes: True Positive (TP), False Positive (FP), True Negative (TN) and False Negative (FN), where TP is correct positive prediction, FP presents incorrect positive prediction, TN gives correct negative prediction and FN determines incorrect negative prediction. To compare the performances of our proposed method, we calculate six assessment metrics: accuracy (Acc), sensitivity (Sen), specificity (Spe), error (Err), precision (Pre) and F1-Score (FSc) which can be defined by:

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \tag{12}$$

$$Sen = \frac{TP}{TP + FN} \tag{13}$$

$$Spe = \frac{TN}{TN + FP} \tag{14}$$

$$Err = \frac{FP + FN}{TP + TN + FP + FN} \tag{15}$$

$$Pre = \frac{TP}{TP + FP} \tag{16}$$

$$FSc = 2 \times \frac{Pre \times Sen}{Pre + Sen} \tag{17}$$

Where:

- Acc is the proportion of the total number of predictions that were correct.
- Sen is the proportion of positive cases that were correctly identified.
- Spe is the proportion of negative cases that were correctly identified.
- Err is the proportion of the total number of predictions that were incorrect.
- Pre is the proportion of correctly predicted Positive divided by total predicted Positive.
- FSc is a harmonic mean of sensitivity and precision.

The second method for evaluating our proposed system is ROC curve, which is created by plotting the sensitivity as the y coordinate against the 100-specificity as the x axis, is a practical process and reliable method of reporting the performance of the clinical prediction tests, and is widely used in medical field to validate the robustness of a CAD system for abnormal pattern recognition and classification. With regard to our study, ROC plays a vital role in measuring performance of a diagnostic system for medical interpretation and comparing different dermoscopic researches when each study is carried out on the same case and dataset. It is often exploited in binary classification tasks to evaluate and measure the effectiveness of a classifier. For extending ROC curve to multi-class and multi-output problem, an average ROC process generated from ROC curves of multiple classes has attracted the attention of many researchers. In this study, we implemented a new technique for averaging ROC curves suggested by authors of [16] based on non-parametric method. The metric ROC-AUC, which is the area between the curve and the x coordinate (100-specificity), is employed to summarize the performance of dermoscopic image analysis systems and used to compare our proposed system with the existing techniques.

To facilitate class-wise visual interpretation, we used PR curve, which is defined as a plot of precision versus recall, to typically illustrate the performance differences between seven dermoscopic categories: MEL, NV, BCC, AKIEC, BKL, DF and VASC. Similar to ROC-AUC, the Precision-Recall Area Under Curve (PR-AUC) is calculated to report the performance of an algorithm on a given dataset. In this study, PR-AUC is used to evaluate class-wise performance and reveal the robustness of the proposed system to maximize the inter-class variability. In our experiments, we take the average PR curves and the corresponding PR-AUCs of all plots as the final results for measuring the overall classification performance across all classes using micro-averaging method [75].

4.5 Quantitative evaluation

As described in Section 3.2, low-level features were processed simultaneously in three parallel paths of the CNN as a function of the number of frozen blocks n; VGG_n, ResNet_n and DenseNet_n. The performance of the different level on the test dataset is shown in Table 6. From these results, it can be seen that the best configuration of fine-tuning process for the VGG-16 and ResNet-18 achieved when n=4 with freezing the weights of the first four blocks (Level_4), yielding classification accuracy of 87.23% with ROC-AUC of 88.71% for VGG-16 and classification accuracy of 89.45% with ROC-AUC of 87.82% for ResNet-18. For the DenseNet-121, the fine-tuning with freezing the weights of the first three blocks

Table 6 Quantitative result of different fine-tuning levels on testing subset in terms of accuracy and ROC-AUC metrics exploring transferability of VGG-16, ResNet-18 and DenseNet-121 pre-trained models

Level$_n$	VGG$_n$		ResNet$_n$		DenseNet$_n$	
	Acc. %	ROC-AUC %	Acc. %	ROC-AUC %	Acc. %	ROC-AUC %
Level$_0$	79.52	81.42	80.17	80.96	81.60	83.27
Level$_1$	83.07	84.28	83.62	82.40	86.13	85.98
Level$_2$	85.31	86.66	85.11	86.34	86.44	87.32
Level$_3$	86.70	85.84	87.58	86.41	90.37	92.52
Level$_4$	87.23	88.71	89.45	87.82	86.97	86.02
Level$_5$	87.06	85.98	86.69	86.21	85.11	84.70

(Level$_3$) is the best with 90.37% classification accuracy and 92.52% ROC-AUC rate on the test set.

As mentioned in Section 3.4, the proposed multi-layer feature fusion network has three CNN models; VGG$_n$-CFU, ResNet$_n$-CFU and DenseNet$_n$-CFU. By comparing the results in Table 6, the experiments suggest that four blocks from each VGG-16 and ResNet-18 models and three blocks from DenseNet-121 can design low-level features of these three models: VGG$_4$-CFU, ResNet$_4$-CFU and DenseNet$_3$-CFU. As shown in Fig. 5, we evaluated the training quality of our proposed framework using training and validation subsets. These curves show the predictive generalization performance of our proposed model as a function of the number of epochs. At epoch 54, the training and validation curves are identical which reach same accuracy and loss values. Hence, this model can typically predict new dermoscopic pattern without suffering from overfitting. The experimental evaluation as shown in Tables 7 and 8 demonstrates that the multi-layer feature fusion network achieved the promising performance on testing dataset, yielding an average accuracy of 98.09% with sensitivity of 93.35%, specificity of 98.88%, precision of 93.36%, F1-Score of 93.35% and error of 1.90%. From Fig. 6, the average ROC shows the best performance of our proposed method in terms of ROC-AUC rate with 96.51%.

As illustrated in figures (Figs. 7, 8, 9, 10, 11, 12 and 13) and listed in Table 9, the class-wise performance was evaluated using PR curve and the corresponding PR-AUC metric.

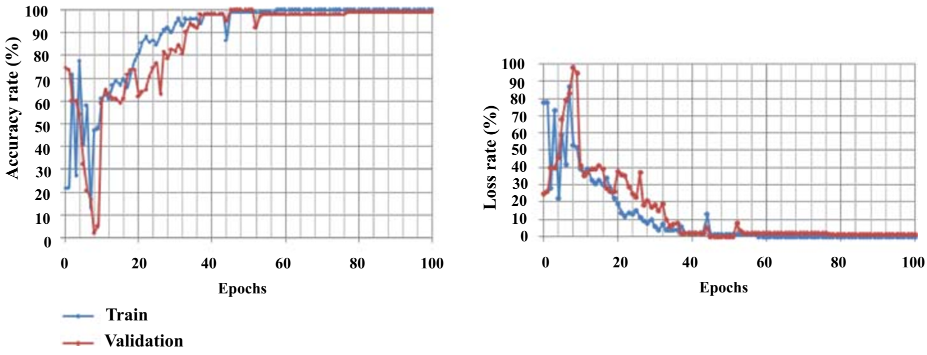

Fig. 5 Learning curves of the proposed model. The comparison of the performance on training and validation dermoscopic data over the varying number of epochs show that at epoch 54, the training and validation curves converge at similar values, and they show a very small gap between training and validating loss rates

Table 7 Confusion matrix of skin abnormality classification given by our proposed method on testing subset

Classes	TP	TN	FP	FN
MEL	376	2369	31	24
NV	372	2369	31	28
BCC	375	2367	33	25
AKIEC	369	2372	28	31
BKL	371	2376	24	29
DF	377	2380	20	23
VASC	374	2381	19	26

The reported system has an overall PR-AUC of 95.73% on testing dataset. During diagnosis procedure, recognition of AKIEC, NV and BCC is the most challenging and very tedious task. From the results summarized in Table 9, the proposed system achieved the PR-AUCs of 92.38%, 92.65% and 93.14% for recognition of AKIEC, NV and BCC, respectively. When the proposed system was tested against MEL, BKL, VASC and DF abnormalities, the results were notable with significant improvements in terms of PR-AUC, yielding 94.27%, 94.62%, 95.04% and 95.91% for MEL, BKL, VASC and DF recognition, respectively.

4.6 Qualitative evaluation

The measurement of the effectiveness of our proposed multi-layer feature fusion network was confirmed by comparing the performance of three CNN medels; VGG-16, ResNet-18 and DenseNet-121. Based on the summary results discussed in Section 4.5, it is clear that the fine-tuning technique combined with hierarchical multi-feature fusion network achieved a promising performance. To evaluate the visual results, we displayed in Fig. 14 the qualitative recognition results of the most classified skin patterns as MEL, NV, BCC, AKIEC, BKL, DF and VASC with softmax probability of 100%. It is remarkable, that the network learned typical textural filters in different CNN model. Hence, we can conclude that these three cooperative CNN models (VGG$_4$-CFU, ResNet$_4$-CFU, and DenseNet$_3$-CFU) are performing better in recognizing and classifying all dermoscopic patterns as MEL, NV, BCC, AKIEC, BKL, DF and VASC than the existing other methods. As was mentioned earlier (Table 8), our proposed study achieved an average of 1.90% in terms of classification error rate. The most common dermoscopic pattern classification error was misclassifying a BCC as MEL (Fig. 15a), a NV as AKIEC tissue (Fig. 15b). In addition, most of AKIEC (Figs. 15c

Table 8 Validity assessment measures for seven classes given by our proposed method on testing subset

Classes	Acc. %	Sen. %	Spe. %	Err. %	Pre. %	FSc. %
MEL	98.03	94.00	98.70	01.97	92.38	93.18
NV	97.89	93.00	98.70	02.11	92.30	92.64
BCC	97.92	93.75	98.62	02.08	91.91	92.82
AKIEC	97.89	92.25	98.83	02.11	92.94	92.59
BKL	98.10	92.75	99.00	01.90	93.92	93.33
DF	98.46	94.25	99.16	01.54	94.96	94.60
VASC	98.39	93.50	99.20	01.61	95.16	94.32
Average	98.09	93.35	98.88	01.90	93.36	93.35

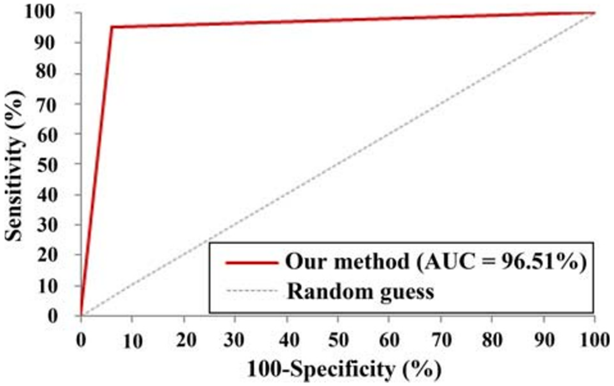

Fig. 6 The average ROC curve of the CAD system for seven classes on the testing dataset. For predicting the seven skin lesion categories, the proposed model had a ROC-AUC of 96.51%

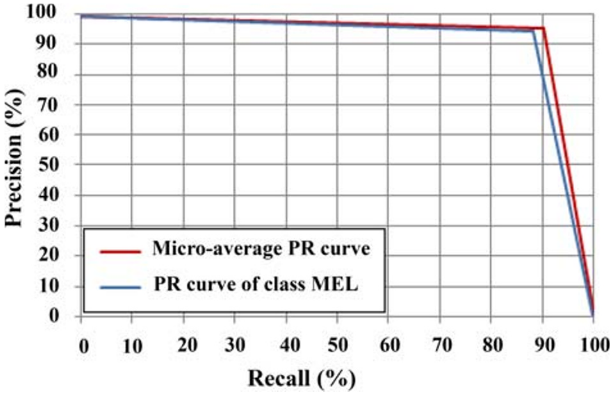

Fig. 7 The precision-recall curve of melanoma abnormality compared with micro-average precision-recall curve of our proposed model. For predicting the melanoma class, the proposed model had a PR-AUC of 94.27%

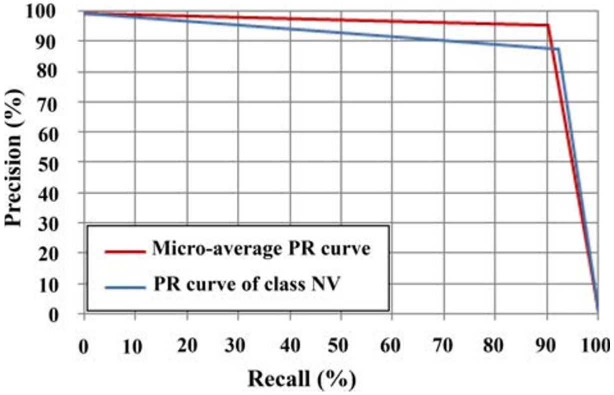

Fig. 8 The precision-recall curve of melanocytic nevus abnormality compared with micro-average precision-recall curve of our proposed model. For predicting the melanocytic nevus class, the proposed model had a PR-AUC of 92.65%

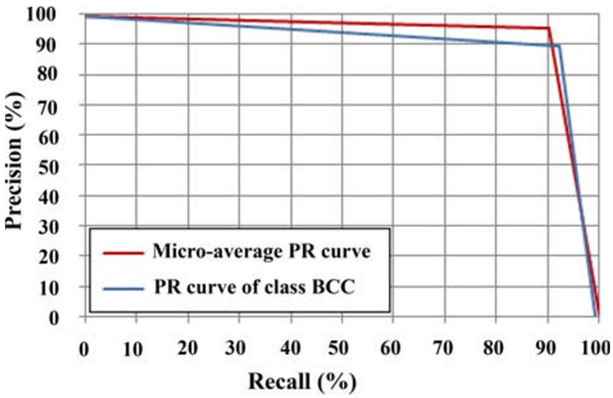

Fig. 9 The precision-recall curve of basal cell carcinoma abnormality compared with micro-average precision-recall curve of our proposed model. For predicting the basal cell carcinoma class, the proposed model had a PR-AUC of 93.14%

and d) are confused with the BKL structure. These confusions occur due to similar morphological features, the image acquisition systems which captured some dermoscopic patterns with small scale and weakly contrasted texture and the insufficiency of these patterns in our training set. Figure 15 shows the most dermoscopic abnormalities misclassified by our proposed approach.

4.7 Comparison with the state of the art

As shown in Table 10, we make a comparison between the proposed CAD system and existing state-of-the-art in order to evaluate and confirm the quality of this study. Based on the summary and comparison of the dermoscopic research results discussed in Section 2, it is difficult to evaluate the CAD system and ensure its quality that achieves high effectiveness in terms of medical diagnosis, because the dataset and quantitative evaluation methods are not standardized. As a result, we should assure that for a dependable comparison and reliable

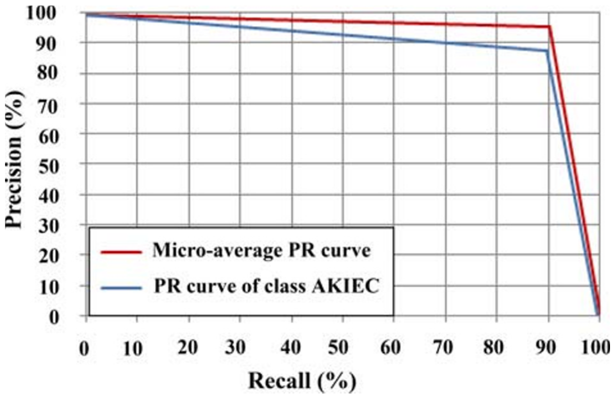

Fig. 10 The precision-recall curve of actinic keratosis abnormality compared with micro-average precision-recall curve of our proposed model. For predicting the actinic keratosis class, the proposed model had a PR-AUC of 92.38%

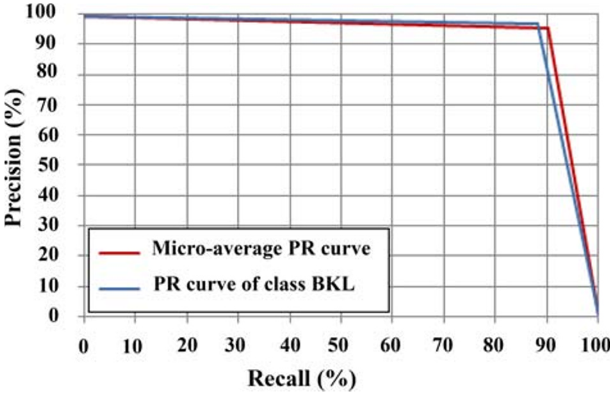

Fig. 11 The precision-recall curve of benign keratosis abnormality compared with micro-average precision-recall curve of our proposed model. For predicting the benign keratosis class, the proposed model had a PR-AUC of 94.62%

clinical research, it would be necessary to use the same database (HAM10000) and the same metrics (Accuracy, Sensitivity, Specificity, Precision, F1-score and ROC-AUC) for validating the performance of our methodology. However, research on this dermoscopic dataset (HAM10000) is still relatively scarce. To overcome this limitation, we tried to implement some existing algorithms used for skin lesion recognition and test them on our dermoscopic dataset. Then, the [10, 25, 28, 37, 57, 97] were implemented and tested on HAM10000 dataset. The visual comparison given in Table 10 demonstrates that our best performing network outperforms the pre-trained AlexNet with an error-correcting output coding classifier combined with support vector machine [25], convolutional neural-adaptive networks [10], pre-trained GoogleNet inception-V3 [28], pre-trained very deep residual networks with 152 layers [37], lesion feature network [57] and pre-trained VGG-16 [97].

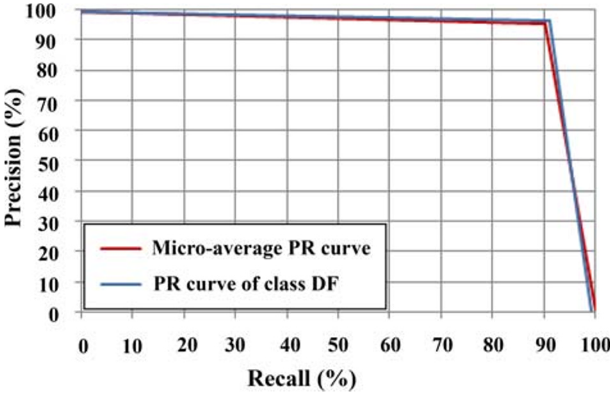

Fig. 12 The precision-recall curve of dermatofibroma abnormality compared with micro-average precision-recall curve of our proposed model. For predicting the dermatofibroma class, the proposed model had a PR-AUC of 95.91%

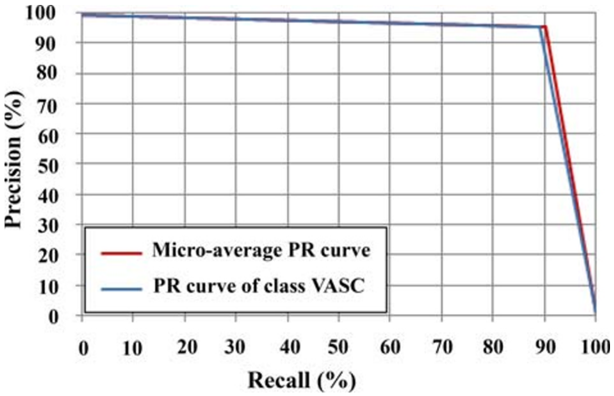

Fig. 13 The precision-recall curve of vascular lesion abnormality compared with micro-average precision-recall curve of our proposed model. For predicting the vascular lesion class, the proposed model had a PR-AUC of 95.04%

5 Discussion

Pattern recognition in dermoscopic image is an area of great research interest due to its importance in skin cancer prevention, as well as in the early diagnosis. Therefore, several automatic diagnosis systems have been proposed using models with a single CNN stream to generate discriminative features for skin lesion classification. As mentioned in Section 2.3, current studies in pattern recognition research field reported that multiple descriptor fusion performed better than single feature map for multi-class classification problem. However, research on multi-feature fusion topic for dermoscopic pattern recognition is still relatively scarce. In addition, dermoscopic imaging field suffers from limited small number of available samples to build an efficient deep CNN models. Thus, one of the major challenges in building deep CNN architecture using dermoscopic data appears from the limited number of training images to create CNN models without occurring overfitting. This paper presented several gaps in dermoscopic domain as well a robust system for investigating them. The proposed approach was evaluated on HAM10000 dataset [82] which includes seven skin lesion categories: MEL, NV, BCC, AKIEC, BKL, DF and VASC. Among several evaluation procedures, the randomly splitting dataset into 60% for training, 20% for validation and 20% for testing [54, 70] is the most commonly used in the literature to evaluate the results of skin

Table 9 System performance in PR-AUC for each dermoscopic class

Classes	PR-AUC %
MEL	94.27
NV	92.65
BCC	93.14
AKIEC	92.38
BKL	94.62
DF	95.91
VASC	95.04
Micro-average	95.73

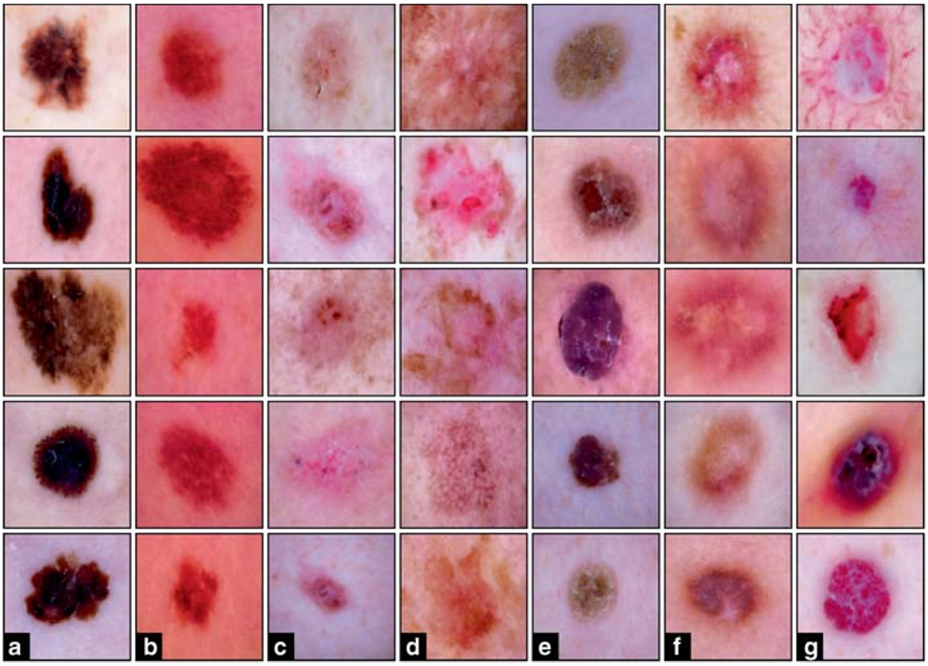

Fig. 14 An illustrative example showing the most classified skin lesions with softmax probability of 100%. **a** Melanoma. **b** Melanocytic nevus. **c** Basal cell carcinoma. **d** Actinic keratosis. **e** Benign keratosis. **f** Dermatofibroma. **g** Vascular lesion

lesion classification, since it avoids overfitting while testing the capacity of the classifier to generalize.

Motivated by the outstanding performance of transfer learning in visual pattern recognition area, fine-tuning is proposed to solve the data deficiency problem in dermoscopic object recognition. As [3, 10, 15, 56, 89, 105] pointed out, lower layers in the convolutional neural network learn fundamental features which are common in many objects, while higher layers learn semantic features which are specific to the target task. Therefore, a good transfer learning scheme should preserve the ability of learning common features and tune the high features to the target object categories. To achieve this goal, we carried out extensive experiments to explorer the right choice of frozen blocks by continuous fine-tuning technique using three pre-trained CNN models; VGG-16 [74], ResNet-18 [39] and DenseNet-121 [41]

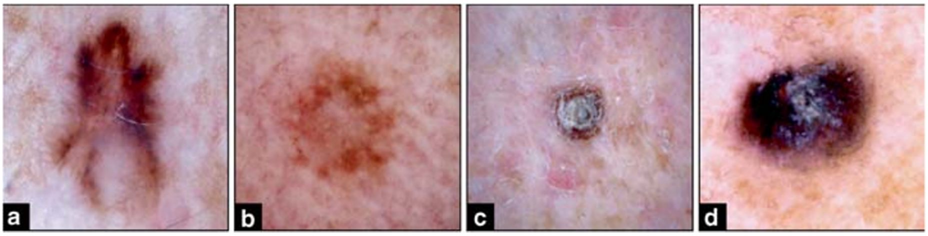

Fig. 15 The most skin abnormalities misclassified by the proposed approach and confused with other dermoscopic patterns. **a** Basal cell carcinoma. **b** Melanocytic nevus. **c**–**d** Actinic keratosis

Table 10 Comparison between the proposed method and other algorithms using the HAM10000 dataset

Methods	Acc. %	Sen. %	Spe. %	Pre. %	FSc. %	ROC-AUC %
Method [28]	91.56	87.26	92.48	92.52	89.81	91.83
Method [25]	83.91	81.47	84.20	81.76	81.61	81.55
Method [10]	84.28	79.34	84.79	84.01	81.60	83.62
Method [37]	86.77	84.91	86.67	85.37	85.13	85.47
Method [57]	83.71	78.15	86.41	80.46	79.28	83.03
Method [97]	87.06	85.75	86.32	86.32	86.03	85.98
Ours	98.09	93.35	98.88	93.36	93.35	96.51

which can be considered the most efficient architectures owing to their high performance achievement on different datasets for object recognition tasks. We based our experiments on these three pre-trained models because they are deep enough that we can investigate the impact of the proposed method on the dermoscopic pattern recognition task, and they are also shallow enough that we can conduct experiments quickly. To decide the best transfer scheme, we do extensive experiments by retraining different blocks in order. Firstly, the whole network is trained on HAM10000, and then different retraining scheme is carried out (Fig. 3). For example, retraining the last block, the last two, the last three and so on. Details about the results are given in Table 6. The best performance was achieved when n=4 for VGG-16 and ResNet-18 with freezing the first four blocks, and n=3 for DenseNet-121. As we are practically choosing the best-performing blocks to design low-level features when presenting the fine-tuning results, we are fixing the same convolutional fusion unit denoted by CFU for all selected blocks to generate high-level features. To make our work more robust and efficient, we proposed multi-layer feature fusion network, which consists of three parallel CNN paths; VGG$_4$-CFU, ResNet$_4$-CFU and DenseNet$_3$-CFU. By comparing the results in Tables 6 and 8, the proposed fusion technique increases the accuracy at least by 7.87% comparing the results of the best configuration of fine-tuning without fusion. Our results using fusion were also compared with the previous works in Table 10, it is remarkable that our proposed system achieved the promising results with an average accuracy of 98.09%, sensitivity of 93.35%, specificity of 98.88%, precision of 93.36%, F1-Score of 93.35%, ROC-AUC of 96.51% and PR-AUC of 95.73%.

From the analysis of the results, it is evident that our proposed algorithm is more robust as compared to fine-tuning configurations and state of the art works mentioned in Tables 6 and 10, respectively. All the previous works used SGD as optimization algorithm, which is very slow, and causes convergence to the local minima rather than the global minima [46]. In addition, the ReLU was used as activation function, which replaces the negative component by zero, occurring the bias shift of the outputs and the death of some neurons, which leads to the impeding learning, vanishing gradient and oscillation problem in optimization algorithms [101]. In our study, we solved these issues by using Adam optimization algorithm [45] to train CFU architecture, which combines the advantages of two recently popular methods; AdaGrad [27] and RMSProp [69]. So, Adam is a good alternative to SGD since it can handle sparse gradients on noisy problems and alleviate the convergence to local minima with little memory consuming. To overcome the limitations of ReLU, LReLU [64] is proposed as activation function that can avoid the zero gradients in negative part and the bias shift of the ReLU function. In order to address the problem of overfitting, we used three techniques; inserting a dropout [12] layer after each FCF layer with a probability of 0.5, using L2 regularization [54] with factor ϵ=0.05, and using the pre-trained blocks from three models (VGG-16, ResNet-18 and DenseNet-121) as hierarchical low-level feature extractors of the proposed system. Additionally, the hyper-parameters setting also play a substantial role in the final performance of the systems. In this work, the hyper-parameters were experimentally chosen until the network began to converge effectively. From the practical point of views [23, 42, 53, 84, 95], the best way to alleviate the problem of semantic feature ambiguity and high inter-class visual similarities between the classes is to employ hierarchical multi-feature fusion technique to ensure maximum information flow between layers in the network. The misclassification of some skin abnormalities occur due to similar morphological appearances (Fig. 15). Another issue that needs to be taken into account is the technological limitations of the image acquisition systems. Better performance may be achieved with higher resolutions which may capture skin features better, such as asymmetrical characteristic, irregular edge of the lesion, different color composition and diameter size, which play a key role in the classification of dermoscopic patterns [61].

The computational time is one of the most essential factors of deep CNN. Consequently, we have always tried to insert after each convolutional block and fusion layer a down-sampling layer [33]. The max pooling layer reduces the dimension of the features with completely keeping a balance between high performance achievement and low computational times. The model build time for the proposed strategy is evaluated by using NVIDIA DIGITS system based on caffe framework [92] for rapidly training the highly accurate deep CNN and a single NVIDIA GeForce GTX 1060 with 6 GB memory. Our experiments demonstrate that the proposed system took around 9 hours and 27 minutes. For each dermoscopic image, the test takes, on average, 4.02 seconds in GPU and 7.45 seconds in CPU.

To the best of our knowledge, this is the first study that implements a very deep hierarchical multi-feature fusion network for the automated recognition system of MEL, NV, BCC, AKIEC, BKL, DF and VASC dermoscopic patterns. Based on the results, the summary of proposed algorithm can be drawn as:

- It can be argued that the proposed algorithm is significantly robust, reliable and efficient comparing to other state of the art classification algorithms.
- The proposed algorithm overcomes many problems that confronted most of previous algorithms such as vanishing gradient, requirement of too much memory, decreasing the bias shift, convergence to local minima and overfitting issue.
- The computational cost of our algorithm is lower than other algorithms that used CNN.
- The proposed CAD system can serve as a complete expert tool to help clinicians in confirming their diagnosis and analysis.

Finally, this computational method based on transfer learning and fusion technique can perform better and more effectively in recognizing skin lesions in dermoscopic images. In addition, such method may fix and control the various gaps in the skin pattern recognition area, starting from image size normalization, class balancing, hair artifact removal, contrast enhancement and ending with the complex skin lesion classification, which convert CAD systems into more complete expert systems for diagnosing such lesions based on dermoscopic images.

6 Conclusion

In this study, we proposed a novel CAD system to classify skin pattern as MEL, NV, BCC, AKIEC, BKL, DF or VASC. Firstly, by integrating continuous fine-tuning CNN models:

VGG-16, ResNet-18 and DenseNet-121, the right frozen blocks were chosen to design efficient low-level features which not only reduce the computational cost with considering high performance achievement, but also robustly deal with limited availability of the training dermoscopic samples and overfitting problem. Then, CFU architecture was implemented to generate high-level features and overcome the limitation of VGG-16, ResNet-18 and DenseNet-121. It takes advantage of using the LReLU activation function that can avoid the zero gradients in negative part and the bias shift, Adam optimization algorithm to handle sparse gradients on noisy problems and the convergence to local minima with little memory consuming, and fusion layers to extract the latent feature representations. Finally, the high inter-class visual similarity problem was alleviated by fusing the three high-level semantic feature maps (VGG$_4$-CFU, ResNet$_4$-CFU, and DenseNet$_3$-CFU). The experimental results show that our proposed method outperforms the most representative state-of-the-art strategies in dermoscopic pattern recognition.

The proposed system is completely automatic, which provides a complete functionality from data pre-processing to the skin abnormality diagnosis, fast and easy for configuration, which requires minimum interaction from the dermatologists for use in clinical applications and making the final decision for the further treatment.

Using insufficient data with small scale is a major problem in dermoscopic abnormality recognition which may increase the classification error and harm the performance of the CAD system. Therefore, as a future research issue, we plan to collect and construct new balanced dermoscopic data including more challenging images, improve the proposed method by taking into consideration multi-scale analysis of skin pattern in order to boost class-wise performance, and evaluate our approach on more datasets.

References

1. Abd-Ellah M, Awad A, Khalaf A, Hamed H (2018) Two-phase multi-model automatic brain tumour diagnosis system from magnetic resonance images using convolutional neural networks. EURASIP Journal on Image and Video Processing 2018(97):1–10. https://doi.org/10.1186/s13640-018-0332-4
2. Agrawal P, Girshick R, Malik J (2014) Analyzing the performance of multilayer neural networks for object recognition. In: Computer vision-ECCV 2014 329-344. https://doi.org/10.1007/978-3-319-10584-0_22
3. Akcay S, Kundegorski M, Devereux M, Breckon T (2016) Transfer learning using convolutional neural networks for object classification within X-ray baggage security imagery. In: 2016 IEEE international conference on image processing (ICIP). https://doi.org/10.1109/icip.2016.7532519
4. Anic G, Sondak V, Messina J, Fenske N, Zager J, Cherpelis B, Lee J, Fulp W, Epling-Burnette P, Park J, Rollison D (2013) Telomere length and risk of melanoma, squamous cell carcinoma, and basal cell carcinoma. Cancer Epidemiol 37(4):434–439. https://doi.org/10.1016/j.canep.2013.02.010
5. Antropova N, Huynh B, Giger M (2017) A deep feature fusion methodology for breast cancer diagnosis demonstrated on three imaging modality datasets. Med Phys 44(10):5162–5171. https://doi.org/10.1002/mp.12453
6. Anwar S, Majid M, Qayyum A, Awais M, Alnowami M, Khan M (2018) Medical image analysis using convolutional neural networks: A review. J Med Syst 42(11):1–13. https://doi.org/10.1007/s10916-018-1088-1
7. Aubreville M, Knipfer C, Oetter N, Jaremenko C, Rodner E, Denzler J, Bohr C, Neumann H, Stelzle F, Maier A (2017) Automatic classification of cancerous tissue in Laserendomicroscopy images of the oral cavity using deep learning. Sci Rep 7(1):1–10. https://doi.org/10.1038/s41598-017-12320-8
8. Azizpour H, Razavian A, Sullivan J, Maki A, Carlsson S (2015) From generic to specific deep representations for visual recognition. In: 2015 IEEE conference on computer vision and pattern recognition workshops (CVPRW). https://doi.org/10.1109/cvprw.2015.7301270
9. Bakkouri I, Afdel K (2017) Breast tumor classification based on deep convolutional neural networks. In: 2017 international conference on advanced technologies for signal and image processing (ATSIP). https://doi.org/10.1109/atsip.2017.8075562

10. Bakkouri I, Afdel K (2018) Convolutional neural-adaptive networks for melanoma recognition. In: International conference on image and signal processing 453-460. https://doi.org/10.1007/978-3-319-94211-7_49
11. Bakkouri I, Afdel K (2018) Multi-scale CNN based on region proposals for efficient breast abnormality recognition. Multimed Tools Appl 78(10):12939–12960. https://doi.org/10.1007/s11042-018-6267-z
12. Baldi P, Sadowski P (2014) The dropout learning algorithm. Artif Intell 210:78–122. https://doi.org/10.1016/j.artint.2014.02.004
13. Banerjee I, Crawley A, Bhethanabotla M, Daldrup-Link H, Rubin D (2018) Transfer learning on fused multiparametric MR images for classifying histopathological subtypes of rhabdomyosarcoma. Comput Med Imaging Graph 65:167–175. https://doi.org/10.1016/j.compmedimag.2017.05.002
14. Byra M, Styczynski G, Szmigielski C, Kalinowski P, Michalowski L, Paluszkiewicz R, Ziarkiewicz-Wroblewska B, Zieniewicz K, Sobieraj P, Nowicki A (2018) Transfer learning with deep convolutional neural network for liver steatosis assessment in ultrasound images. Int J Comput Assist Radiol Surg 13(12):1895–1903. https://doi.org/10.1007/s11548-018-1843-2
15. Castrejon L, Aytar Y, Vondrick C, Pirsiavash H, Torralba A (2016) Learning aligned cross-modal representations from weakly aligned data. In: 2016 IEEE conference on computer vision and pattern recognition (CVPR). https://doi.org/10.1109/cvpr.2016.321
16. Chen W, Samuelson F (2014) The average receiver operating characteristic curve in multireader multicase imaging studies. Br J Radiol 87(1040):1–8. https://doi.org/10.1259/bjr.20140016
17. Chen Y, Xie H, Shin H (2018) Multi-layer fusion techniques using a CNN for multispectral pedestrian detection. IET Comput Vis 12(8):1179–1187. https://doi.org/10.1049/iet-cvi.2018.5315
18. Chougrad H, Zouaki H, Alheyane O (2018) Deep convolutional neural networks for breast cancer screening. Comput Methods Programs Biomed 157:19–30. https://doi.org/10.1016/j.cmpb.2018.01.011
19. Chu B, Madhavan V, Beijbom O, Hoffman J, Darrell T (2016) Best practices for fine-tuning visual classifiers to new domains. In: European conference on computer vision 9915:435-442. https://doi.org/10.1007/978-3-319-49409-8_34
20. Chu J, Guo Z, Leng L (2018) Object detection based on multi-layer convolution feature fusion and online hard example mining. IEEE Access 6:19959–19967. https://doi.org/10.1109/access.2018.2815149
21. Coates A, Huval B, Wang T, Wu D, Catanzaro BY, Ng A (2013) Deep learning with COTS HPC systems. In: International conference on machine learning
22. Codella N, Gutman D, Celebi M, Helba B, Marchetti M, Dusza S, Kalloo A, Liopyris K, Mishra N, Kittler H, Halpern A (2018) Skin lesion analysis toward melanoma detection: A challenge at the 2017 International symposium on biomedical imaging (ISBI), hosted by the international skin imaging collaboration (ISIC). In: 2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018). https://doi.org/10.1109/isbi.2018.8363547
23. Dai X, Ng J, Davis L (2017) FASON: First and second order information fusion network for texture recognition. In: 2017 IEEE conference on computer vision and pattern recognition (CVPR). https://doi.org/10.1109/cvpr.2017.646
24. Ding S, Zhu H, Jia W, Su C (2011) A survey on feature extraction for pattern recognition. Artif Intell Rev 37(3):169–180. https://doi.org/10.1007/s10462-011-9225-y
25. Dorj U, Lee K, Choi J, Lee M (2018) The skin cancer classification using deep convolutional neural network. Multimed Tools Appl 77(8):9909–9924. https://doi.org/10.1007/s11042-018-5714-1
26. Du C, Gao S (2017) Image segmentation-based multi-focus image fusion through multi-scale convolutional neural network. IEEE Access 5:15750–15761. https://doi.org/10.1109/access.2017.2735019
27. Duchi J, Hazan E, Singer Y (2011) Adaptive subgradient methods for online learning and stochastic optimization. J Mach Learn Res 12:2121–2159
28. Esteva A, Kuprel B, Novoa R, Ko J, Swetter S, Blau H, Thrun S (2017) Dermatologist-level classification of skin cancer with deep neural networks. Nature 542(7639):115–118. https://doi.org/10.1038/nature21056
29. Ge Y, Jiang S, Xu Q, Jiang C, Ye F (2017) Exploiting representations from pre-trained convolutional neural networks for high-resolution remote sensing image retrieval. Multimed Tools Appl 77(13):17489–17515. https://doi.org/10.1007/s11042-017-5314-5
30. Gibson E, Li W, Sudre C, Fidon L, Shakir D, Wang G, Eaton-Rosen Z, Gray R, Doel T, Hu Y, Whyntie T, Nachev P, Modat M, Barratt D, Ourselin S, Cardoso M, Vercauteren T (2018) NiftyNet: a deep-learning platform for medical imaging. Comput Methods Programs Biomed 158:113–122. https://doi.org/10.1016/j.cmpb.2018.01.025
31. Gogate M, Adeel A, Hussain A (2017) Deep learning driven multimodal fusion for automated deception detection. In: 2017 IEEE symposium series on computational intelligence (SSCI). https://doi.org/10.1109/ssci.2017.8285382

32. Golrizkhatami Z, Acan A (2018) ECG classification using three-level fusion of different feature descriptors. Expert Syst Appl 114:54–64. https://doi.org/10.1016/j.eswa.2018.07.030
33. Gong Y, Wang L, Guo R, Lazebnik S (2014) Multi-scale orderless pooling of deep convolutional activation features. In: Computer vision-ECCV 2014 392-407. https://doi.org/10.1007/978-3-319-10584-0_26
34. Guo Y, Liu Y, Oerlemans A, Lao S, Wu S, Lew M (2016) Deep learning for visual understanding: A review. Neurocomputing 187:27–48. https://doi.org/10.1016/j.neucom.2015.09.116
35. Gutman D, Codella N, Celebi E, Helba B, Marchetti M, Mishra N, Halpern A (2016) Skin lesion analysis toward melanoma detection: A challenge at the international symposium on biomedical imaging (ISBI) 2016, hosted by the International Skin Imaging Collaboration (ISIC). arXiv:1605.01397
36. Han L, Hu J, Li L (2013) Structure descriptor for articulated shape analysis. In: International symposium on visual computing 171-180. https://doi.org/10.1007/978-3-642-41914-0_18
37. Han S, Kim M, Lim W, Park G, Park I, Chang S (2018) Classification of the clinical images for benign and malignant cutaneous tumors using a deep learning algorithm. J Investig Dermatol 138(7):1529–1538. https://doi.org/10.1016/j.jid.2018.01.028
38. Havaei M, Davy A, Warde-Farley D, Biard A, Courville A, Bengio Y, Pal C, Jodoin P, Larochelle H (2017) Brain tumor segmentation with deep neural networks. Med Image Anal 35:18–31. https://doi.org/10.1016/j.media.2016.05.004
39. He K, Zhang X, Ren S, Sun J Deep Residual Learning for Image Recognition. In: 2016 IEEE conference on computer vision and pattern recognition (CVPR, vol 2016. https://doi.org/10.1109/cvpr.2016.90
40. Hou Q, Cheng M, Hu X, Borji A, Tu Z, Torr P Deeply supervised salient object detection with short connections. In: 2017 IEEE conference on computer vision and pattern recognition (CVPR), vol 2017. https://doi.org/10.1109/cvpr.2017.563
41. Huang G, Liu Z, Maaten L, Weinberger K, vol 2017, Densely connected convolutional networks. In: 2017 IEEE conference on computer vision and pattern recognition (CVPR). https://doi.org/10.1109/cvpr.2017.243
42. Huo L, Rao T, Zhang L (2018) Fused feature encoding in convolutional neural network. Multimed Tools Appl 78(2):1635–1648. https://doi.org/10.1007/s11042-018-6249-1
43. Ide H, Kurita T (2016) Low level visual feature extraction by learning of multiple tasks for convolutional neural networks. In: 2016 International joint conference on neural networks (IJCNN). https://doi.org/10.1109/ijcnn.2016.7727665
44. Jia Y, Shelhamer E, Donahue J, Karayev S, Long J, Girshick R, Guadarrama S, Darrell T (2014) Caffe: convolutional architecture for fast feature embedding. In: ACM international conference on multimedia (MM). https://doi.org/10.1145/2647868.2654889
45. Kingma D, Ba J (2015) Adam: A Method for Stochastic Optimization. 2015 International Conference on Learning Representations (ICLR). arXiv:1412.6980
46. Kleinberg R, Li Y, Yuan Y (2018) An alternative view: When Does SGD escape local minima?. In: 2018 International conference on machine learning (ICML). arXiv:1802.06175
47. Krizhevsky A, Sutskever I, Hinton G (2017) ImageNet classification with deep convolutional neural networks. In: Proceedings of the 25th international conference on neural information processing systems, vol 1, pp 1097–1105. https://doi.org/10.1145/3065386
48. Kuai Y, Wen G, Li D (2018) Hyper-Siamese network for robust visual tracking. SIViP 13(1):35–42. https://doi.org/10.1007/s11760-018-1325-6
49. Lacy K, Alwan W (2013) Skin cancer. Medicine 41(7):402–405. https://doi.org/10.1016/j.mpmed.2013.04.008
50. Lecun Y, Bottou L, Bengio Y, Haffner P (1998) Gradient-based learning applied to document recognition. In: Proceedings of the IEEE, vol 86, pp 2278–2324. https://doi.org/10.1109/5.726791
51. Lee H, Mendes A, Spolaôr N, Oliva J, Sabino Parmezan A, Wu F, Fonseca-Pinto R (2018) Dermoscopic assisted diagnosis in melanoma: Reviewing results, optimizing methodologies and quantifying empirical guidelines. Knowl-Based Syst 158:9–24. https://doi.org/10.1016/j.knosys.2018.05.016
52. Lee T, Ng V, Gallagher R, Coldman A, McLean D (1997) Dullrazor: A software approach to hair removal from images. Comput Biol Med 27(6):533–543. https://doi.org/10.1016/s0010-4825(97)00020-6
53. Li E, Xia J, Du P, Lin C, Samat A (2017) Integrating multilayer features of convolutional neural networks for remote sensing scene classification. IEEE Trans Geosci Remote Sens 55(10):5653–5665. https://doi.org/10.1109/tgrs.2017.2711275
54. Li F, Zurada J, Liu Y, Wu W (2017) Input layer regularization of multilayer feedforward neural networks. IEEE Access 5:10979–10985. https://doi.org/10.1109/access.2017.2713389
55. Li H, Weng J, Shi Y, Gu W, Mao Y, Wang Y, Liu W, Zhang J (2018) An improved deep learning approach for detection of thyroid papillary cancer in ultrasound images. Sci Rep 8(1):1–12. https://doi.org/10.1038/s41598-018-25005-7

56. Li Y, Charalampaki P, Liu Y, Yang G, Giannarou S (2018) Context aware decision support in neurosurgical oncology based on an efficient classification of endomicroscopic data. Int J Comput Assist Radiol Surg 13(8):1187–1199. https://doi.org/10.1007/s11548-018-1806-7
57. Li Y, Shen L (2018) Skin lesion analysis towards melanoma detection using deep learning network. Sensors 18(2):1–16. https://doi.org/10.3390/s18020556
58. Liu B, Wei Y, Zhang Y, Yang Q (2017) Deep neural networks for high dimension low sample size data. In: Proceedings of the 26th international joint conference on artificial intelligence. https://doi.org/10.24963/ijcai.2017/318
59. Liu M, Cheng D, Wang K, Wang Y (2018) Multi-modality cascaded convolutional neural networks for Alzheimer's disease diagnosis. Neuroinformatics 16(3/4):295–308. https://doi.org/10.1007/s12021-018-9370-4
60. Liu W, Wang Z, Liu X, Zeng N, Liu Y, Alsaadi F (2017) A survey of deep neural network architectures and their applications. Neurocomputing 234:11–26. https://doi.org/10.1016/j.neucom.2016.12.038
61. Lynn N, Kyu Z (2017) Segmentation and Classification of Skin Cancer Melanoma from Skin Lesion Images. In: 2017 18th international conference on parallel and distributed computing, applications and technologies (PDCAT). https://doi.org/10.1109/pdcat.2017.00028
62. Ma C, Huang J, Yang X, Yang M (2015) Hierarchical convolutional features for visual tracking. In: 2015 IEEE International Conference on Computer Vision (ICCV). https://doi.org/10.1109/iccv.2015.352
63. Ma C, Huang J, Yang X, Yang M (2018) Robust visual tracking via hierarchical convolutional features. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1–1. https://doi.org/10.1109/tpami.2018.2865311
64. Maas A, Hannun A, Ng A (2013) Rectifier nonlinearities improve neural network acoustic models. In: 2013 international conference on machine learning (ICML)
65. Majumder N, Hazarika D, Gelbukh A, Cambria E, Poria S (2018) Multimodal sentiment analysis using hierarchical fusion with context modeling. Knowl-Based Syst 161:124–133. https://doi.org/10.1016/j.knosys.2018.07.041
66. Mash R, Borghetti B, Pecarina J (2016) Improved aircraft recognition for aerial refueling through data augmentation in convolutional neural networks. In: Advances in visual computing 113-122. https://doi.org/10.1007/978-3-319-50835-1_11
67. Matsugu M, Cardon P (2004) Unsupervised feature selection for multi-class object detection using convolutional neural networks. In: International symposium on neural networks 864-869. https://doi.org/10.1007/978-3-540-28647-9_142
68. Menegola A, Fornaciali M, Pires R, Bittencourt F, Avila S, Valle E (2017) Knowledge transfer for melanoma screening with deep learning. In: 2017 IEEE 14th international symposium on biomedical imaging (ISBI 2017). https://doi.org/10.1109/isbi.2017.7950523
69. Mukkamala M, Hein M (2017) Variants of RMSProp and Adagrad with logarithmic regret bounds. In: 2017 international conference on machine learning (ICML). arXiv:1706.05507
70. Nickerson P, Tighe P, Shickel B, Rashidi P (2016) Deep neural network architectures for forecasting analgesic response. In: 2016 38th annual international conference of the ieee engineering in medicine and biology society (EMBC). https://doi.org/10.1109/embc.2016.7591352
71. Popescu D, Ichim L (2015) Image recognition in UAV application based on texture analysis. In: International conference on advanced concepts for intelligent vision systems 693-704. https://doi.org/10.1007/978-3-319-25903-1_60
72. Shie C, Chuang C, Chou C, Wu M, Chang E (2015) Transfer representation learning for medical image analysis. In: 2015 37th annual international conference of the ieee engineering in medicine and biology society (EMBC). https://doi.org/10.1109/embc.2015.7318461
73. Shin H, Roth H, Gao M, Lu L, Xu Z, Nogues I, Yao J, Mollura D, Summers R (2016) Deep convolutional neural networks for computer-aided detection: CNN Architectures, dataset characteristics and transfer learning. IEEE Trans Med Imaging 35(5):1285–1298. https://doi.org/10.1109/tmi.2016.2528162
74. Simonyan K, Zisserman A (2015) Very deep convolutional networks for large-scale image recognition. In: 2015 International conference on learning representations (ICLR). arXiv:1409.1556
75. Sun A, Lim E (2001) Hierarchical text classification and evaluation. In: Proceedings 2001 IEEE international conference on data mining. https://doi.org/10.1109/icdm.2001.989560
76. Sun J, Wan C, Cheng J, Yu F, Liu J (2017) Retinal image quality classification using fine-tuned CNN. In: International workshop on ophthalmic medical image analysis 126-133. https://doi.org/10.1007/978-3-319-67561-9_14
77. Tang J, Mat Isa N (2014) Adaptive image enhancement based on bi-histogram equalization with a clipping limit. Comput Electr Eng 40(8):86–103. https://doi.org/10.1016/j.compeleceng.2014.05.017

78. Taqi A, Awad A, Al-Azzo F, Milanova M (2018) The impact of multi-optimizers and data augmentation on tensorflow convolutional neural network performance. In: 2018 IEEE conference on multimedia information processing and retrieval (MIPR). https://doi.org/10.1109/mipr.2018.00032
79. Thao L, Quang N (2017) Automatic skin lesion analysis towards melanoma detection. In: 2017 21st Asia pacific symposium on intelligent and evolutionary systems (IES). https://doi.org/10.1109/iesys.2017.8233570
80. Toader M, Esanu I, Taranu T, Toader S (2017) Utility of polarized dermoscopy in the diagnosis of cutaneous lupus erythematosus and morphea. In: 2017 E-health and bioengineering conference (EHB). https://doi.org/10.1109/ehb.2017.7995491
81. Tripathi G, Singh K, Vishwakarma D (2018) Convolutional neural networks for crowd behaviour analysis: a survey. Vis Comput 35(5):753–776. https://doi.org/10.1007/s00371-018-1499-5
82. Tschandl P, Rosendahl C, Kittler H (2018) The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Scientific Data 5:1–9. https://doi.org/10.1038/sdata.2018.161
83. Tschandl P, Sinz C, Kittler H (2019) Domain-specific classification-pretrained fully convolutional network encoders for skin lesion segmentation. Comput Biol Med 104:111–116. https://doi.org/10.1016/j.compbiomed.2018.11.010
84. Vu T, Ho N, Yang H, Kim J, Song H (2018) Non-white matter tissue extraction and deep convolutional neural network for Alzheimer's disease detection. Soft Comput 22(20):6825–6833. https://doi.org/10.1007/s00500-018-3421-5
85. Wan M, Lai Z, Yang G, Yang Z, Zhang F, Zheng H (2017) Local graph embedding based on maximum margin criterion via fuzzy set. Fuzzy Set Syst 318:120–131. https://doi.org/10.1016/j.fss.2016.06.001
86. Wan M, Li M, Yang G, Gai S, Jin Z (2014) Feature extraction using two-dimensional maximum embedding difference. Inform Sci 274:55–69. https://doi.org/10.1016/j.ins.2014.02.145
87. Wan M, Yang G, Gai S, Yang Z (2015) Two-dimensional discriminant locality preserving projections (2DDLPP) and its application to feature extraction via fuzzy set. Multimed Tools Appl 76(1):355–371. https://doi.org/10.1007/s11042-015-3057-8
88. Wang X, Hua X, Xiao F, Li Y, Hu X, Sun P (2018) Multi-object detection in traffic scenes based on improved SSD. Electronics 7(11):1–28. https://doi.org/10.3390/electronics7110302
89. Wei Y, Zhao Y, Lu C, Wei S, Liu L, Zhu Z, Yan S (2016) Cross-modal retrieval with CNN visual features: A new baseline. IEEE Transactions on Cybernetics 47(2):449–460. https://doi.org/10.1109/tcyb.2016.2519449
90. Wen J, Ye F, Huang X, Li S, Yang L, Xiao X, Xie X (2015) Prognostic significance of preoperative circulating monocyte count in patients with breast cancer. Medicine 94(49):1–7. https://doi.org/10.1097/md.0000000000002266
91. Wu Z, Singh B, Davis L, Subrahmanian V (2017) Deception detection in videos. arXiv:1712.04415
92. Xu J, Zhang Z, Friedman T, Liang Y, Broeck G (2018) A semantic loss function for deep learning with symbolic knowledge. In: Proceedings of machine learning research (PMLR). arXiv:1711.11157
93. Yang X, Bian C, Yu L, Ni D, Heng P (2018) Class-balanced deep neural network for automatic ventricular structure segmentation. In: International workshop on statistical atlases and computational models of the heart 152-160. https://doi.org/10.1007/978-3-319-75541-0_16
94. Yang Y, Shah M (2014) Learning discriminative features and metrics for measuring action similarity. In: 2014 IEEE international conference on image processing (ICIP). https://doi.org/10.1109/icip.2014.7025311
95. Ye F, Pu J, Wang J, Li Y, Zha H (2017) Glioma grading based on 3D multimodal convolutional neural network and privileged learning. In: 2017 IEEE international conference on bioinformatics and biomedicine (BIBM). https://doi.org/10.1109/bibm.2017.8217751
96. Yosinski J, Clune J, Bengio Y, Lipson H (1792) How transferable are features in deep neural networks? In: International Conference on Neural Information Processing Systems. arXiv:1411.1792
97. Yu C, Yang S, Kim W, Jung J, Chung K, Lee S, Oh B (2018) Acral melanoma detection using a convolutional neural network for dermoscopy images. PLOS ONE 13(3):1–14. https://doi.org/10.1371/journal.pone.0196621
98. Yu D, Liu Y, Pang Y, Li Z, Li H (2018) A multi-layer deep fusion convolutional neural network for sketch based image retrieval. Neurocomputing 296:23–32. https://doi.org/10.1016/j.neucom.2018.03.031
99. Yu W, Yang K, Yao H, Sun X, Xu P (2017) Exploiting the complementary strengths of multi-layer CNN features for image retrieval. Neurocomputing 237:235–241. https://doi.org/10.1016/j.neucom.2016.12.002
100. Yu Z, Jiang X, Zhou F, Qin J, Ni D, Chen S, Lei B, Wang T (2018) Melanoma recognition in dermoscopy images via aggregated deep convolutional features. IEEE Trans Biomed Eng 66(4):1006–1016. https://doi.org/10.1109/tbme.2018.2866166

101. Zhang K, Guo L, Gao C (2018) Optimization method of residual networks of residual networks for image classification. In: 2018 IEEE international conference on big data and smart computing (BigComp). https://doi.org/10.1109/bigcomp.2018.00054

102. Zhang X, Wang S, Liu J, Tao C (2018) Towards improving diagnosis of skin diseases by combining deep neural network and human knowledge. BMC Med Inform Decis Mak 18(2):1–8. https://doi.org/10.1186/s12911-018-0631-9

103. Zhang Y, Muhammad K, Tang C (2018) Twelve-layer deep convolutional neural network with stochastic pooling for tea category classification on GPU platform. Multimed Tools Appl 77(17):22821–22839. https://doi.org/10.1007/s11042-018-5765-3

104. Zhao R, Ouyang W, Li H, Wang X (2015) Saliency detection by multi-context deep learning. In: 2015 IEEE conference on computer vision and pattern recognition (CVPR). https://doi.org/10.1109/cvpr.2015.7298731

105. Zheng L, Zhao Y, Wang S, Wang J, Tian Q (2016) Good practice in CNN feature transfer. arXiv:1604.00133

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

Ibtissam Bakkouri is a Ph.D student at Faculty of Sciences, Ibn Zohr University, Agadir - Morocco. Since October 2015, she has prepared her thesis in computer vision and medical imaging at Ibn Zohr University. Her research interests cover biomedical imaging, deep learning, and artificial intelligence.

Karim Afdel received his Doctorat from the University of Aix Provence France in Computer Engineering, Analysis and Medical Image Processing in 1994. Since 1995, he has been Professor at Ibn Zohr University, Faculty of Sciences, Agadir - Morocco. His research interests include medical image analysis and artificial intelligence.