

Malignant Melanoma Classification Using Cross-Platform Dataset with Deep Learning CNN Architecture



Soumen Mukherjee, Arunabha Adhikari and Madhusudan Roy

Abstract Melanoma, a malignant skin lesion, is the deadliest of all types of skin cancer. Deep learning has been shown to efficiently identify patterns from images and signals from various application domains. Use of deep learning in medical image analysis is, however, limited till date. In the present paper, two well-known malignant lesion image datasets, namely Dermofit and MEDNODE, are both separately and together used to analyze the performance of a proposed deep convolutional neural network (CNN) named as CNN malignant lesion detection (CMLD) architecture. When Dermofit and MEDNODE datasets are used separately with tenfold data augmentation, the CNN gives 90.58 and 90.14% classification accuracy. When the datasets are mixed together the CMLD gives only 83.07% accuracy. The classification accuracy of the MEDNODE dataset using deep CNN is considerably high in comparison with the results found in the related literature. The classification accuracy is also high in case of Dermofit dataset in comparison with the traditional feature-based classification.

Keywords ABCD rule · Image augmentation · MNIST · Convolutional layer · CMLD · MEDNODE · Dermofit · Stochastic gradient descent with momentum

S. Mukherjee

Department of Computer Application, RCC Institute of Information Technology, Kolkata, India
e-mail: soumou601@gmail.com

A. Adhikari (✉)

Department of Physics, West Bengal State University, Barasat, West Bengal, India
e-mail: arunabha.adhikari@gmail.com

M. Roy

Surface Physics and Material Science Division, Saha Institute of Nuclear Physics, Kolkata, India
e-mail: roy.madhusudan57@gmail.com

© Springer Nature Singapore Pte Ltd. 2019

S. Bhattacharyya et al. (eds.), *Recent Trends in Signal and Image Processing*,
Advances in Intelligent Systems and Computing 922,
https://doi.org/10.1007/978-981-13-6783-0_4

1 Introduction

Melanoma is the most [1] deadly type of skin cancer which requires surgery to cure in most of the cases [1]. In nearly 20% cases, surgery also fails to cure from this fatal disease. Out of all the skin cancer death, 75% happens for melanoma itself. It is the only cause for more than 50,000 cancer deaths around the world [1, 2]. Ultraviolet radiation exposure in bare skin is the leading cause of melanoma. In most of the European countries like Switzerland and Denmark, patients are diagnosed with melanoma at a much high rate. It is found from the literature that early diagnosis of melanoma plays a significant role for the successful treatment [1]. Melanoma is a skin lesion type which can only be identified by expert dermatologists because it is a close similarity with benign lesions. Dermatologists use several quantitative marking techniques like ABCD rule, 7-point checklist, 3-point checklist, etc. [3] other than the gold-standard skin biopsy to identify melanoma in an early stage. In recent year, computer vision, machine learning, and deep learning play a vital role in the detection of melanoma. In the present work, use of deep learning in malignant melanoma detection is carried out with a proposed CNN architecture with benchmark datasets. The paper is organized in the following way. In the related work section, two different approaches, i.e., identification of melanoma using the traditional feature and using deep learning CNN, are discussed. In the next section, a small introduction to deep learning is given. The proposed work description is given after that with dataset description and discussion on proposed CNN malignant lesion detection (CMLD) architecture. Finally, in the conclusion part, the importance of the work and the future scope is discussed.

2 Related Work

Melanoma and other malignant skin cancer detection and identification can be done using hand-crafted traditional features and also with deep learning features. In this section, both aspects are discussed in detail.

2.1 Melanoma Classification Using Traditional Feature

Malignant skin cancer detection using image processing and machine learning technique is primarily started with hand-crafted features, where a considerably large amount of traditional features like shape, size, color, texture, etc. are extracted from each lesion image. Then, a more computationally efficient subset of the extracted feature set is found with different feature ranking algorithms and feature optimization algorithms, and finally, different supervised and unsupervised classification algorithms are used on the feature subset to identify the proper class of the skin lesion

image. Quite a large number of researches are carried out using the hand-crafted features with well-known Dermofit and MEDNODE datasets which are used in this work. Using Dermofit dataset, Tan et al. [4] got 88% classification accuracy with optimized 1472 traditional features by a genetic algorithm with SVM classifier. Mukherjee et al. [5] used four features ranking algorithms and found selected 163 features which ranked very high in all four ranking. These features with multilayer perceptron (MLP) classifier yielded 86.2% classification accuracy. Ballerini et al. [6] achieved 93% with only non-melanoma skin cancer and benign lesion images and 74% accuracy using 960 images of Dermofit dataset with 2 class and 5 class, respectively, with several color features and 15,552 texture features using K-NN classifier with threefold cross-validation. Laskaris et al. [7] found 80.64% classification accuracy with 31 images of Dermofit dataset. MEDNODE dataset is also used by different researchers in this domain. Giotis et al. [8] got 81% accuracy with 675 features using the 170 images of MEDNODE dataset. Mukherjee et al. [9] got 83.33% classification accuracy with only ten highest ranked traditional features using MLP. The ranking of the features is done by ReliefF algorithm. In the same work, they have shown that with only the best 25 PCA components, the classification accuracy level can be increased by up to 87%. Mukherjee et al. [10] got 91.02% accuracy when they used 1886 features but optimized the number of neurons in the two hidden layers of MLP by particle swarm optimization (PSO) using the MEDNODE dataset. In their work, they have also worked with Dermofit dataset where they have found 88% accuracy with 1886 features. In conclusion, the highest accuracy achieved in full Dermofit (1300 images) and MEDNODE (170 images) dataset is 88 and 91.02% so far. Other than Dermofit and MEDNODE dataset, DermIS, DermQuest, PH2, etc. are the other popular datasets in this domain.

2.2 Melanoma Classification Using Deep CNN Architecture

In recent times, several researchers have worked on malignant lesion identification using deep learning architecture. Nasr-Esfahani et al. [11] in their work used MEDNODE dataset with 170 color images. They have augmented images to make it total 6120 including the original ones. Different image augmentation techniques were used like image cropping (5 and 10% cropping) and image rotation (with an angle 0° , 90° , 180° , and 270°) generating 35 augmented images from a single original image. Then, each image is rescaled to 188×188 pixels. The authors have used deep CNN to classify the images into two classes, malignant and benign. In their work, they have used two convolution layers with 20 and 50 feature maps of 5×5 kernel each in the first and second layer, respectively. One max-pooling layer is present between each convolution layer. Finally, fully connected two-class layer is given for the classification task. The system is implemented in Intel i7 processor with 32 GB RAM and with two linked GPU of Titan and GeForce series. They have maintained 80/20% training and testing ratio. They have achieved 81% classification accuracy. Pomponiu et al. [12] have proposed a method of deep CNN, where they used the images from

DermIS and DermQuest image library. Their original datasets are divided into 217 benign and 182 malignant images which are augmented to make it around 10,000 images. They introduced Gaussian noise, histogram equalization, and compression and motion blur to augment the images. The images are rescaled to 224×224 pixels to feed into the CNN. The CNN they have used has eight layers, which has five convolution layers at the beginning and three fully connected layers at the end. They have got a two-class classification accuracy of 93.64% in their work. Ayan and Unver [13] in their work have shown how data augmentation helps to improve the classification rate of melanoma using deep learning. In their work, they have used different sized 1000 original images from popular ISIC dataset which are divided into two equal set of 500 images of malignant and benign class. All the images are resized to 224×224 pixels before feeding to the CNN. The CNN architecture they have used has total 11 layers with four convolution layers, two max-pool layers, three dropout layers, and two fully connected dense layers. They have achieved 78% classification accuracy without data augmentation, and when data augmented with five times of the original images to make the total image count up to 5000 with 75/25% training and testing ratio, the two-class classification accuracy increased to 81% with NVIDIA K4200 GPU card and 64 GB RAM. The entire program of their proposed system is implemented in Python. Kwasigroch et al. [14] made a comparison of the performance of three popular CNN architectures in melanoma classification including VGG19, ResNet50, and VGG19-SVM. In their work, they have used ISIC dataset with different types of image augmentation scheme like transformation (i.e., rotation, flip, shift, etc.). They have used an imbalanced dataset of 9300 benign and 670 malignant images. With VGG19, ResNet50, and VGG19-SVM, they have achieved 81.2, 75.5, and 80.7% classification accuracy. They have used GeForce GPU with 6 GB RAM in their work, and the program implemented in Python. Maia et al. [15] got 92.5% accuracy with PH2 dataset consisting of 200 images (40 malignant and 160 benign) using popular VGG19 CNN architecture. Lopez et al. [16] also used ISBI 2016 challenge dataset with training and testing images of 900 and 379 images and got 81.33% classification accuracy. They have used VGG16 architecture for their deep CNN network. Ali and Al-Marzouqi [17] have used ISBI 2016 challenge dataset with 727 benign and 173 malignant images to achieve 81.6% two-class classification accuracy with modified LightNet architecture, consisting of 17 layers including five blocks. Every block has convolution, ReLU, max-pooling, and dropout layer. Last block has fully connected and soft-max classification layer in this deep CNN.

3 Deep Learning

Deep learning is the most popular technique for image recognition in recent times for its efficiency and accuracy. It is not only used in the domain of image processing but also in natural language processing, bioinformatics, audio processing, etc. Deep learning application in image processing acts very similar to the visual cortex of the brain and finds out deep features from the images. Convolutional neural network

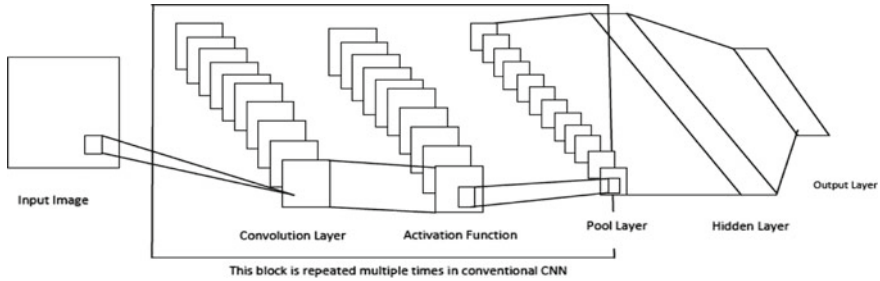


Fig. 1 Deep CNN architecture

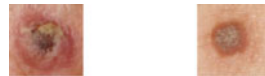
(CNN) is one of the mostly used deep learning techniques [18–21]. It is first developed in 1980s, but around 2012 the use of CNN in deep learning application comes into reality for the development of graphics processing unit (GPU) as CNN requires high computational time. The main area of concern is the training time required by CNN, which is considerably very high in complex CNN rather than traditional techniques. CNN has two major components, convolutional layer and pooling layer. Each convolutional layer produces a new image from the original image known as feature map. Convolutional layer contains small square shape convolutional filters which are applied to the input image. The size of the filter used varies and depends on different types of application. After convolutional layer, a batch normalization layer is present which normalizes the output of the previous convolutional layer. Activation function like ReLU (rectified linear unit) or sigmoid function is used at the end of the convolutional layer. The pooling layer is used to reduce the size of the feature map generated from the convolutional layer. There are different types of pooling like max-pooling, mean-pooling, and average pooling which are also application-specific. Filter shift step is known as stride. In the last stage of CNN, one or few fully connected layers, usually soft-max layer of classification, are present. Figure 1 shows small sample deep CNN architecture. More complex architecture has multiple numbers of convolutional and pooling layers. Different convolutional neural networks (CNNs) are used for the image classification task, e.g., AlexNet, VGGNet, GoogLeNet, ResNet, CIFAR10/100, MNIST, SVHN, etc. The CNN architecture varies for different types of application [18, 19].

4 Dataset

In the present work, two different popular malignant lesion image datasets are used. The first one is Dermofit [6] image set, which contains 1300 malignant (both melanoma and non-melanoma) and benign skin lesion optical images. Every image of this dataset is tagged by dermatopathologists, and it is identified as a gold-standard dataset. Images in this dataset are in “.png” color format with varying image pixel

Table 1 Details of the two datasets used in the work

Dataset	Total images	Malignant	Benign	Format	Size
Dermofit	1300	526	774	.png color	Varying size
MEDNODE	170	70	100	.jpg color	Varying size

Fig. 2 a, b Malignant and benign skin lesion images from Dermofit dataset

(a) Malignant

(b) Benign

size. The second dataset used in the work is MEDNODE [8] dataset, which has total 170 number of “.jpg” color malignant (melanoma) and benign optical images. MEDNODE is collected by University Groningen. Nikon D3 or D1x cameras are used to take these images. Every image of this dataset is validated by a dermatologist and identified as a gold-standard dataset. Details of the two datasets are given in Table 1. In Fig. 2, the malignant melanoma and benign skin lesions are shown from Dermofit dataset.

5 Proposed Work

In this proposed work, MATLAB 2018a is used for writing code for image augmentation and for creating proposed deep learning CNN architecture. The codes are executed in Intel i5 processor with 4 GB RAM without GPU. The original images are augmented with replication of images by making three types of transformations, namely horizontal flip, vertical flip, and both horizontal and vertical flip together and introducing four types of noise, i.e., Gaussian, Poisson, Salt-and-Pepper, and speckle in the original image. Two other ways of augmentation done are increment and decrement of brightness by a value of 50 in the original image. This image augmentation makes malignant images to 5260 (526×10) images and benign image to 7740 (774×10) images for Dermofit dataset. Similarly, MEDNODE dataset [8] original images (170) are also extended by ten times (1700) of the original dataset after data augmentation. Figure 3 shows the results of different transformations on one of the single malignant melanoma images from MEDNODE dataset.

6 CNN Malignant Lesion Detection (CMLD) Architecture

In the present work, a new CNN malignant lesion detection (CMLD) architecture is proposed which is lightweight and similar to the MNIST deep CNN architecture [22]. Several deep CNN architectures are available at present like AlexNet, VGGNet,

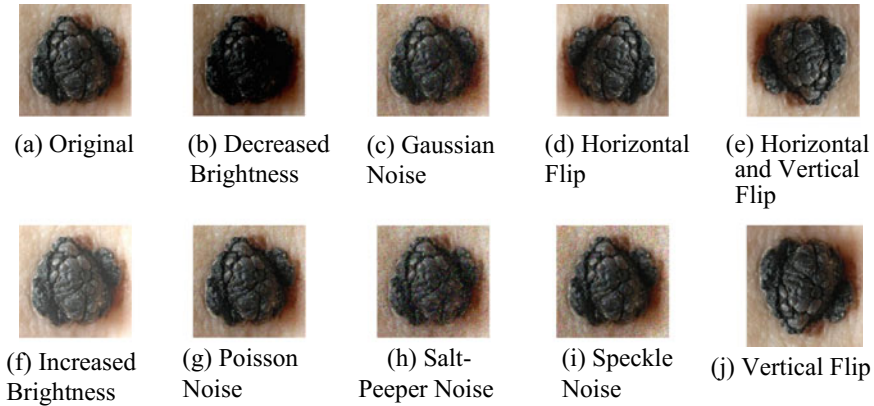


Fig. 3 a–j Original and different types of augmented images of MEDNODE melanoma dataset

ResNet, GoogleNet, etc. Each of them requires a considerable amount of time and space overhead and hence may be called heavyweight. Some CNN architectures require less computational load and branded as lightweight. In comparison with the lightweight CNNs like CIFAR10/100, MNIST, SVHN, etc., the above-mentioned heavyweight CNNs give very small enhancement in classification accuracy. In this proposed CNN malignant lesion detection (CMLD) architecture after the input image layer in the top layer, three 2D convolution layers are present and separated by two max-pool layers in between, and finally, fully connected layers and soft-max layer are present for classification. The detail of the architecture is given in Table 2. Architecture of CMLD is similar to MNIST deep CNN was originally developed to classify a 10-class (digits of 0–9) problem of handwritten digit identification.

In this proposed work, the authors have used input image layer of size 128×128 to minimize data loss unlike 28×28 input image size used in MNIST deep CNN architecture. Increase in image size will increase the training time of the deep CNN. Yu et al. [23] have shown that increasing the input image size will increase the complexity of the network but may not increase the classification accuracy.

For the present work, 75% of the total images are taken for training, and 25% of images are taken for testing. The training options used for the whole experiment are given in Table 3.

7 Result and Discussion

Two datasets, Dermofit and MEDNODE, are separately classified with the deep learning CNN architecture. Using augmented MEDNODE (1700 images) and Dermofit (13,000 images) datasets, an accuracy of 90.14 and 90.58% are found. When these two datasets are merged together, i.e., then a total image set of 14,700 gives

Table 2 Details of the layers present in CNN malignant lesion detection (CMLD) architecture from top to bottom

[128 × 128] Input image layer
Two-dimensional convolve layer with eight number of [3 × 3] filters with padding so that output size is same like input
Batch norm layer
ReLU layer
Two-dimensional max-pool layer with [2 × 2] pool and [2 × 2] stride (step in CNN)
Two-dimensional convolve layer with 16 number of [3 × 3] filters with padding so that output size is same like input
Batch norm layer
ReLU layer
Two-dimensional max-pool layer with [2 × 2] pool size and [2 × 2] stride (step in CNN)
Two-dimensional convolve layer with 32 number of [3 × 3] filters with padding so that output size is same like input
Batch norm layer
ReLU layer
Fully connected layer (2)
Soft-max layer
Classify layer

Table 3 Showing different values of the options chosen for training CMLD

Method	Learning rate	Max. epochs	Validation frequency
Stochastic gradient descent with momentum	0.01	500	100

Table 4 Results found from the CMLD architecture

Proposed scheme	MEDNODE dataset (1700 images) (%)	Dermofit dataset (13,000 images) (%)	MEDNODE and Dermofit dataset both (14,700 images) (%)
Accuracy	90.14	90.58	83.07

an accuracy of 83.07%, which is much lesser than the accuracy of the individual dataset. The results are given in Table 4. In Table 5, comparison of different results found from the present work and related work are shown.

Table 5 Comparison of results with other related works using two datasets

Results found with hand-crafted features			
Dermofit dataset		MEDNODE dataset	
Related work	Accuracy (%)	Related work	Accuracy (%)
Tan et al. [4] with 1472 features	88	Giotis et al. [8] with 675 features	81
Mukherjee et al. [5] with 163 features	86.2	Mukherjee et al. [10] with 1886 features	91.02
Mukherjee et al. [10] with 1886 features	88	Mukherjee et al. [9] with 10 features	83.33
		Mukherjee et al. [9] with 25 PCA components	87
Results found with deep learning architecture			
Proposed work	90.58	Nasr-Esfahani et al. [11]	81
		Proposed work	90.14 (without GPU)

8 Conclusion

It is given in Table 5 that deep learning gives 2.5% better accuracy result in case of Dermofit and 1% lesser accuracy in MEDNODE in comparison with the best results obtained from hand-crafted features. The classification accuracy of the present CMLD fares much better on MEDNODE than CNN used by Nasr-Esfahani et al. [11] although no GPU is used in CMLD and two linked GPUs are used by Nasr-Esfahani et al. When the mixed platform datasets are used, the classification accuracy drops through around 7% as it is commonly seen with hand-crafted features. It is a common belief that data augmentation improves classification accuracy in deep CNN [13], but the present work shows that when instead of synthetic data replication data is augmented from another dataset deep CNN does not always perform better and accuracy may decrease. It also shows deep CNN is not always superior to hand-crafted features in terms of classification accuracy but definitely is much more computationally expensive. Opportunity for the traditional machine learning approaches using the hand-crafted features and image preprocessing to find out an algorithm for higher accuracy and search for a feature set which may work across the datasets is still open. On the other hand, CNNs may have better classification accuracy if more complex network similar to ImageNet, GoogleNet, AlexNet is applied and more augmented data being introduced in the network. In conclusion, research for both the approach is still worth perusing.

References

1. Corrie P, Hategan M, Fife K, Parkinson C (2014) Management of melanoma. *Br Med Bull* 111(1):149–162. <https://doi.org/10.1093/bmb/ldu019>
2. Riker AI, Zea N, Trinh T (2010) The epidemiology, prevention, and detection of melanoma. *Ochsner J* 10(2):56–65
3. Marks R (2000) Epidemiology of melanoma. *Clin Exp Dermatol* 25:459–463. <https://doi.org/10.1046/j.1365-2230.2000.00693.x>
4. Tan TY, Zhang L, Jiang M (2016) An intelligent decision support system for skin cancer detection from dermoscopic images. In: 12th International conference on natural computation, fuzzy systems and knowledge discovery (ICNC-FSKD), pp 2194–2199
5. Mukherjee S, Adhikari A, Roy M (2018) Malignant melanoma identification using best visually imperceptible features from Dermofit dataset. Accepted and presented in 1st international conference on emerging trends in engineering and science (ETES-2018)
6. Ballerini L, Fisher RB, Aldridge RB, Rees J (2013) A color and texture based hierarchical K-NN approach to the classification of non-melanoma skin lesions, color medical image analysis. In: Celebi ME, Schaefer G (eds) *Lecture notes in computational vision and biomechanics*, vol 6, pp 63–86
7. Laskaris N, Ballerini L, Fisher RB, Aldridge B, Rees J (2010) Fuzzy description of skin lesions. In: *SPIE conference proceedings*, vol 7627. <https://doi.org/10.1117/12.845294>
8. Giotis I, Molders N, Land S, Biehl M, Jonkman MF, Petkov N (2015) MED-NODE: a computer-assisted melanoma diagnosis system using non-dermoscopic images. *Expert Syst Appl* 42:6578–6585
9. Mukherjee S, Adhikari A, Roy M Malignant melanoma detection using multi layer perceptron with visually imperceptible features and PCA components from MED-NODE dataset. *Int J Med Eng Inf Inderscience (UK) Enterprise Ltd. Publishers (in press)*
10. Mukherjee S, Adhikari A, Roy M (2018) Melanoma identification using MLP with parameter selected by metaheuristic algorithms, intelligent innovations in multimedia data engineering and management. Published from IGI Global, pp 241–268. ISBN 9781522571070
11. Nasr-Esfahani E, Samavi S, Karimi N, Soroushmehr SMR, Jafari MH, Ward K, Najarian K (2016) Melanoma detection by analysis of clinical images using convolutional neural network. In: 38th Annual international conference of the IEEE engineering in medicine and biology society (EMBC), pp 1373–1376
12. Pomponiu V, Nejati H, Cheung NM (2016) DEEPMOLE: deep neural networks for skin mole lesion classification. *IEEE ICIP*, pp 2623–2627
13. Ayan E, Unver HM (2018) Data augmentation importance for classification of skin lesions via deep learning. In: *Electric electronics, computer science, biomedical engineering's meeting (EBBT)*, Istanbul, pp 1–4
14. Kwasigroch A, Mikolajczyk A, Grochowski M (2017) Deep neural networks approach to skin lesions classification—a comparative analysis. In: 2017 IEEE conference, pp 1069–1074
15. Maia LB, Lima A, Pereira RMP (2017) Evaluation of melanoma diagnosis using deep features. In: 2017 IEEE conference, pp 1069–1074
16. Lopez AR, Giro-i-Nieto X, Burdick J, Marques O (2017) Skin lesion classification from dermoscopic images using deep learning techniques. In: *Proceedings of the IASTED international conference biomedical engineering (BioMed 2017)*, Innsbruck, Austria, pp 49–54
17. Ali AA, Al-Marzouqi H (2017) Melanoma detection using regular convolutional neural networks. In: 2017 International conference on electrical and computing technologies and applications (ICECTA). IEEE, pp 1–5
18. Voulodimos A, Doulamis N, Doulamis A, Protopapadakis E (2018) Deep learning for computer vision: a brief review. *Hindawi Comput Intell Neurosci* 2018(Article ID 7068349):13. <https://doi.org/10.1155/2018/7068349>
19. Pak M, Kim S (2017) A review of deep learning in image recognition. In: 4th International conference on computer applications and information processing technology (CAIPT), Kuta Bali, pp 1–3

20. Patterson J, Gibson A (2017) Deep learning: a practitioner's approach, 1st edn. O'Reilly, p 507
21. Buduma N (2017) Fundamentals of deep learning, 1st edn. O'Reilly, p 283
22. Hasanpour SH, Rouhani M, Fayyaz M, Sabokrou M (2016) Let's keep it simple, using simple architectures to outperform deeper and more complex architectures. [arXiv:1608.06037](https://arxiv.org/abs/1608.06037)
23. Yu Z, Jiang X, Zhou F, Qin J, Ni D (2018) Melanoma recognition in dermoscopy image via aggregated deep convolutional features. IEEE Trans Biomed Eng. <https://doi.org/10.1109/tbme.2018.2866166>