Diagnosis of Skin Cancer Using Machine Learning and Image Processing Techniques

Prashant Kaler
Student, Department of CSE (MCA), Visvesvaraya Technological University, Centre for PG Studies, Kalaburagi, India
E-mail:prashantkaler3137@gmail.com

Shilpa Kodli
Assistant Professor, Department of CSE (MCA), Visvesvaraya Technological University, Centre for PG Studies, Kalaburagi, India
E-mail:shilpakodli@gmail.com

Sudhir Anakal
Research Scholar, Department of CSE (MCA), Visvesvaraya Technological University, Centre for PG Studies, Kalaburagi, India
E-mail:sudhir.anakal@gmail.com

Abstract: Skin Lesion is a part of the skin that can be caused by abnormal growth in the epithelium layer on the skin. There are nine types of skin lesion like Actinic Keratoses (AK), Basal Cell Carcinoma (BCC), Dermatofibroma (DF), Melanoma (MEL), Melanocytic Nevi (MV), Benign Keratosis (BK), Vascular Lesions (VASC), Squamous Cell Carcinoma (SCC), and Pigmented Benign Keratosis (PBK). The aim of this study is to spotlight on the problem of skin lesion classification based on early detection of the disease using deep learning techniques. This approach is used to work out the problem of classifying a dermoscopic image. The dermoscopic is a digital device; in this case Smartphone is attached to a lens and collects the images through the device. The proposed spotlight is built in the region of using Convolutional neural network architecture and ResNet-50 module is used to predict Skin-Lesion classification. The dataset used in this research was taken from kaggle repository. The proposed work uses ResNet-50 CNN model which has yielded 93% of accuracy for detecting Skin Cancer, previous work was carried out using Visual Geometry Group model which yielded 73% accuracy. In the proposed work we have considered 25,000 images of skin lesion. Hence we are able to attain this accuracy with more reliable Machine Learning algorithms compared to the previous work.

Index Terms: Dermoscopic Images, Machine Learning, Convolutional Neural Networks, Skin Lesion, ResNet-50.

1. Introduction

Skin lesion arising from the skin can cause abnormal growth in the epithelium layer on the outermost layers of the skin, which is often diagnosed as a skin lesion disease. It triggers mutations and also spreads to whole parts of the body, especially often appearing on painless raised areas of the skin, such as smooth, pearly bumps due to sun-damage due to small blood vessels. In skin lesions, there are nine types of skin cancer namely AK, BCC, DF, MEL, MV, BK, VASC, SCC, and PBK. The BCC, SCC, and MEL are considered to be the most important types of skin cancer, and non-melanoma skin cancer is one of the most common skin cancer types.

This basically means we see reactions on the shoulders, neck, and head or sometimes on the small blood vessels. Here, the patient needs to visit the hospital and receive the appropriate treatment. BCC is one type of skin cancer that affects the skin and grows slowly, causing damage to the tissue on your body and causing death. If a patient cannot receive proper treatment at a specific time, the patient can suffer from skin cancer. The BCC symptoms can be seen on the neck, head, and shoulders.

This disease basically causes painless areas to look like smooth, pearly bumps due to sun-damage. SCC is the second most common skin cancer that grows large in the squamous cells in the middle and outer layers of the skin and can spread to other parts of the body, causing serious complications and injuries to the patient. He can suffer from skin cancer if he does not receive proper treatment. By avoiding ultraviolet light, we can reduce the risk of squamous cell carcinoma of the skin.

This is the most serious type of skin cancer which develops in the cells that create melanin (pigment), which gives color to the skin. MEL can be in your eyes and rarely into your body, such as your nose or throat. By avoiding ultraviolet light, we can reduce the risk of MEL in the skin. Here, we are classifying using diagnosed dermoscopic images. It is used to visualize the epidermis of the skin. The use of dermoscopic images is to get better diagnostic accuracy of the skin lesion. It is described above that intelligent medical images are based on skin lesion diagnosis systems that assist a physician in classifying the skin lesion. In this project, methods or models are used to classify skin lesion classification using deep learning.

Here we are utilizing the dermoscopic image dataset from kaggle to implement the model, which is a well-known platform for dataset collection. In this dataset we have over 25,000 images of various skin conditions. To classify these images, we have used a ResNet-50 model, which is used to train the test model to get better accuracy.

The highlights of the model are:

- A key objective of this research is to provide interactive support tool with straight forward and easy to use.
- To provide accurate screening of pigmented skin lesion to improve early detection.
- And also, we are focusing on reducing unnecessary referrals to hospitals and also reduce the number of effect of skin lesion.

The objectives for this research is to come up with a better deep learning model which can help us to get better accuracy for the diagnosis of Skin Cancer. The problem found in the previous work is the author has used a VGG deep learning model and achieved 73% of accuracy in predicting the disease and also has classified the skin cancer disease of only two types i.e. Melanoma and Benign respectively. So in our work we have used ResNet50 which is a deep learning CNN model which is used to solve the vanishing gradient problem which is better suited for classification problems. As compared to the previous work where the author has considered only two types of skin lesion in our work we are considering nine types of skin lesion namely Actinic Keratosis, Basal Cell Carcinoma, Dermatofibroma, Melanoma, Nevus, Pigmented Benign Keratosis, Seborrheic Keratosis, Squamous Cell Carcinoma, Vascular Lesion.

2. Literature Review

Adria Romero et.al [1] has used Convolutional Neural Network (Con.Net) architecture; it is used for large scale Visual Recognition. In this research, they have worked on Melanoma and Benign disease and, the sample images taken from the ISIC Archives dataset consist of dermoscopic images. In this study, there are used CNN model to analyze and taken skin lesion images of Melanoma and Benign and pre-trained using VGGNet on large dataset. The evaluation of the training dataset is 76.88% and the model evaluation of the test dataset is 81.33%.

Ranpreet kaur et.al [2] has done a comprehensive research solution as proposed using computer vision algorithm. The different algorithms that have been used include Decision Tree (DT), Support Vector Machine (SVM), and Artificial Neural Network (ANN). This machine learning algorithm is used to remove constraints in processing data and to remove such contrast, noise-free and cleaned images. In this paper, Deep Convolutional Neural Network (DCNN) algorithm is used for first breakthrough on skin cancer. It is used to pre-train the InceptionV3 Model on 129,450 clinical images that are used to perform classification on 2032 different diseases and they utilize 68 Convolutional layers.

In Manu Goyal et.al [3] has proposed a solution using deep learning methods. They have trained the dataset using Mask R-CNN and DeepLabV3+ methods. In this paper, the proposed results are based on a specific score of 97.98% based on clinically benign cases; 97.30% for melanoma cases; and 98.58% on seborrhoeic keratosis cases. This model over-performed other CNN models such as FrCN, FCNs, U-Net, and SegNet.

Julie Ann A. Salido et.al [4] has worked on skin lesions using dermoscopic images and they have preprocessed the images. And, then they classify the images using CNN and they test these classifiers using both preprocessed and unprocessed images from the PH2 dataset. Here the proposed result with 93% accurate and absence of melanoma at the range of 86% to 94%.

Haseebyounis et.al [5] has discussed classification of skin cancer dermoscopic images using transfer learning. The algorithm used to pre-trained, MobileNet Convolution neural network. The images trained more than 1000 skin lesion datasets. The weighted average accuracy of precision and recall is 0.90. This help for prognosis skin cancer at early phase.

The paper published on skin lesion by Dr. Ahlam Fadhil Mahmood et.al [6] explains that they have developed different pre and post approaches for diagnosis of melanoma using various dermoscopic images. Initially, they have alienated unwanted features from the images in preprocessing phase and adapted two different approaches for detection; first approach is to use a fuzzy based rule system and in the second testing approach is to use new modeled local colors with bag of features.

Deep Learning has been used in the detection of many diseases and even used for object detection. Neema M et.al [7] Ventured into study for classifying the skin lesion. In this paper they have adapted a deep learning approach which mainly focuses on skin lesion classification. This model helps in classifying between melanoma and benign with the accuracy of 70%.

The main objective of this paper is to classify the skin lesion images into three different classes, i.e. normal, abnormal and melanoma. This paper was published by the Murat Koklu et.al [8], in which they have taken 4 different machine learning algorithms, namely, SVM, KNN, decision tree, and ANN. In this model they have utilized PH2 dataset and have achieved the accuracy of 89%.

In this paper, the author R.D.Seeja et.al [9], came up with a proposed solution using deep learning methods. The objective of this paper is to classify the melanoma using dermoscopy images. Initially, they came up with a deep learning based U-Net algorithm and Convolutional Neural Network used to segment the melanoma lesions. The VGG16 algorithm is used to classify every lesion in a dermoscopic image as a melanoma. This model helps in classifying melanoma with the accuracy of 83.18%

Dermoscopy and the dermatologists are considered to be the renowned approaches for skin lesion classification. Nowadays, everyone who immerses themselves in the research of image classification or object detection is mostly using image processing and computer vision techniques for improving the accuracy of the model. Bhuvaneshwari Shetty et.al [10], has published a study in which they are analyzing the accuracy of skin lesion classification using both CNN and Machine learning algorithms, and they have found out that the CNN provides the better accuracy than other machine learning algorithms.

In the current study they haven't used machine learning algorithm for accurate detection of skin cancer and they have worked only on a few selected diseases like MEL and BK, and the patient's needs to follow radiologist's words for the diagnosis and detection of the diseases. Occasionally physicians can umpire it incorrectly and classify the disease wrongly where the patient has to undergo whole life. Therefore to overcome this rigid issue we are developing a machine learning model with 9 different types of skin cancer, which helps the radiologist's to detect the skin cancer at the earliest. In this study the model has attained the accuracy of 93% which is higher compared to current model.

3. Methodology Used

Determining the skin disease devoid of any sophisticated deep learning technique is delineated somewhere between 60 to 80%. However the use of skin lesion images taken from the dermoscopy may increase the detection of skin cancer but skin lesions are very subtle, that can certainly catapult any dermatological expert into difficulty in discriminating the disease. Deep learning has gained so much advancement in the medical field or other fields so that it has been successful in attracting many specialists from different fields. Swift enlargement of the deep learning into image classification and biomedical data processing has influenced many specialists to acquire this technique for more precision and accurate results. When there is a chance of human errors in determining the presence of the disease then comes the computer aided disease diagnosis, which could predict the results in little period of time with high accuracy. In this research, demography images are pre-trained with the ResNet-50 a CNN model which has given the accuracy rate of 93%. This type of techniques will give the doctors more precision in determining the diagnosis of the skin disease.
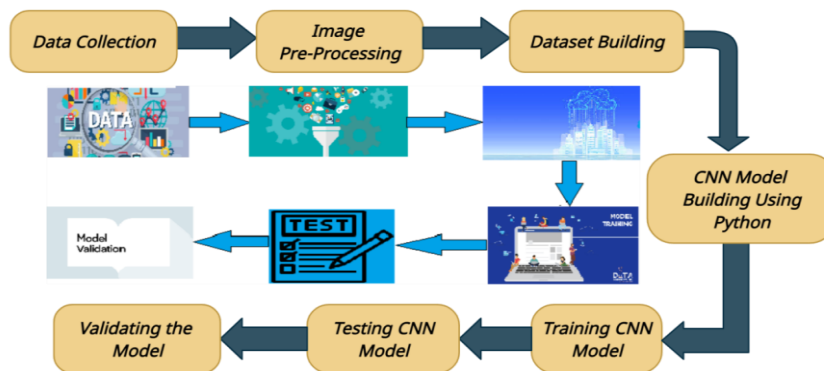

Fig. 1. Proposed Methodology

In the above the proposed methodologies flow is represented. Here, first the data is collected. The data is in the form of an image. We then apply image processing concepts to clean the images (like aspect ratio of the image, resolution of the images are done). Once the images are preprocessed we then build a data set of refined images. Later, the CNN model was built using Python as a coding language. The CNN model is then trained with the dataset. After training the model with the dataset, it is then tested. After testing the model, its validation is performed.

The Convolutional Neural Network (CNN) is commonly used to visualize images or objectives. CNN is a division of Artificial Neural Network (ANN). A basic CNN consist of three layers Convolutional Layer, Pooling Layer, And Fully Connected Layer. CNN is used in various areas like image classifying, computer vision, natural language processing, soon. CNN was developed in the early 1980s. We have also used Matplotlib library for displaying the

images and depicting plots for showing the validation and training losses. CNN is typically used for object detection, image classification, speech recognition, and for some other applications.

Residual Network (ResNet-50) is a type of neural network, which is used for building the model, training, and testing the ResNet-50 model. The ResNet-50 model makes the steer clear of degradation problems which are faced by other CNN models. With the advent of CNN, we are obtaining cutting edge results in the issues such as image classification, etc. but researchers were mulling to build some profound neural networks by adding some more layers to it, when developers started to add more layers in CNN model to get more accuracy and that certain point the accuracy stated to saturated. ResNet-50 is a pre-trained model of CNN, which comes with numerous varieties such as ResNet-18, ResNet-101, ResNet-152, etc. There are many architectures in ResNet-50 as stated before, but we have chosen ResNet-50, it has displayed much better accuracy and efficiency compared to other models. Below, architecture gives you a basic overview of how a ResNet-50 model works.
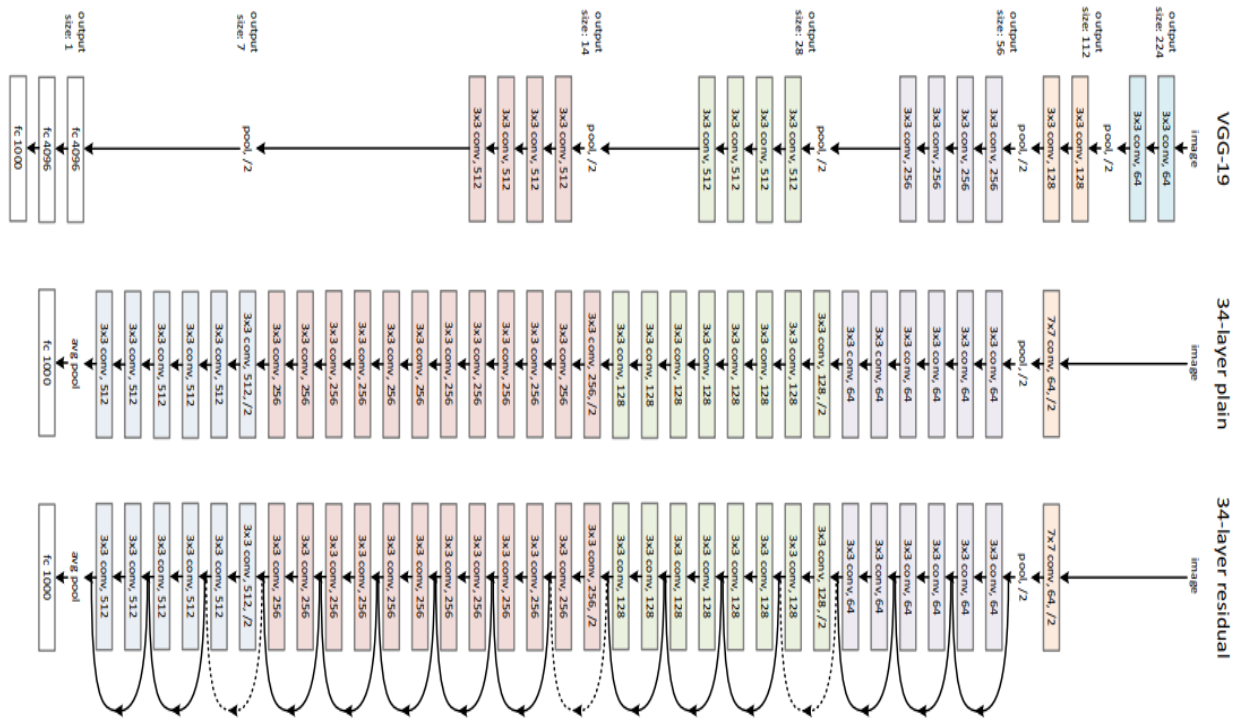

Fig. 2. ResNet-50 Architecture

Below table 1 shows the number of images used for both training and testing of the model.

Table 1. Images used for Training and Testing of the Model

Class	Number of Training Images	Number of Testing Images
Actinic keratoses	114	16
Basal cell carcinoma	376	16
Dermatofibroma	95	16
Melanoma	438	16
Melanocytic nevi	357	16
Benign keratosis	77	16
Vascular lesions	139	03
Squamous cell carcinoma1	81	03
Pigmented benign keratosis	462	16

4. Results and Discussions

This model is developed using python programming language and Jupyter notebook has been chosen as a tool for executing this project. Jupyter notebook facilitates us to write code in each individual slides, which helps us to find any snippet of the code easily. This notebook also provides ample space for depicting the graphs and results of the project in a beautiful and attractive manner. The below picture shows how the images are displayed in the Jupyter Notebook.
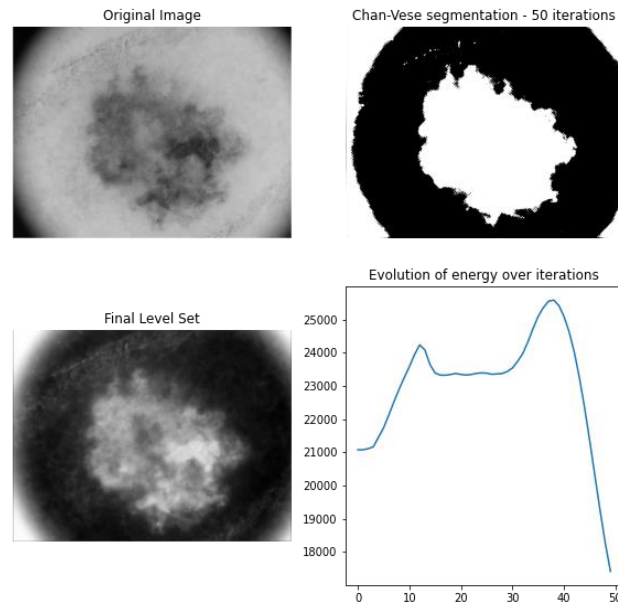
Fig. 3. Image Segmentation

In fig. 4, we are showing all nine types of skin lesion images namely AK, BCC, DF, MEL, MV, BK, VASC, SCC, and PBK.
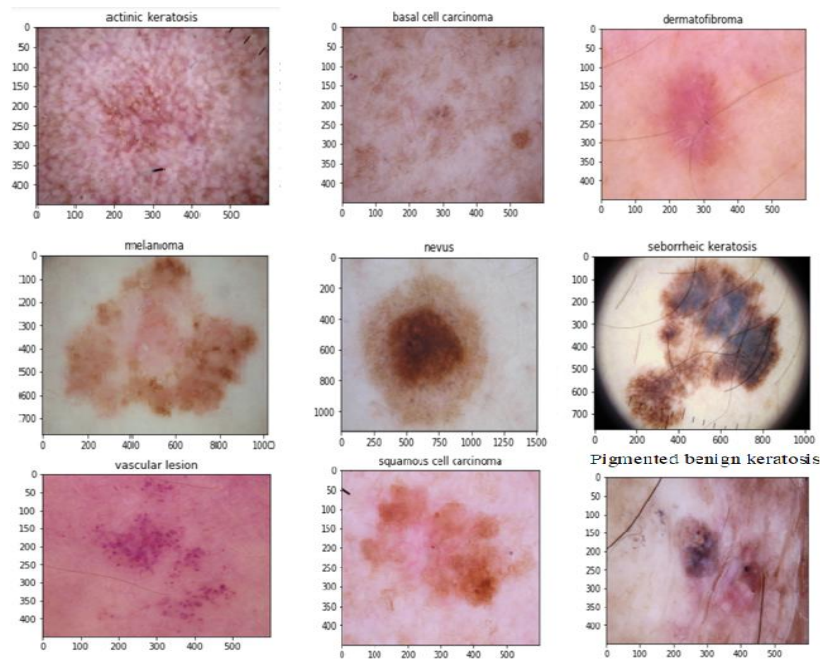

Fig. 4. Skin Lesion Types

Below points provides a brief explanation about the Fig. 4.

➢ Actinic Keratosis lesion image, you can see rough, scaly patch on the skin which comes through sun exposure, mostly you can find at face like lips, ears, forearms and back of the hands. This is also one of the skin lesion type used in this research.
➢ Basal Cell Carcinoma. It caused by exposing the skin to ultraviolet radiation through sunlight, mostly commonly you can see head, neck etc.
➢ Dermatofibroma skin, it commonly occurring cutamous entity. The Dermatofibroma referred to as benign fibrous histiocytomus of skin.
➢ Melanoma is effective and most dangers type of disease, it begins in cell Melanocytic and present melanin. Most commonly you can see nose or throat of the skin.
➢ Melanocytic Nevi, this is usually small in size and circular in shape. It cause through abnormal cells in skin pigmentation in some part of the body.

➢ Seborrheic Keratosis, this is usually in brown color and rounded like oval, most commonly spread on the face, chest and back.
➢ Vascular, it is usually in red in color and grown in faint areas and abnormalities of the skin. Most commonly known as birthmarks.
➢ Squamous cell carcinoma, the Squamous cell carcinoma grows in large and spread all over the body, it makes serious complications and it begins the squamous cells in middle and outer layer of the skin.

In fig. 3, it shows that the image is converted into the gray scale image to find the disease contour on the skin. By converting the image into the gray scale, it becomes very easy to find out the disease spread area with its intensity. After training the model, the accuracy gradually increased to 93% percentage. The predictions made by the model were accurate. In figure 4 you can view the confusion matrix of the model, where it has predicted all the skin diseases classes successfully.
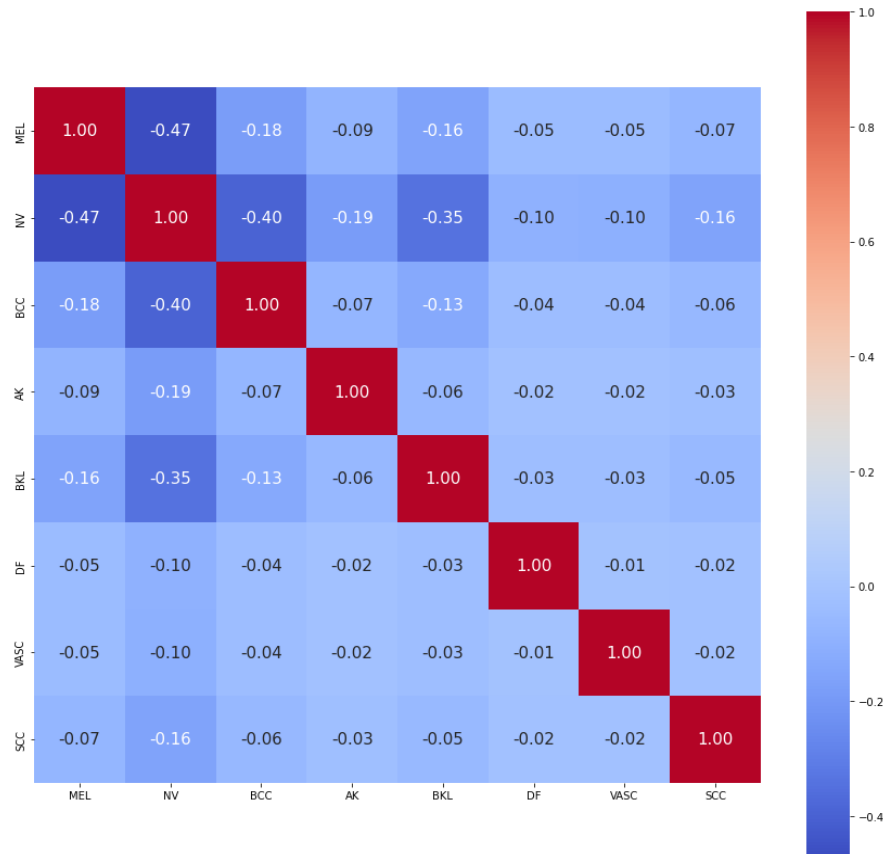

Fig. 5. Confusion Matrix

5. Conclusion

Here, we are proposing a solution for a major public health problem of skin lesion. We need to put effort to address skin lesion risk factor. We want to control skin lesion through giving information to people and raising awareness about the disease. The youth must be aware and protected from harms like Actinic Keratosis, Squamous Cell Carcinoma, Melanocytic Nevi, Pigmented Benign Keratosis and soon. This model will greatly aid doctors in the detection of skin disease. In this research we provide 9 classes of skin disease images as input to the model and using the ResNet-50 model we are training these images and testing for better accuracy. After training and testing the model we got 93% accuracy. For the detection of the skin lesion, a dermatologist needs maximum experience for correct diagnosis of the disease. Incorrectly judging the disease may alienate the right treatment from the patient. Therefore, these types of sophisticated models will help doctors correctly detect skin lesion.

As a future work, we are developing an application using this model, which will be made available for both the android and the ios users. This application would be useful to hospitals and some other medical facilities. This model should be further developed with large datasets with some other sophisticated algorithms to overcome even sensitive diseases which are very elusive to the current technologies.

References

[1] Adria Romero Lopez, Xavier Giro-i-Nieto Universitat Politecnica de Catalunya Barcelona, Catalunya, Spain, "Skin Lesion Classification from Dermoscopic Images Using Deep Learning Techniques".

[2] Ranpreet Kaur, Hamid Gholam Hosseini, Roopak Sinha and Maria Linden, "Melanoma Classification Using a Novel Deep Convolutional Neural Network with Dermoscopic Images".

[3] Manu Goyal, Amanda Oakley, Priyanka Bansal, Darren Dancey and Moi Hoon Yap, "Skin Lesion Segmentation in Dermoscopic Images With Ensemble Deep Learning Methods".

[4] Julie Ann A. Salido and Conrado Ruiz Jr, "Using Deep Learning to Detect Melanoma in Dermoscopy Images".

[5] Haseeb younis, Muhammad Hamza Bhatti, Muhammad Azeem, "Classification Of Skin Cancer Dermoscopic Images Using Transfer Learning".

[6] Dr. Ahlam Fadhil Mahmood, Hameed Abdulaziz Mahmood, "Appending global to local features for skin lesion classification on dermoscopic images".

[7] Neema M, Arya S Nair, Annette Joy, Amal Pradeep Menon, Asiya Haris, "Skin Cancer Detection using Deep Learning".

[8] Ilker Ali Ozkan, Murat Koklu, "Skin lesion classification using Machine Learning Algorithms".

[9] R.D.Seeja, A.Suresh "Melanoma Segmentation and Classification using Deep Learning".

[10] Bhuvaneshwari Shetty, Roshan Fernandes, Anisha P Rodrigues, and Kuruva Lakshmanna, "Skin Lesion Classification of Dermoscopic images using Machine Learning and Convolutional Neural Network".

[11] H. Kittler, H. Pehamberger, K. Wolff, and M. Binder, "Diagnostic accuracy of Dermoscopy".

[12] E.H.Page, "Description of skin lesion" https://google/m9ybFp.

[13] Skin Cancer Foundation. (June 2017). [Online]. Available: http://www.skincancer.org/skin-cancer-information/melanoma.

[14] N. K. Mishra and M. E. Celebi. ―An overview of melanoma detection in dermoscopy images using image processing and machine learning,‖ eprint arXiv: 1601.07843 2016.

[15] A. A. Abder-Rahman and T. M. Deserno, ―A systematic review of automated melanoma detection in dermatoscopic images and its ground truth data,‖ in Proc. SPIE, Medical Imaging 2012: Image Perception, Observer Performance, and Technology Assessment, 2012, vol. 8318.

[16] "Usage of objectives," https://keras.io/objectives/.

[17] Sudhir Anakal, P Sandhya, "Decision Support System for Drug-Drug Interaction Pertaining to COPD and its Comorbidities", International Journal of Education and Management Engineering (IJEME), Vol.12, No.2, pp. 1-6, 2022. DOI: 10.5815/ijeme.2022.02.01

Authors' Profiles

Prashant Kaler is a student studying Masters of Computer Applications at Visvesvaraya Technological University, Center for PG Studies, Kalaburagi, India. Research areas are Machine Learning, Data Mining, Artificial Intelligence, Deep Learning and Data Visualization.

Shilpa Kodli, (PhD), is working as Assistant Professor at Department of Computer Science and Engineering (MCA), Visvesvaraya Technological University, Center for PG Studies, Kalaburagi, India. Research areas are Cloud Computing, Networks, and Artificial Intelligence. She has published more than 19 peer reviewed research articles.

Sudhir Anakal, (PhD), is a Research Scholar at Department of Computer Science and Engineering (MCA), Visvesvaraya Technological University, Center for PG Studies, Kalaburagi, India. Research areas are Machine Learning, Data Mining, Artificial Intelligence, Data Analytics and Cyber Security. He has published more than 12 peer reviewed research articles.