PERFORMANCE ANALYSIS OF TRADITIONAL DEEP LEARNING-BASED MODEL: RF-DCNN CLASSIFIER WITH CCNN

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

Abstract
Data mining is essential in the healthcare industry. In this study, we created predictive ensemble models that are well-suited for multiclass dermatitis by combining several basic and weighted rules. The results obtained in this research were contrasted with other information from previous studies to showcase the efficacy of our method. We conducted numerous technical studies utilizing the identical data but different classification methods to evaluate the effectiveness of the suggested skin classification treatment, and subsequently established a multi-model ensemble approach to assess its efficiency. Following this, we employ a multi-model ensemble method to merge the two data mining techniques in order to achieve a maximum accuracy.
Keywords: Accuracy, Dermoscopy, Texture, Classifier, Conventional.

Introduction

The skin is the primary component of the human body. The skin protects the body from UV rays, diseases, injuries, temperature, and harmful radiation, while also assisting in the production of vitamin D3. Since the skin plays a crucial role in regulating body temperature and defending against skin issues, it's essential to maintain its health. Skin conditions may seem benign, but they can pose a threat if not managed correctly. Numerous illnesses exhibit initial symptoms that are often similar, posing a challenge in early diagnosis. Skin disorders can lead to psychological issues as well as physical problems, especially for individuals with facial scarring or deformities. Skin can be affected by a variety of external and internal factors. Some of the factors that impact skin disorders include artificial skin damage, harsh chemicals, challenging illnesses, an individual's immune system, and genetic abnormalities. Skin conditions greatly impact individuals' overall well-being. Treating skin conditions is particularly difficult in the field of science due to the complexities involved in managing symptoms and their variability in different circumstances. Skin conditions are common among various ailments, and using inappropriate treatments for those conditions can lead to adverse consequences. People often contract skin diseases that require prompt treatment.

Literature Review

Imaging Techniques for Skin Lesion

Multiple imaging methods have been identified for examining skin lesions. Photography is the simplest perception method, providing an image of the skin's outer layer. The next phase involves using a polarized light camera in dermoscopy techniques, which reduces surface reflections and highlights the image of the epidermis, the second skin layer. Because of this, it is simpler to predict the formations of lesions like dots, globules, and networks, which are the primary signs of diagnosing skin lesions. Dermoscopy is a technique that enhances the link between medical science and specific visual traits, creating a new form of imaging called epiluminescence microscopy (ELM). This approach has provided a new way to explore pigmented skin lesions through clinical examination of morphological markers. Additionally, ultrasound can visually perceive the lesion, providing another method of diagnosis. Ultrasound imaging is commonly employed to measure the depth of skin lesions, and if there are no differences from healthy skin, then both regions will appear identical. At this stage, the expert examines lesions and employs high-frequency ultrasound (over 30 MHz) to assess how deeply they have infiltrated in order to ensure precise incision and proper surgical approach. On the other hand, optical coherence tomography (OCT) and confocal laser scanning microscopy (CLSM) are also available modalities. Examining with the naked eye has limited sensitivity in detecting early abnormalities such as a lesion. In this specific situation, several non-invasive techniques for examining skin lesions like dermoscopy are utilized. The focus is on body photography while the other is the confocal technique.

Medical Diagnostic Criteria in Dermatology

Dermatologists use certain diagnostic criteria to distinguish between various types of lesions. Some of these criteria are related to the set of patterns and colors that are observed in the skin lesion images. The first method was proposed in 1985 by Friedman et al. [1] to diagnose skin lesions is called pattern analysis. This method considers a set of patterns, also called global features, that can be found in each type of skin lesions. A specific pattern is characterized by one or more fundamental skin structures by local features(image descriptors) that can cover parts or the entire lesion. In addition, there are a variety of diagnostic approaches such as ABCD rule, Menzies technique, 3-point and 7-point checklist to evaluate lesion structures and patterns, using non-invasive imaging system[2,3]. The standard diagnostic rule of skin lesion is ABCD in the domain of dermatological was suggested in 1994 by Stolz et al [4]. Precisely, the following rule used to evaluate the properties of lesion as asymmetrical, border, color and,diameter [4,5].

Computer Aided Diagnostic System

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 [13] 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 [14]. Abbas et al. [15], 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) [16]. 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 [17,18].

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 specifies 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.

Proposed Methodology

This chapter provides an overview of ensemble learning for classification tasks. Specifically, it focuses on structured skin disease analysis for disease classification. In this chapter, two phases are proposed: Phase one, identifies the most appropriate datasets. Phase two proposed competitive ensemble classification model. It consists of multiple classification models which compete with one another to choose the most appropriate classifier for dataset. These two phases help to improve the classification performance of structured skincare data.

Classification is one of the most important decision-making techniques for selecting data. In this method, the main aim of research is to build Intelligent Skin Disease Prediction System to predict the data as presence of skin disease for improving the classification accuracy. This section comprises of proposed ensemble learning method for classification tasks to prediction of skin disease by classifier are briefly explained.

Convolution models in melanoma diagnosis

In simple words, CNNs are a type of feed-forward neural network that uses a mathematical convolution operation to extract features from images automatically. These features are passed to successive layers that learn more abstract features from previous ones until a final output is yielded, simulating some of the human visual cortex. Consequently, CNNs can build more complex concepts from simpler ones, e.g., they can learn to detect a human face based on more straightforward ideas as a nose and mouth, which are known from much simpler ones such as corners and contours. As a matter of example, Figure illustrates a simple CNN model that learns to predict whether or not a patient has melanoma utilizing mapping the extracted features into more abstract feature spaces. The general layout of CNN architecture is significant, including layers, activation function, and hyper-parameters. Layers in CNN are mainly categorized into convolutional, pooling, and fully connected layers. An activation function is a transformation function that maps the input signals into output signals required for the neural network to function. Popular activation functions include linear activation, Sigmoid functions (logistic and hyperbolic tangent functions), Rectified linear units (ReLU), also known as piecewise linear functions, Exponential Linear Unit, and Softmax. Hyper-parameters include filter kernel, batch size, padding, learning rate, and optimizers, etc. Optimizers are used to produce maximum performance from a network model. Examples include Adam, Rmsprop, Nesterov, and Sobolev gradient based optimizer.

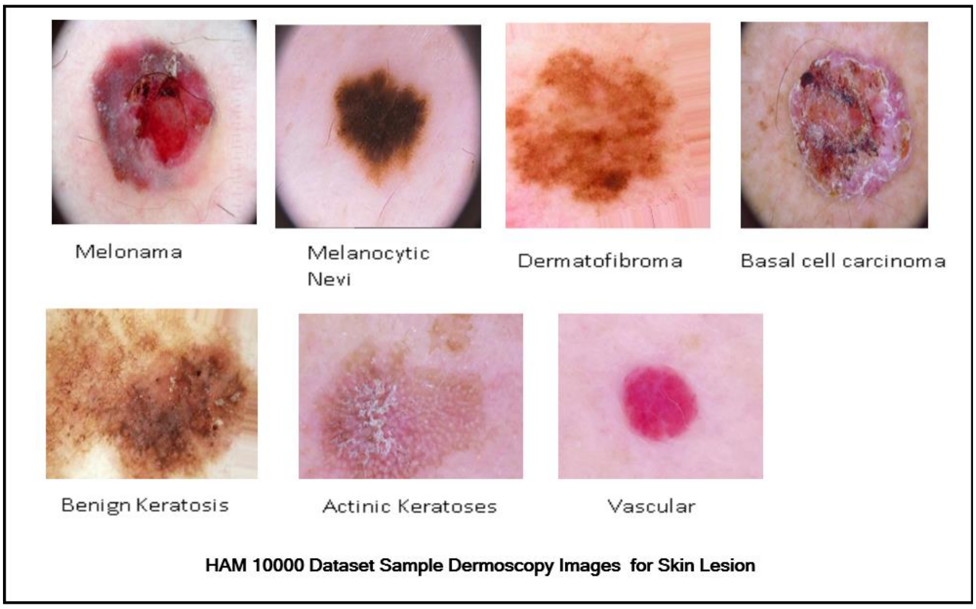
Image datasets for the melanoma diagnosis

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.

Dermatoscopy 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. The actual facts connected with the training dataset are used to evaluate the model. The image in our suggested model should be 224 224 pixels wide. The focus of this research is to see how accurate our suggested technique is in diagnosing skin cancer on dermatoscopic data. Figure 2 shows the HAM 10000 Dataset Sample Images.



HAM 10000 Dataset Sample Dermoscopy Images for Skin Lesion

HAM 10000 Dataset Sample Images

Pre-processing Skin Lesion Images

By boosting image quality, the model's ability to generalize can be increased. Preprocessing can reduce the quantity of redundant data in a visual while increasing the intensity of the vital data, simplifying the data, and improving reliability. The word "image resizing" refers to the process of resizing an image. An image size is either boosted or decreased in magnitude to solve the issue of varied pattern sizes in the database. All image will have the same amount of characteristics if the image size is reduced. Additionally, reducing the visual reduces processing time, improving system speed.

Data preparation is the first step in the approach provided in this research article. The preparation of data step includes, the information gathered from records is not always accurate and may contain noise, erroneous or missing numbers, or data that is inconsistent. For our studies, each input sample image is required to obtain features such as colour, texture, and form. As a result, we must use a variety of data cleaning techniques to eliminate such abnormalities. Even after cleaning, the data are not ready for mining since they are in multiple formats that cannot be utilized directly, thus the data must be transformed into mining-friendly formats. Normalization is the transformation used to achieve this; smoothing, aggregation, and other techniques are also utilized.

Ensembles Method for skin images for a fine-tuned models

The ensemble approach is employed in this research work to increase the performance of algorithms by determining the correctness of the skin disease dataset. Using the Random Forest technique and Convolutional Neural Networks (CNNs), framework analyze two alternative ensemble machine learning techniques.

Random Forest Technique

The Random Forest technique is suitable for the construction and analysis of very large datasets. To improve accuracy and performance, the Random Forest classifier was used. We suggest using random forest algorithms in this study to increase the reliability of skin image segmentation and classification, and comparing them to HAM 10000 data sets. The suggested approach may produce high-resolution feature maps that can aid in the preservation of picture spatial information. This model is both simple and smart, and it performed well in even the most difficult cases. Random forest aids in the construction of a number of decision trees and finally outputs the class to which the input variable belongs, according to Man et al. In our study, random forest helps to address the over-fitting problem that is common with decision tree training set data. To tree learners, they employ bootstrap aggregating approaches. There is a potential of forming a link between the kernel techniques and the random forests, which is commonly referred to as the kernel random forest, which is easier to grasp and interpret. This algorithm's predictive enactment can compete with the greatest supervised learning algorithms, and it provides a constant feature location estimate.

Random forests are a type of meta estimator that seeks to train a number of decision tree-based classifiers on distinct sub-samples of a given dataset while also using averaging principles to improve predicted accuracy and avoid over-fitting. The random forest algorithm entails a large number of different decision trees that work together as an ensemble, and provide a class prediction, with the most votes being the model prediction.

Any of the unique structural model will be outclassed by a large number of substantially uncorrelated trees acting as a committee. The poor correlation across different models is the key issue. The random forest algorithm's trees will protect each other from their discrete faults. Individual decision trees have been proven to be effective for skin lesion classification, however a collection of trees can progress in the right direction even if some of them are incorrect, resulting in total accuracy that is greater than individual decision trees. Because of the large number of decision trees that contribute to the process, the random forests method is regarded a very exact and robust model. Furthermore, it avoids the problem of overfitting since it takes the mean of all the predictions, which cancels the biases.

Algorithm for RF-DCNN Classifier:

Input: HAM 10,000 image dataset
Output: classified into different skin disorder classes
Begin
To extract multi-level features, map the features of skin visuals from the input domain (HAM10000) to the ensemble deep neural network (RF-DCNN);
Get the multi-level distinctive features that were extracted;
$\boldsymbol{for}(each\ input\ skin\ image)$
Train the RF-DCNN from start to the end;
Get the skin visuals with the two-step feature extracted;
Provide the RF-based DCNN classifier with the features extracted skin data;
Sort the various types of skin conditions into categories;
$\boldsymbol{endfor}$
End

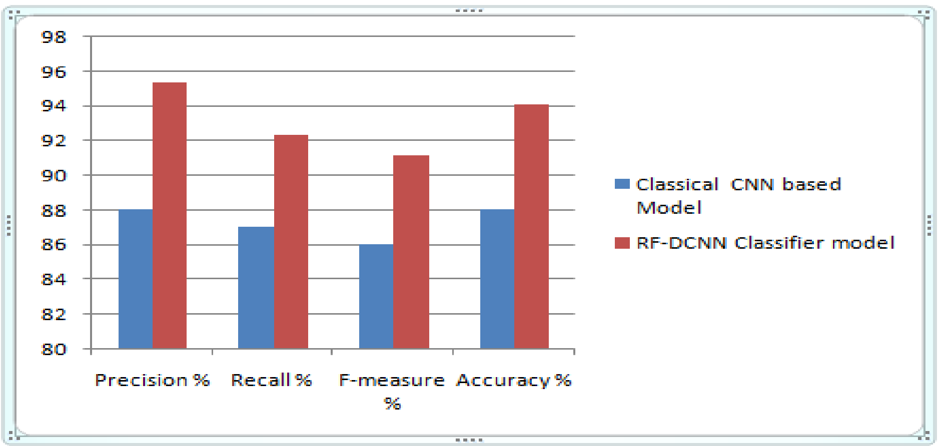
Findings of the Research

The performance measures of the suggested models, mean-accuracy, precision, re-call, and f1-score, were utilized to evaluate the performance of the RF-DCNN Classifier model. The performance of the RF-DCNN model is evaluated in this section using the HAM10,000 database using Python 3.6, with HAM visual images from each class being randomly picked for training. This study aided in the development of a system for forecasting skin disorders. Because regulators and medical institutions have never had a comprehensive plan for establishing information systems, this research is the most recent discovery. This might be due to a lack of human resources with formation technological competence and insufficient human resources for information systems. When applied to dermatological datasets, the ensemble technique and feature selection produce better results than individual classifier systems. The ensemble technique is more accurate and effective at predicting skin diseases. Table 4.1 shows how individual deep learners from classic CNN compare to the proposed ensemble classifier in terms of classification performance. Figure shows the performance of the Data mining is essential in the healthcare industry. In this study, we created predictive ensemble models that are well-suited for multiclass dermatitis by combining several basic and weighted rules. The results obtained in this research were contrasted with other information from previous studies to showcase the efficacy of our method. We conducted numerous technical studies utilizing the identical data but different classification methods to evaluate the effectiveness of the suggested skin classification treatment, and subsequently established a multi-model ensemble approach to assess its efficiency. Following this, we employ a multi-model ensemble method to merge the two data mining techniques in order to achieve a maximum accuracy of 96.1 percent. We reached the highest accuracy in the literature on the skin disease dataset.Learning model rather than the newly presented model.

Performance analysis with traditional deep learning-based models

Percentage of performance for various metrics	Classical CNN based Model	RF-DCNN Classifier model
Precision %	88	95.42
Recall %	87	92.33
F-measure %	86	91.17
Accuracy %	88	94.10

Effectiveness of Performance Metrics on RF-DCNN and CNN based Models



The integration model enhanced some of older CNN models with mean-accuracy, precision, f1-score and re-call, as shown in the table 1. Performance evaluation criteria such as accuracy, sensitivity, and specfcity may be used to compare skin lesion segmentation and classification performance. We identified various state-of-the-art methodologies for performance comparison in this research, and

the results are displayed in Table 2 below. In this comparison, words like true positive (TP), false positive (FP), true negative (TN), and false negative (FN) are commonly employed (FN). Since the layers are maintained and imported from a previously trained network. The study uses two separate classification methods: the random forest Classifier and the Deep Convolutional Neural Networks Classifier. We reached the greatest mean-accuracy of 96.1 percent after using these strategies.
model we showed in the study stated in Table employed a lot of the same data segmentation. Combined models have a more accuracy, indicating that the combination method identifies seven types of skin disease more effectively than the deep neural network alone. The proposed approach combines the possibilities of predicting different models and incorporates in-depth features from many well-trained and well-configured DCNNs at multiple output levels. This study discusses several data mining strategies for predicting skin diseases. To identify the prediction of skin illness, two machine learning approaches are used: random forest classifier and CNN.

Conclusion

This study aided in the development of a system for forecasting skin disorders. Finally, in the healthcare business, data mining is crucial. In this work, integrated Predictive heterogeneous ensemble models that work well for multiclass dermatitis are developed using a number of simple and weight-bearing rules. 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 percent.

Acknowledgement

I would like to express my sincere gratitude to NGM College Management for offering SEED Money support for this Project work. I also extend my gratitude to Principal, Director, Dean, Manager, and HoD for their continual support in accessing the college's resources.

References

1. Okuboyejo D. , O. Olugbara O. , and A. Odunaike S. , CLAHE Inspired Segmentation of Dermoscopic Images Using Mixture of Methods. Springer, Editors: H. K. Kim et al, 05 2014.
2. Jayaraman P. and Ravichandran K. S. , “Novel approaches for diagnosing melanoma skin lesions through supervised and deep learning algorithms,” Journal of Medical Systems, vol. 40, 04 2016.
3. Kiani K. and Sharafat A. R. , “E-shaver: An improved dullrazor (r) for digitally removing dark and light-colored hairs in dermoscopic images,” Computers in biology and medicine, vol. 41, pp. 139–45, 02 2011.
4. Giotis I. , Molders N. , Land S. , Biehl M. , Jonkman M. , and Petkov N. , “Med-node: A computer-assisted melanoma diagnosis system using non-dermoscopic images,” Expert Systems with Applications, vol. 42, 05 2015
5. Oliveira R. , Filho M. , Ma Z. , Papa J. , S Pereira A. , Manuel J. , and Tavares J. , “Computational methods for the image segmentation of pigmented skin lesions: A review,” Computer Methods and Programs in Biomedicine, vol. 131, pp. 127–141, 07 2016.
6. Abbas Q. , Celebi M. E. , and Fondon I. , ´ “Hair removal methods: A comparative study for dermoscopy images,” Biomedical Signal Processing and Control, vol. 6, pp. 395–404, 10 2011.
7. Suer S. , Kockara S. , and Mete M. , “An improved border detection in dermoscopy images for density based clustering,” BMC bioinformatics, vol. 12 Suppl 10, p. S12, 10 2011.
8. Mete M. , Kockara S. , and Aydin K. , “Fast density-based lesion detection in dermoscopy images,” Computerized medical imaging and graphics : the official journal of the Computerized Medical Imaging Society, vol. 35, pp. 128–36, 03 20

9. Silletti A. , Peserico E. , Mantovan A. , Zattra E. , Peserico A. , and Fortina A. B. , “Variability in human and automatic segmentation of melanocytic lesions,” in 2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Sep. 2009, pp. 5789–5792.
10. 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.
11. 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.
12. 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.
13. 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.
14. 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.
15. 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.
16. Kittler J. and Illingworth J. , “On threshold selection using clustering criteria,” IEEE Transactions on Systems, Man, and Cybernetics, vol. SMC-15, no. 5, pp. 652–655, 1985.
17. Sezgin M. and Sankur B. , “Survey over image thresholding techniques and quantitative performance evaluation,” Journal of Electronic Imaging, vol. 13, pp. 146–168, 01 2004.
18. Barata C. , Marques J. , and Rozeira J. , “Evaluation of color based keypoints and features for the classification of melanomas using the bag-of-features model,” in Advances in Visual Computing, vol. 8033, 07 2013, pp. 40–49.
19. Stanley R. , Stoecker W. , Moss R. , S Rabinovitz H. , B Cognetta A. , Argenziano G. , and Soyer P. , “A basis function feature-based approach for skin lesion discrimination in dermatology dermoscopy images,” 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. 14, pp. 425–35, 12 2008.