A Systematic Analysis and Design of Skin Melanoma Detection System using Machine Learning

Husam K Salih Juboori[1], Mohanad F Jwaid[2], Mohammed Alaa H. Altemimi[3]

[1]Pharmacy Department, Al Rasheed University College, Baghdad, Iraq
[2]Assistant Lecturer, Al Immam University College, Balad, Iraq.
[3]Computer Techniques Engineering Dept, Al Rasheed University College, Baghdad, Iraq.
Email:Husam.K.Salih@alrasheedcol.edu.iq

Issue Details
Issue Title: Issue 1
Received: 15 January, 2021
Accepted: 08 February, 2021
Published: 31 March, 2021
Pages: 3573 – 3582

Abstract

Now a days, skin cancer is well known reason for human death. abnormal skin cells growth is known as skin cancer ,these skin cells generated on human body which exposed to the sunlight, it can generate anywhere on the human body. At early stage, most of the cancers are curable. Hence, it is required to detect skin cancer at early stage to save patient life. It is possible to recognize skin cancer at early stage with advanced technology. We describe a novel framework for dermoscopy image identification that employs a deep learning algorithm and an objects encoding scheme. The deep representations of a rescaled dermoscopy picture, in particular, are first extracted using an unusually deep residual neural network that has been pre-trained on a large natural picture dataset. After that, local deep descriptors are gathered via order less visual statistic characteristics, which are then used to generate a global picture representation using fisher vector encoding. Finally, using a convolution neural network, we used the fisher matrix encoded models to organize melanoma photos (CNN). With limited training data, our suggested system may produce additional discriminative information to deal with big differences among melanoma groups as well as minor differences between melanoma and non-melanoma classes.

Keywords

Dermoscopic Image Recognition, CNN Algorithm, Melanoma Detection, Segmentation

1. Introduction

Even experienced dermatologists find it difficult to forecast skin lesions due to minor differences in surrounding skin and injuries, apparent similarity amongst skin sores, stupefied lesion outskirts, and other factors. A technological and mechanical detection method with provided photos can assist physicians in diagnosing malignant skin lesions at an early stage. Dilated convolution, a deep learning innovation, is reported to have improved accuracy while using the same amount of processing difficulties as regular CNN.

To give proper treatment recognition of skin lesion is important. Hence, the survival rate is increased due to early recognition of melanoma in dermoscopic images. The accurate detection of melanoma skin lesions is possible to highly trained dermatologists. Therefore, it is very challenging task to detect melanoma due to very close similarity among skin and lesion, and similarity which is visible between the non-melanoma and melanoma lesions, etc. Due to the less expertise is in available, efficient and accurate automatic system for detection of skin melanoma i.e. a system that can automatically analyses and detects the skin cancer, will be really advantageous as far as enhancing the accuracy and efficiency of pathlabs.

Overall, this paper presents a methodology for locating the problems for automatic and reliable melanoma diagnosis in dermoscopy pictures in order to address these concerns. This research makes two important contributions. An efficient framework was presented based on the deep CNN and higher than the corresponding technique. It is beneficial to build suitable parameters for some more reliable melanoma recognition when training data is low.

Segmentation is used to extract the feature, and CNN is used to analyze melanoma and non-melanoma pictures in the suggested framework. Then, using the open ISBI 2016 skin lesion data, a series of tests were conducted to compare the proposed strategy to existing CNN-based approaches.

The organization of this document is as follows. In Section 2 (Literature Survey), we enlisted details of all research existed. In Section 3 (System Architecture), present architecture of proposed system. In section 4 (Result and Discussion) discussed experimental setup and dataset used is discussed. In section 5 (Conclusion) discussed conclusion and future work required to improve our system.

2. Literature Survey

In this study [1], author used the MED-NODE dataset. From the segmented part colour feature are extracted and for performance checking of system three classifiers are used by the author viz.KNN, Decision Tree, and Naïve Bayes. Proposed methodology consists of following steps:

1. Preprocessing: The digital images contains artifacts .Thresholding algorithm is used for removing artifacts.
2. Segmentation: Here, the segmentation algorithm has been utilized to find the region of interest.
3. Feature Extraction: Colour feature are extracted from segmented part
4. Classification: For classify the image whether it is melanoma or benign author used three classifiers.
5. The accuracy of 81.95% has been achieved by decision tree algorithm and it is greater as compared to the other two algorithms.

In this study, Andre Esteva et al. [2] introduced a completely automated skin lesion segmentation methodology based on trained which is a 19-layer deep learning model, convolutional neural network (CNNs) is used. No conceptual understanding of the data is required to use this model. They employ a range of tactics to enhance information learning, despite the reality that they would have little training data. Because of the considerable imbalance in the quantity of background and rear pixels that occurs when cross entropy is used as the error function for management discussion and analysis, an innovative probability distribution based on Cosine similarity is designed to reduce the necessity for sample re-weighting. The author employed two publicly accessible databases, namely ISBI 2016 and PH2, to measure the efficacy, efficiency, and generalization potential of the proposed framework. According to the author's experiments, using these two datasets, the suggested strategy outperforms existing state-of-the-art approaches.

The author developed a Convolutional Neural Network model for melanoma classification in this research study [3], as well as comparing its performance to current models. According to Author, the architecture they devised is basic and only requires a few parameters to work. We show in this paper that by employing a typical convolutional neural network with minimal parameters, the suggested system may obtain equivalent results in terms of accuracy and specificity. For categorizing the ISBI 2016 challenge dataset, Lequan Yu et al. created a model with and without a segmentation module. A convolutional neural network is utilized to offer precise segmentation. Softmax classifier and support vector machine classifier are the two classifiers utilized here for categorization. This classifier determines the average results. Data augmentation is applied to the input picture in the form of rotations and shifts. The results of categorization with and without segmentation are nearly identical, according to the report. The classification accuracy is determined to be 85.5 percent with segmentation and 82.8 percent without segmentation.

In this research [4], author proposed a new method for recognition of melanoma using convolutional neural networks (CNNs) and it is compared with existing methods. Author says that their system, substantially deeper networks can accomplish prosperous and more discriminative features for more accurate and exact recognition. To take advantage of deeper systems, the author presented a number of strategies for ensuring successful preparation and learning in the face of little training data. The following advancements are used in the technique:

a. When a network expands, use the remaining time to figure out how to adjust to the fraud and overfitting issues. By increasing network depth, it will ensure that the presentation is successful.
b. Create a fully convolutional residual network (FCRN) for exact skin lesion segmentation, then consolidate a multi-scale contextual information integration approach to boost its capacity.
c. Finally, to create a two-stage framework, coordinate the suggested FCRN (for segmentation) and back propagation net-works (for classification).

This methodology enables the classifying network to extract more representative and explicit feature extraction methods on segmented results rather than the entire dermoscopy picture, minimizing the shortage of training data even more. The Melanoma Detection Challenge dataset was evaluated using ISBI 2016 Skin Lesion Analysis, according to the study's author.

Here [5], the deep learning system is implemented on computer with GPU. The clinical images are used instead of dermoscopic images. The input clinical images contains noise effects and illumination, these effects are pre-processed to enhance the images using pre-processing. These images are fed to CNN classifier(Convolutional Neural Network) for the classification purpose. Proposed System is consist following methodology:

1. Preprocessing: The clinical images taken by digital cameras contains noise and illumination. In this technique these effects are reduced to enhance the image.
2. CNN Proposed: After noise reduction, the clinical pictures from the training dataset are put into the proposed CNN. Melanoma is detected in clinical samples using the CNN approach (non-dermoscopic images).
3. The results reveal that the suggested strategy outperforms existing approaches through respect of predictive performance.

The author of this paper [6] presented a deep Siamese CNN (SCNN) architecture, which is trained using just binary image pair information and requires less supervision to learn picture representations. The fundamental goal is to confine the majority of experiments to a single trained convolutional neural network approach (CNN). Using an accessible to the public multiclass retinal fundus picture dataset, the authors evaluate the learnt picture descriptions on a task of bandwidth medical picture retrieval. The findings of the experiment reveal that the author's system, i.e. deep SCNN, is equivalent to current single supervised CNNs and requires significantly less supervision during training.

In this paper [7], the author used the support vector machine for classification. For accurate and fast evaluation of lesion the automatic image analysis tool provides the techniques like non- invasive medical computer vision or medical image processing. Proposed System involves following steps:

1. Collect the image dataset. This images are captured by using dermoscopy.
2. Preprocessing
3. Here, the Thresholding has been used for segmentation.
4. Gray level co-occurrence matric(GLCM) and ABCD(Asymmetry, border, colour, diameter) rule, etc. is used for Statistical feature extraction.
5. PrincipalComponent Analysis(PCA) is used for the selection of the feature.
6. The exact dermoscopy score will be calculated and thenSupport vector machine will be used for classification.
7. 92.1 % accuracy is achieved by using this classification methodology.

The author of this publication [8] discusses the potential of developing a global skin disease diagnostics system using Deep Convolutional Neural Networks (CNN). The system initially trained the CNN architecture using 23,000 skin disease photos from the Dermnet dataset, then assessed its performance using photos from the Dermnet and OLE datasets. When tested on the Dermnet dataset, experimental findings suggest that the suggested methodology can achieve as high as 73.1 percent Top-1 accuracy and 91.0 percent Top-5 accuracy. Top-1 and Top-5 accuracies are 31.1 percent and 69.5 percent, respectively, in the OLE dataset test. The author also suggests that if additional training photos are used, accuracies can be improved even more.

The spread of melanoma occurred via metastasis hence it is proven that to be very fatal. As per statistical proof or evidences the most of deaths occurred due to skin cancer are as a result of melanoma.

For discovering clusters in huge spatial data bases with noise, the DBSCAN clustering technique is used. In this work [9], a pre-processing phase was added to a slightly different version of a well-known density-based clustering technique, DBSCAN. This method is called fast density-based lesion detection (FDBLD), and it eliminates repetitive calculations in DBSCAN by choosing questioning locations and core points carefully. The goal of this study is to improve the performance of the prediction for detecting lesion borders in dermoscopy pictures by using FDBLD. The pre-processing stage is extremely important for FDBLD. The primary goal of this research is to eliminate FDBLD reliance in the pre-processing stage, allowing color information to be used, and to improve the accuracy of the findings. FDBLD has been updated to include a new distance metric to do this.

This study [10] revealed a unique hair-restoration algorithm that can preserve skin lesion traits including color and texture while also segmenting both dark and light hairs. The proposed approach is built on three fundamental steps:

a) Segmented the rough hairs using a form of commitment with this first Gaussian derivation (MF-FDOG) with Thresholding, which resulted in greater resistance for both light and darkness hairs.
b) Hairs were treated utilizing morphological edge-based techniques before being corrected with a rapid marching in painting technique.
c) The precision of the diagnosis (DA) and the method for texture-quality measurements (TQM) relies on hand-made skin masks dermatologists which are used as the basis for an evaluation of system performance.

The suggested method is very accurate, resilient, and capable of restoring hair pixels without causing damage to the lesion texture, according to their testing results. This method is completely automated and simple to include into a CAD system.

The author of this study [11] advocates using a more comprehensive perspective in order to give a better starting point for developing a comprehensive grasp of Deep Learning. In particular, this study aims to give a more detailed examination of the most essential components of Deep Learning, as well as any current advancements in the area. This research covers, in specific, the importance of deep learning and the different sorts of deep learning methodologies and structures. It then introduces new networks (CNNs), the sort of Deep Learning network most commonly employed, and explores the development of CNN design and its essential features, e.g. starting with AlexNet and concluding in Elevated Network (HR.Net).

The author proposes a unique transfer learning strategy in this research [12] to solve the aforementioned problems by first training the deep learning model on huge unlabeled medical picture datasets and then transferring the information to train the deep learning model on a limited number of tagged medical pictures. The author also proposes a novel deep learning model (DCNN) model that incorporates current breakthroughs in the field.

3. System Architecture

Following fig. 1 shows the proposed framework design. The system includes various modules such as input dataset, pre-processing, image segmentation, feature extraction, training and testing using cnn algorithm. In our work we used CNN algorithm for classification of melanoma and non-melanoma images.

Following are the steps involves in execution of our proposed system.

1. Input image dataset to the system.
2. Pre-processing is performed to enhance image quality and removal of hairs from image.
3. Several features are extracted from input image dataset from which training file is generated.
4. Generated training file dataset and new test input images are pass to CNN classification algorithm.
5. The output of CNN algorithm is melanoma detection i.e the input test show melanoma or not.
6. At the end graphical evaluation is perform to check the performance of proposed system.

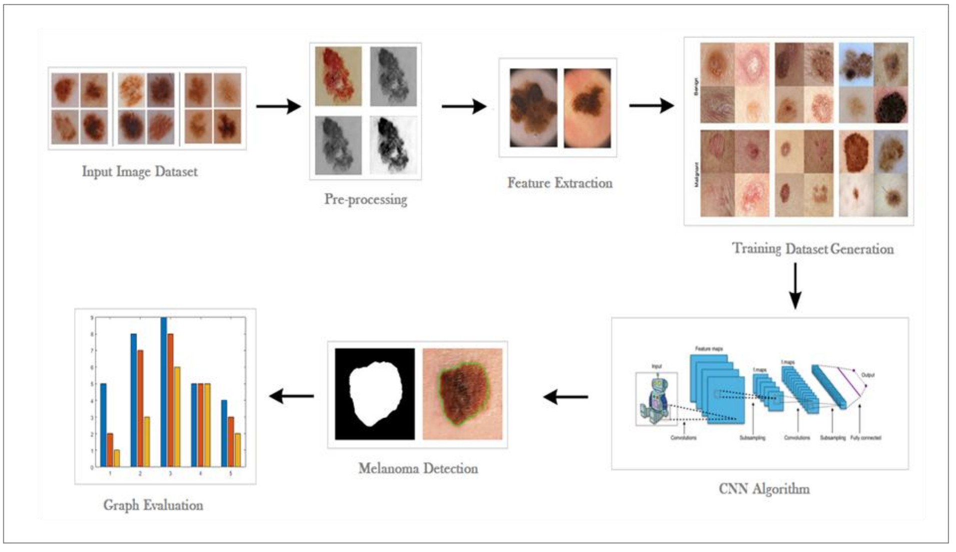

Figure 1. System Architecture

ISSN: 0304-2294

3.1 Mathematical Formulation

System S is represented as

S = {ID, P, F, T, CNN, M}

1. Input Dataset

ID = $\{\delta_1, \delta_2, \delta_3 \ldots \delta_n\}$

Where ID is the input image dataset and $\delta_1, \delta_2 \ldots \delta_n$ are the number of images.

2. Preprocessing

PR = {pr1, pr2, pr3}

PR is preprocessing and pr1, pr2 and pr3 are the steps to be carried out during preprocessing.

pr1 be the reading of input dataset

pr2 be the enhancement of image input and

pr3 be the removal of hair from image.

3. Feature Extraction

$\beta = \{\beta_1, \beta_2, \beta_3 \ldots \beta_n\}$

Where F is the set of features extracted from the image and $\beta_1, \beta_2, \beta_3 \ldots \beta_n$ are the extracted features such as border, thickness, color, etc.

4. Training and Testing file generation

T = $\{T_1, T_2\}$

Where T is the set of Training and Testing file and T_1 is Training file and T_2 is Testing file both the files contains various extracted features values while training file contains class of each image as 0 or 1.

5. Convolutional Neural Network (CNN).

CNN = {C, RL, PO, FC, LS}

Where CNN is algorithm consisting of various stages as

C is convolutional operation

RL be the ReLU activation layer

PO be the Pooling layer

FC be the Full Connection layer and

LS be the Loss function.

6. Melanoma Detection

M= {0,1}

M is the set of Class having value 0 or 1

0 be the absent of Melanoma and

1 be the present of Melanoma

4. Result Analysis

4.1 Dataset / Database used

For validating, testing, and training we used the ISIC[11] dataset, which consists of 10015 dermoscopic pictures of seven skin lesion classes with significant class imbalances.

4.2 Experimental Setup

All of the experiments are written in Python using Anaconda (Juypter) tools, methods, and strategies, as well as a competing classification method using feature extraction techniques, and run in a system with a Core I7, 2.30 GHz Windows 10 (64 bit) PC with 8GB of RAM.

4.3 Result

Fig. 2 shows the performance analysis Graph. We summarize the performance result of the three machine learning methods in term of spam sensitivity and specificity. Table 1 shows the reading from which fig.2 graph is plot. In terms of accuracy, the SVM approach is the most accurate, while the k nearestNeighbor approach gives us a lesser percentage, but in terms of spam precision, the suggested CNN algorithm has superior accuracy than the Convolution Neural Network and Support Vector Machine with softmax algorithms. In following graphs x-axis show different classification algorithms while y-axis show percentage.

Table 1. Performance parameters reading

Algorithms	AC	SE	SP
CNN (Softmax)	0.850	0.500	0.934
SVM	0.844	0.520	0.824
CNN (Proposed)	0.939	0.507	0.854

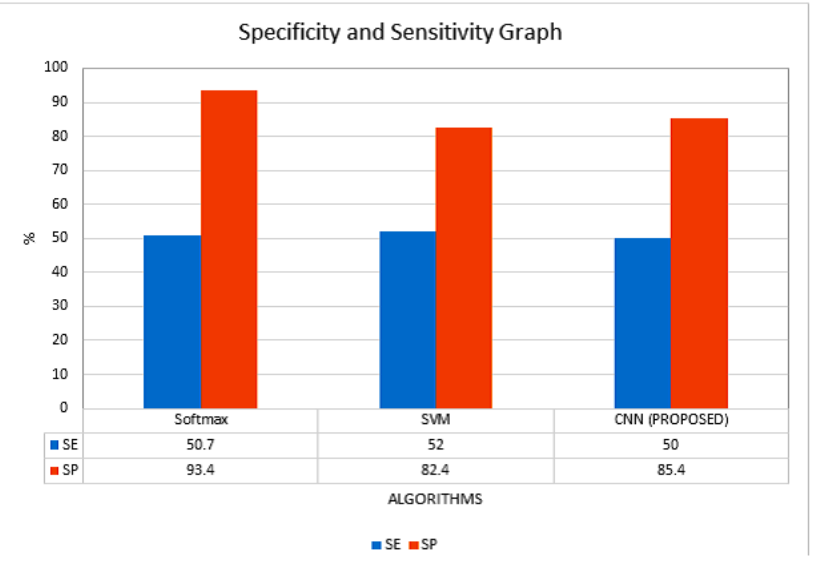

Figure 2: Specificity and Sensitivity Graph

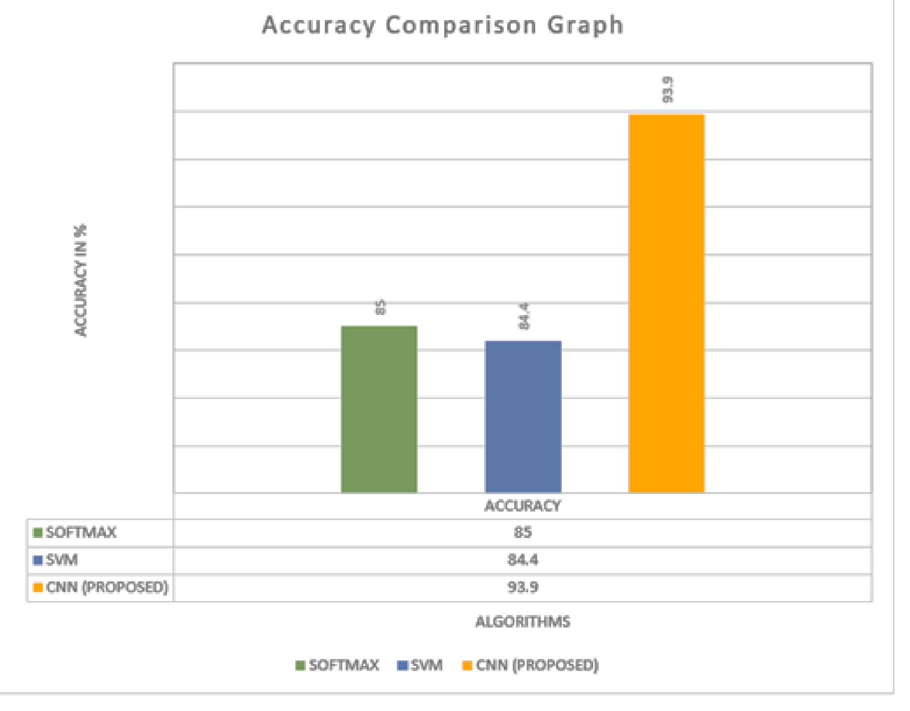

Figure 3: Accuracy Comparison Graph

Conclusion

Skin cancer is considered one of the most lethal cancers in the world. We use multiple deep neural network topologies, transfer learning techniques, and segmentation to create a computer-assisted skin lesion classification system. In addition, a CNN classifier was used to recognize melanoma and non-melanoma photos, as well as other pre-processing and augmentation procedures to reduce the influence of the ISIC's class imbalance.

References

[1] Shalu, Aman Kamboj, “A Color-Based Approach for Melanoma Skin Cancer Detection", International Conference on Secure Cyber Computing and Communication(ICSCCC),2018.

[2] Andre Esteva, Brett Kuprel, Roberto A. Novoa, Justin Ko, Susan M. Swetter, Helen M. Blau, and Sebastian Thrun,"Dermatologist-level classification of skin cancer with deep neural networks", Vol 542, p-115-127, Springer Nature Feb-2017.

[3] Aya Abu Ali, Hasan Al-Marzouqi ,“Melanoma Detection Using Regular Convolutional Neural Networks”,IEEE Conference on ECTA 2017

[4] Lequan Yu, Hao Chen, Qi Dou, Jing Qin, Pheng-Ann Heng, "Automated Melanoma Recognition in Dermoscopy Images via Very Deep Residual Networks", IEEE Transactions on Medical Imaging, Volume: 36, Issue: 4, April 2017.

[5] E. Nasr-Esfahani, S.Samavi, N. Karimi, S.M.R. Soroushmehr, M.H. Jafari, K.Ward, K. Najarian, "Melanoma Detection by Analysis of Clinical Images Using Convolutional Neural Network", IEEE 2017.

[6] Yu-An Chung, Wei-Hung Weng, "Learning Deep Representations of Medical Images using Siamese CNNs with Application to Content-Based Image Retrieval",31st Conference on Neural Information Processing Systems (NIPS 2017).

[7] Hiam Alquran, Isam Abu Qasmieh, Ali Mohammad Alqudah, Sajidah Alhammouri, Esraa Alawneh,Ammar Abughazaleh , Firas Hasayen, “The Melanoma Skin Cancer Detection and Classification using Support Vector Machine”, IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies.

[8] Haofu Liao,"A Deep Learning Approach to Universal Skin Disease Classification", Graduate Problem Seminar - Project Report, University of Rochester, 2015.

[9] Sait Suer, Sinan Kockara1, Mutlu Mete,"An improved border detection in dermoscopy images for density based clustering", BMC Bioinformatics 2011.

[10] Qaisar Abbas, Irene Fondo´n Garcia, M. Emre Celebi, Waqar Ahmad,"A Feature-Preserving Hair Removal Algorithm for Dermoscopy Images", Skin Research and Technology 2011.

[11] Alzubaidi, L., Zhang, J., Humaidi, A.J. et al. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. J Big Data 8, 53 (2021). https://doi.org/10.1186/s40537-021-00444-8

[12] Alzubaidi, L., Al-Amidie, M., Al-Asadi, A., Humaidi, A.J.; Al-Shamma, O., Fadhel, M.A.; Zhang, J., Santamaría, J., Duan, Y. Novel, Transfer Learning Approach for Medical Imaging with Limited Labeled Data. Cancers 2021, 13, 1590. https://doi.org/10.3390/cancers13071590

[13] https://www.isicarchive.com/#!/topWithHeader/onlyHeaderTop/gallery