A STUDY ON ENHANCED RF-DCNN CLASSIFIER BASED ON CNN

Dr. R. Jayaprakash, Assistant Professor, Department of Computer Technology, NGM College, Pollachi. *jayaprakash@ngmc.org*

Abstract
Ensemble classifications are the most significant techniques in the field of ML and choosing the suitable classifier is a challenging task. As a result, the primary goal of this thesis is to identify the appropriate classifier for competitive among various classifiers. This proposed algorithm helps to identify suitable combinations of methods which provide better classification output in terms of accuracy for different classifications of skincare problems using Dermatology publicly accessible dataset ISIC2019 images. Moreover, exploring different combination methods provides better classification out comes on Dermatology datasets and is able to categorize skin disorders.

Keywords:
Skincare, Dermatology, Scaling, Lesion, Detection

Introduction
To solve the skin lesion localization problem by approximating the correct boundaries of it. The approaches used for boundary detection are divided into three parts as manual, semi-automated, and automated approaches to image segmentation. The adopted approach in this thesis is based on self-learning model of neural network. This approach mimics the behavior of visual cortex to perceive the localized object as a lesion. The procedure is based on automated delineation of lesion boundaries for all class skin lesions. The proposed approach performs the activity in minimum computation time with improved accuracy.
To extract features from the skin lesion images which give the clinically relevant information. This information is useful for the understanding of dermatologists and helps to interpret the abnormalities in skin lesion images. However, the clinical inspired system is difficult to implement, but the approaches discussed in the thesis characterize the clinical attributes of the lesion in image space.
To perform the categorical recognition of all ten classes of skin lesions, more vibrant feature descriptors are required. Therefore, there is a need for a competent approach that can empirically improve the feature extraction process for skin lesion recognition in a diagnostic system of dermatology. The robust approach is able to recognize multi-class samples of skin lesions.

Lesion Segmentation
This is a difficult task that literature has carefully researched. The wide range of lesion shapes, sizes, and colors, as well as various skin types and textures, make it crucial to develop an accurate segmentation technique. A lesion border approximation is a needed step for proper extraction of features and consequently describe the characterization of lesions. To avoid miss-classification, there is a demand of appropriate segmentation approach, due to this, these approaches is categorized into manual, semi-automated and automated for a variety of CAD systems used for skin lesion diagnosis.
The change in human opinion about lesions boundary tracing equally promotes for automated approach related to lesion border approximation using segmentation. Dermatologists have reportedly used higher-level knowledge to estimate the lesion border, resulting in the average reproducibility of segmentation results. However, Silletti et al. argued that the state-of-the-art related to automatic segmentation approaches, with the exception of the Fuzzy C-Means (FCM), were poorly performed in comparison with expert dermatologists. The complexity can be attributed, along with many other things such as low contrasts, damage to surrounding skin, blurred boundaries, the existence of artifacts and irregular structures characterizing the skin lesion images. Readers can refer to the pre-processing techniques in the above section for enhancing image contrast and removing other artefacts typically found in global and local image processing in microscopic images. A part of literature, suggested that the tumor areas manually extracted by dermatologists have been found to be inconsistency, sometimes

by their characterize attributes, to validate the same by automated segmentation approach; can also help to assist in the reproducibility of results. Recently, to achieve better support for computer diagnosis, the current literature has seen a significant improvement in lesion segmentation from the surrounding healthy skin parts. However, Chang et al. argued that the fully automatic segmentation of all types of skin lesion images is not practical due to image modality and that is why the acquisition of skin lesion images becomes even more complex and certainly important. In some of the findings for better segmentation results, Karkunen Loeve Transform (KLT), commonly known as the Principal Component Analysis (PCA), has been used to improve the edges of the lesion image. The literature specify that, in order to achieve comparable lesion segmentation using ensemble methods, bottom-hat and top-hat transformations have been used to enhance the contrast of lesion images. The literature has charted the numerous border detection approaches that can helps to segment the pigmented skin lesion from the neighboring region in an automated mode .

Lesion Boundary Detection

Lesion boundary detection is the initial phase in the automated investigation of skin lesion images. It is essential for the following reasons: (a) The boundary structure gives significant information for accurate detection of lesions, (b) Extraction of other essential features such as various structures and colors of lesions. Lesion structure and variegated color is completely depends upon the accuracy of lesion outskirt detection. So, it is quite a challenging task due to the following reasons:

- Low contrast between the region of interest and normal skin region.
- Irregular and fuzzy boundary description
- Various types of artifacts blood vessels, skin hair , air bubbles and coetaneous regions
- Presence of multiple lesion
- Color variegation

This can be performed in the two phases as 1) pre-processing and 2) segmentation.

Pre-processing speed up the boundary detection mechanism by transformation of color space, image contrast enhancement, region of interest approximation, and removal of artifacts. In continuation of this, segmentation performed by the division of an image into disjoint areas that are homogeneous in relation to a selected attributes i.e. luminance, texture and color.

Pre-processing

Image preprocessing is an essential step, while dealing with the entitled images that do not have the excellent quality to be analyzed. These images have the lack of attributes due to the artefacts' such as color transformation, illumination effects, skin hairs, nails and blood dots, etc. Following methods have been discussed in the literature to overcome the image shortcoming and handle them nicely for better diagnosis of skin lesion and develop adaptable diagnostic system. For the color transform, there are different color spaces which can be frequently used for image processing include Red-Green-Blue (RGB), (CIE-L*a*b, CIE-X*Y*Z, CIE-L*u*v), Luma with chrominance (Y'PbPr, YIQ, Y'CbCr, Y'UV), Hue-Saturation-Intensity, Hue-Saturation Value and Hue-Saturation-Luminance (HSI, HSV, HSL). Usually, digitized lesion images are generated as RGB. These images are primarily transform into grayscale for performing scalar image processing to represent the image intensity instead of being accepted as a standard color image. An effort has been made by of Dobrescu et al.Tto ease down the classification accuracy of a skin lesion by converting the individual image sample used in their study to an image of 256 gray levels. But, over year the requirement of multichannel (vector) processing increases and to take benefit of the lesion's original color information, pigmented skin lesion images are traced directly. Afterwards, the main challenge becomes a computational speed in which there is a requirement of vector images.

The primary phase of homogeneous framework is preprocessing. It usually performs the color transformation, image restoration and removing artefacts. Conditions in which lesion images are acquired, directly influence the key features which can be used to discriminate the samples of lesion. Rahman contentedthat recovery and lesion classification tasks that might be challenging when capturing images are from separate datasets or devices under different conditions (such as lightening).

Due to this, a non-uniform pattern of illumination is generated that can disturb complete diagnostic procedures of skin lesion. One of the approaches proposed in the literature is to tackle such challenges by commencing the color calibration of the image acquisition device. Abbas et al., proposed an approach which can enhance the contrast of skin lesion image by making adjustments and map the pixel intensity values of lesions within the stated range of CIE L*a*b color space. One major flaw in contrast enhancement in the region with a relatively small intensity range is over noise amplification. These limitations may be addressed by using Contrast Limited Adaptive Histogram Equalization (CLAHE). The presence of artifacts, usually called noise, is a major obstacle for successful diagnosis of skin lesion by medical imaging. Furthermore, intact skin artifacts are hair shaft, dermoscopic gels, thin blood vessels, shadows, ruler markings, specular reflections, vignetting and air bubbles that can misrepresent diagnosis and hinder the accuracy of automated diagnosis system.
Reported literature clearly specified that the most frequent artefacts are hair shafts and marks of ruler. It has been observed that a great deal of effort has been made to remove hair shaft and markings by a ruler from skin lesion images. An awesome review by Abbas et al regarding comparative study of state-of-the-art algorithms. After the removal of intact artefacts, there is a challenge of lesion approximation and localization. Although images of skin lesions may be quite large, lesion frequently occurs in a relatively small space. For different reasons, an accurate bounding box (the smallest rectangular axis box that comprises the lesion) may be useful. It gives an estimate of the size of the lesion and its related techniques is discussed in the next section.

Image datasets for the melanoma diagnosis

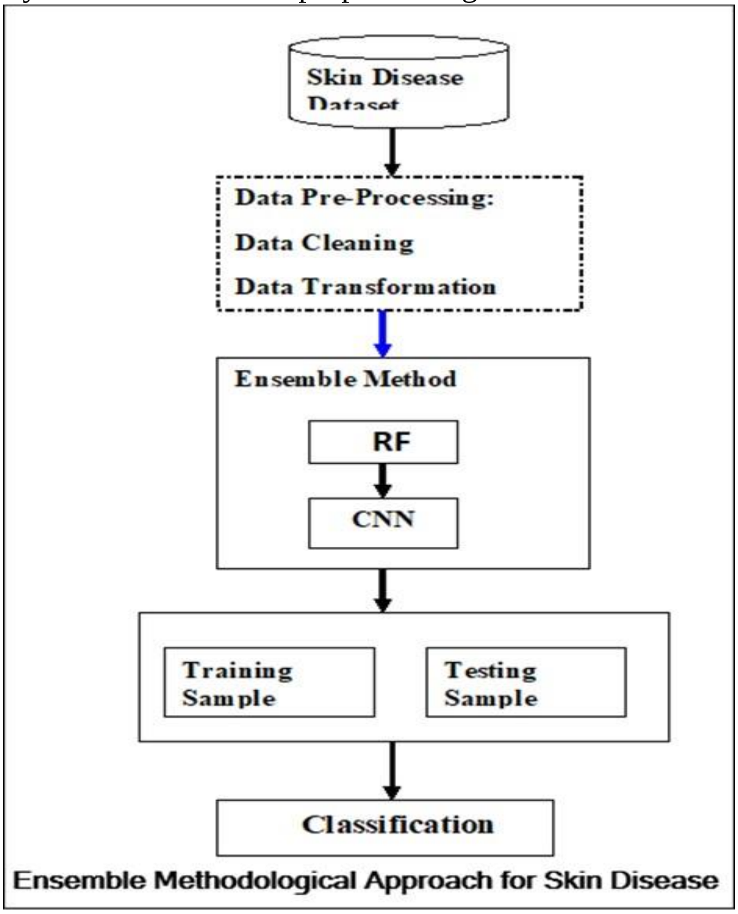
Due to the incidence of melanoma is continuing to increase worldwide, in the last few years, several private and public datasets have been published, thus allowing a better study of this illness and, therefore, the design of better approaches for its automatic diagnosis. The most popular private data collections of dermoscopic images are the Interactive Atlas of Dermoscopy, Dermofit Image Library, and the dataset presented, which conducted a comparison with 21 dermatologists using 129,450 clinical images. Regarding public datasets for studying melanoma, the most extensive collection of datasets can be found in the ISIC repository, which comprises images labeled by expert dermatologists. The 10015 dermatoscopic images of the HAM10000 training set were 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. The Australian site stored images and meta-data in PowerPoint files and Excel databases. The Austrian site started to collect images before the era of digital cameras and stored images and metadata in different formats during different time periods.
Dermatoscopic is a widely used diagnostic technique that improves the diagnosis of benign and malignant pigmented skin lesions in comparison to examination with the unaided eye. Dermatoscopic images are also a suitable source to train artificial neural networks to diagnose pigmented skin lesions automatically. Recent advances in graphics card capabilities and machine learning techniques set new benchmarks with regard to the complexity of neural networks and raised expectations that automated diagnostic systems will soon be available that diagnose all kinds of pigmented skin lesions without the need of human expertise.
The dataset is essential for training the neural networks we propose for automated diagnosis. The dataset entitled HAM10000 is the skin disease dataset that has been taken out from Kaggle, which has functioned as a baseline database retrieved from the source ISIC archive webpages. Furthermore, the vast majority of research do not use a traditional exploratory technique and only use a small number of datasets. The dataset includes age, gender, and cell type in metadata format, such as a comma-separated values file (.CSV). More than 10,000 dermatoscopic data were gathered from people all around the world for this collection. The dataset also includes extra recommendations and methods for dealing with the issues like over-fitting and insufficient data, which will aid in improving the model's accuracy and performance. Melanocytic Nevi (NV), Benign Keratosis-like Lesions (BKL), Dermato-fibroma (DF), Vascular Lesions (VASC), Actinic Keratoses (AKIEC), Basal-Cell Carcinoma (BCC), and Melanoma are the seven categories of skin issues in this dataset. The number of skin samples in

each type of lesion contained in the sample is imbalanced. To eliminate this variance, we used data augmentation techniques to bring all classes of lesions into the same digital image range. To improve model generalization, the dataset is separated into three parts: 85 percent training data, 5 percent validation data, and 10 percent testing data.

Techniques to improve CNN for melanoma diagnosis

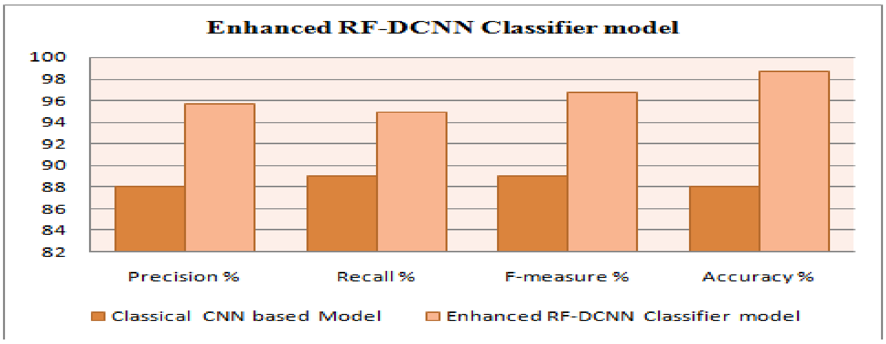
In the last decade, it has been an increasing tendency to develop and use modern CNN architectures to solve complex real-world problems. Researchers also try to apply advanced techniques to train these models better. Here we have summed up four main approaches, namely, transfer learning techniques, data augmentation methods, the development of ensembles of CNNs, and multitask models.

In this section, the Predictive Ensemble Deep Convolution Neural Networks Classifier (RF-DCNN) that is Random Forest Deep Convolution Neural Networks model is briefly described. The purpose of the present study is to develop an effective performance strategy for categorizing ceroscopy images data into seven groups. This section explains how the given approach for identifying, accessing, and analyzing data of skin problems works. The method will be incredibly effective in diagnosing seven different types of skin issues. Preprocessing, feature extraction, and classification are all elements of the architecture that may be subdivided. The proposed diagram's construction is shown in Figure.



Architecture of Enhanced RF-DCNN Classifier based on CNN

Performance for various metrics	Classical CNN based Model	Enhanced RF-DCNN Classifier model
Precision %	88	95.68
Recall %	89	94.93
F-measure %	89	96.78
Accuracy %	88	98.71



Conclusion

The results acquired in this study were compared to other data found in the literature to demonstrate the effectiveness of our technique. We used a large number of technical studies using the same information but various classification techniques to compare the efficiency of the proposed treatment of skin classification, and then developed a multi-model ensemble method to compare the efficiency of the proposed treatment of skin classification. After that, we use a multi-model ensemble approach to combine these two data mining techniques to get the greatest accuracy of 98.71 percent. On the skin illness dataset, we achieve the greatest accuracy in the literature. By employing the first-stage prediction as a feature rather than a separate training, the machine learning-based multi-model collection technique decreases generation mistakes and acquires more information. Furthermore, the complicated interactions between classifiers are automatically learnt using machine learning, allowing the collecting approach to make improved predictions. Furthermore, Enhanced Random Forest Deep Convolution Neural Networks (ERF-DCNN) Classifier has the highest accuracy of any of these approaches, at 98.71 percent. There will be many improvements and extensions in the future. First, the process of getting a skin disease should be in a created smart phone system, then a skin lesion will be found on the skin layer, and finally all the skindiseases in the whole area and the severity of the disease will be detected.

Acknowledgement

I place my special thanks to the management of NGM College, Pollachi, to carry out my research work with support of SEED Money and necessary facilities.

References

1. Chang W.-Y. , Huang A. , Chen Y.-C. , Lin C.-W. , Tsai J. , Yang C.-K. , Huang Y.-T. , Wu Y.-F. , and Chen G.-S. , “The feasibility of using manual segmentation in a multifeature computer-aided diagnosis system for classification of skin lesions: a retrospective comparative study,” BMJ open, vol. 5, p. e007823, 05 2015.
2. Fischer S. , Schmid P. , and Guillod J. , “Analysis of skin lesions with pigmented networks,” in Proceedings of 3rd IEEE International Conference on Image Processing, vol. 1, Sep. 1996, pp. 323–326 vol.1.
3. Mahammed M. , Bessaid A. , and ahmed A. taleb , “Extraction of specific parameters for skin tumour classification,” Journal of medical engineering & technology, vol. 33, pp. 288–95, 02 2009.
4. Zagrouba E. and Barhoumi W. , “An accelerated system for melanoma diagnosis based on subset feature selection,” Journal of Computing and Information Technology, vol. 13, pp. 69–82, 03 2005.
5. Taouil K. and Romdhane N. B. , “Automatic segmentation and classification of skin lesion images,” in The 2nd International Conference on Distributed Frameworks for Multimedia Applications, May 2006, pp. 1–12.
6. Celebi M. Wen Q. , HwangSae , Iyatomi H. , and Schaefer G. , “Lesion border detection in dermoscopy images using ensembles of thresholding methods,” Skin research and technology : official journal of International Society for Bioengineering and the Skin (ISBS) [and] International Society for Digital Imaging of Skin (ISDIS) [and] International Society for Skin Imaging (ISSI), vol. 19, 06 2012.