IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Skin Cancer Detection using Machine Learning And Cloud

Miss. Kakade pratiksha Shanker Miss. Kalse Payal Ramesh

Miss. Sahane Gauri Sunil Miss. Mokal pradnya dnyandev

prof.M.R.Parkhe

B.E Computer Eigg. Dept, P.R.E.C. Loni, Ahmednagar, Maharashtra-413736., Savitribai Phule Pune University.

Abstract -Today as information is in abundance on all the digital mediums as it has become an integral part of our daily life. This information can be used to improve various daily needs of our life from security, banking and healthcare etc. Using internet as a medium we can decrease the load on the human being to handle the data manually.so ,we thought of exploiting this information to help medical field as a medical imaging data is available in abundance on the internet freely Thus, medical image analysis will be very helpful to a medical practitioner. It will help in decreasing the work load of a doctor and him diagnosis the disease more accurately and increase his productivity to a new level. So, we thought of applying our idea on cancer disease. Today cancer has become very common disease and taking more lives yearly. Thus, automaticmedicalimaginganalysiswillbeabooninearlycancerdetection. In cancer we are going to analyze skin cancer images which has become a major health problem. So, we thought of designing a computer-aided diagnosis (CAD) system which will help in skin cancer detection and handling various stages of cancers using skin lesion images as the backend of our application So, we propose a system we are going to download skin lesion database from ISIC. ISIC contains a large free to use various stages of skin lesions. Then the SIFT algorithm will be applied on the images to extract features. The extracted features will then be saved in local database for further use. Then we will create training dataset by studying features with two classes benign and malignant. Thus, finally we apply Naive Bayes for classification and get accurate classification of skin lesion image given for testing. Thus, decision making and productivity of a medical practitioner can be improved so that he can concentrate on treatment and not diagnosis.

Keywords – skin disease, ISIC, SIFT, NB, mobile computing, cloud computing

I. INTRODUCTION

Image Processing and Cloud Computing are hot topics. Today a lot of things can be achieved in medical field by using Image Processing techniques to identify and provide information various diseases that can affect a human being. But a smart system has to be developed which will identify the Skin disease and provide information to the user if he has any. To solve this problem, we thought of designing a Smart Disease Identification and Detection system as our project.

1. To support decsion making in disease diagnosis.
2. To help a medical practioner save time and life of a patient.
3. To make effective use of cloud technology so that a patient and doctor can access system from anywhere and anytime
4. To save a patients life by early diagnosis of a disease
5. To combine image processing, feature extraction, machine learning, desktop computing, mobile computing and cloud computing technogies together to achieve a skin cancer diagnosis system

II. LITERATURE SURVEY

This section describes the fundamentals of various image processing techniques that can be used in designing a new more reliable skin cancer diagnosis system to help medical practioner to diagnoise skin cancer easily. It helps in understanding various ideas put forward by various technical papers published by various authors and how they put forth a more accurate and concrete techniques. Some of the ideas with technique and drawbacks are mentioned below:

1. Paper: - Accurate segmentation and registration of skin lesion images to evaluate lesion change. Year: - 2019. Author: - Fulgencio Navarro, Marcos Escudero-Vinolo and Jes ˜ us Besc ´ os. ´ Technique: - SIFT algorithm to evaluate skin lesion images. Drawback: - This paper gives more dependance on feature extraction, Segmentation and not on prediction usingmachine learning which will increase the accuracy of diagnosis.

2. Paper: - Research on Image Detection and Matching Based on SIFT Features. Year: - 2018. Author: - Feng Guo, Jie Yang, Yilei Chen, Bao Yao. Technique: - Image detection and matching using SIFT. Drawback: - This paper gives more dependance on image detection and matching and not on prediction using machine learning which will increase the accuracy of diagnosis.

3. Paper: - Research on Vehicle Detection Technology Based on SIFT Feature. Year: - 2018. Author: - Sujata Rani and Parteek Kumar. Technique: - Vehicle detection using SIFT.

Drawback: - This paper gives more dependance on vehicle detection and matching and not on prediction using machine learning which will increase the accuracy of diagnosis.

Thus, this kind of attack as shown in Fig.1 can be avoided by designing a mechanism by using machine learning and artificial intelligence techniques together. The first technique is natural language processing which can interpret a meaning of word and its grammar which can be used with various other techniques. The first technique we found useful by studying research papers by various authors is machine learning which can classify a URL in to malicious and safe classes. The other very powerful technique is to make use of artificial intelligence technique called as deep learning which can analyze any data deeply and predict good results thus help in detecting malicious URLs early. The mentioned techniques will also a database to keep track of the URLs and block them, thus making use of cloud computing mandatory for such kind of applications as cloud data can be accessed any time anywhere.

So, in other words the main objective of this paper is to:

- Focus on security of personal data of users.
- Alert a user of malicious URLs at early stage.
- Study Natural Language Processing techniques which is base of text mining.
- Study various machine learning and artificial intelligence techniques.
- Propose a new fusion frame work with a combination of NLP, Machine Learning, Artificial Intelligence and Cloud together.
- Create a blacklist database of our own without depending on the internet for it.
- Evaluate and analyze the new malicious URL detection framework and its strengths.

Thus, the rest of the paper is structured as follows:

- Section II. explains literature survey which studies various techniques with their advantages and drawbacks.
- Section III. explains the methodology i.e., mathematical model and algorithms to be used by the system.
- Section IV. explains proposed system with block diagram or system architecture and working of the system.
- Section V. shows the results of how the application are implemented and how they can be used.

III. METHODOLOGY

Our projects mathematical perspective can be put and described as given below.

1 Relevant Mathematics Associated with the Project A.

1.1 Set Theory Applied to the Project
1. User: - Set(U)={U0, U1, U2, U3, U4}
U0 ∈ U = Capture Skin image using mobile camera.
U1 ∈ U = Establish communication with cloud.
U2 ∈ U = Send image to cloud.
U3 ∈ U = Receive acknowledgement.
U4 ∈ U = Download and view Image prediction from cloud i.e. benign or malignant

2. Admin: - Set(A)={U1, A0, A1, A2, A3, A4, A5}
U1 ∈ A = Establish communication with cloud.
A0 ∈ A = Download skin image from ISIC. A1 ∈ A = Apply SIFT on database images.
A2 ∈ A = Save SIFT features in Oracle DB.
A3 ∈ A = Create training dataset.
A4 ∈ A = Download user test image from cloud.
A5 ∈ A = Apply Naive Bayes and send prediction to cloud i.e. benign or malignant.
.

1.2 Probability, NP-Hard and NP-Complete:

So, by studying the sets as defined above we come to notice that a element U1 is common in both modules and used in coordination in both sets which can be placed as

$x \in U \cap A$ i f $x \in U$ and $x \in A$ (A.1)

the probability of intersection of element in both modules can be given as

$P(U \cap A) = P(U) + P(A)$ (A.2)

So, intersection of common element can be shown as

$U \cap A = \{U1\}$ (A.3)

The conditional probability of both modules using the same element can be shown as

$P(U|A) = P(U \cap A)\ P(A)$ (A.4)

Thus, we conclude that our project "Skin Cancer Detection Using Machine Learning and Cloud" success and failure will depend upon the internet as it send test image from users phone to admin, i.e., if the internet connection is not good or not present the test image will not be fetched and the project won't work, thus this is a case of failure, so our project supports NP-Hard and not NP-Comp

IV. PROPOSED SYSTEM

This This section is mainly divided in 3 main modules with other sub parts in them. The text that follows explains the modules with a block diagram or system architecture as shown in Fig.2. to illustrate them. The working of the framework is explained as:

- **Blacklist URL dataset:**

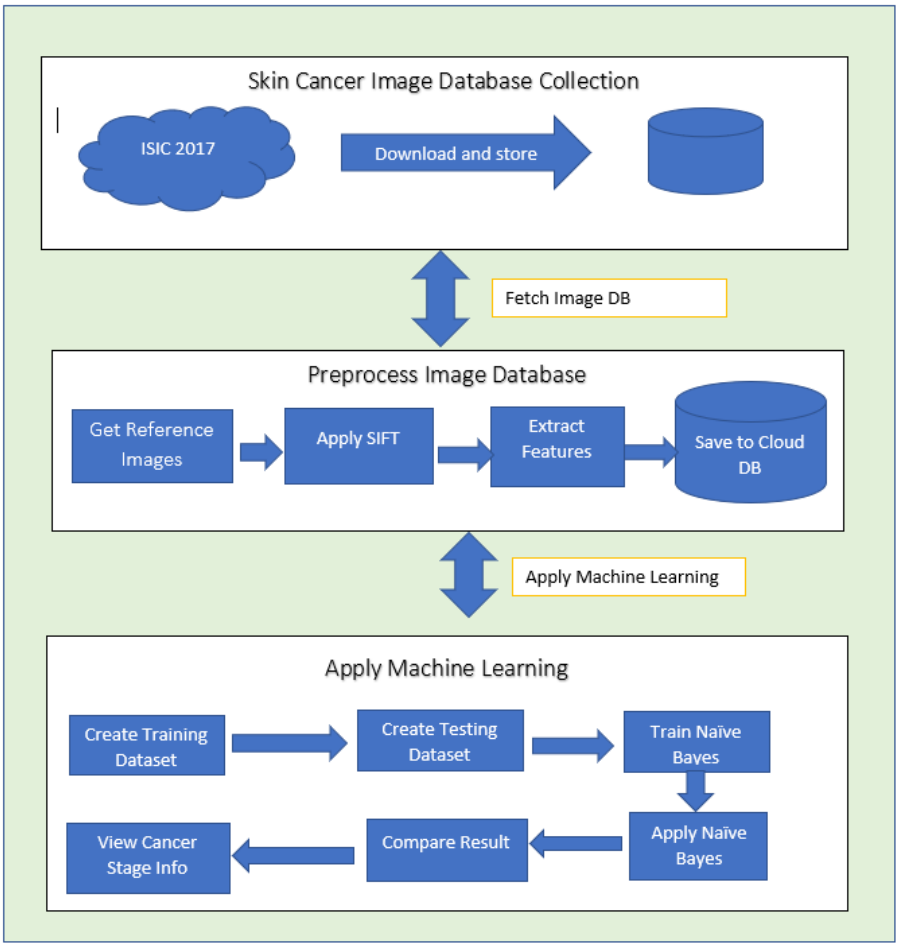
In this module we are proposing to use Google Drive as Cloud. The Google Drive sub cloud Google Sheets will be used to store blacklisted URL's and feature keywords. The google sheets is free cloud where the information can be stored and retrieved at any time anywhere. To store URL and keywords we will have a form in desktop application which will use google sheets as backend. The communication with the cloud and desktop application will be done using Google Drive and Google Sheets API. This cloud model will be used by both techniques i.e., Malicious URL detection using SVM and R-CNN.

- **SVM:**

In this module we first propose to fetch the backlisted dataset from google cloud. Then feature extraction will be done on blacklisted URLs using OPEN-NLP. The extracted feature keywords will be used to create a training dataset which will train SVM. Then a testing URL will be passed from which a testing dataset will be created. Both training and testing datasets will be passed to SVM algorithm which will give a classification results in the form of two classes i.e., safe and malicious.

- **R-CNN**:

This this module we also first propose to fetch the backlisted dataset from google cloud. Then feature extraction will be done on blacklisted URLs using OPEN-NLP. The extracted features will be used to create a text vector which will be used to train a R-CNN model. The R-CNN model will have all the layers necessary to predict the results properly. The layers will be fine-tuned for better predictions if necessary. Then after training the R-CNN a URL from dataset or any other URL will be passed to it for prediction. The prediction will be in the form of two labels i.e., safe or malicious



System Architecture Diagram

V. RESULTS AND DISCUSSION

Fig.1.Login screen

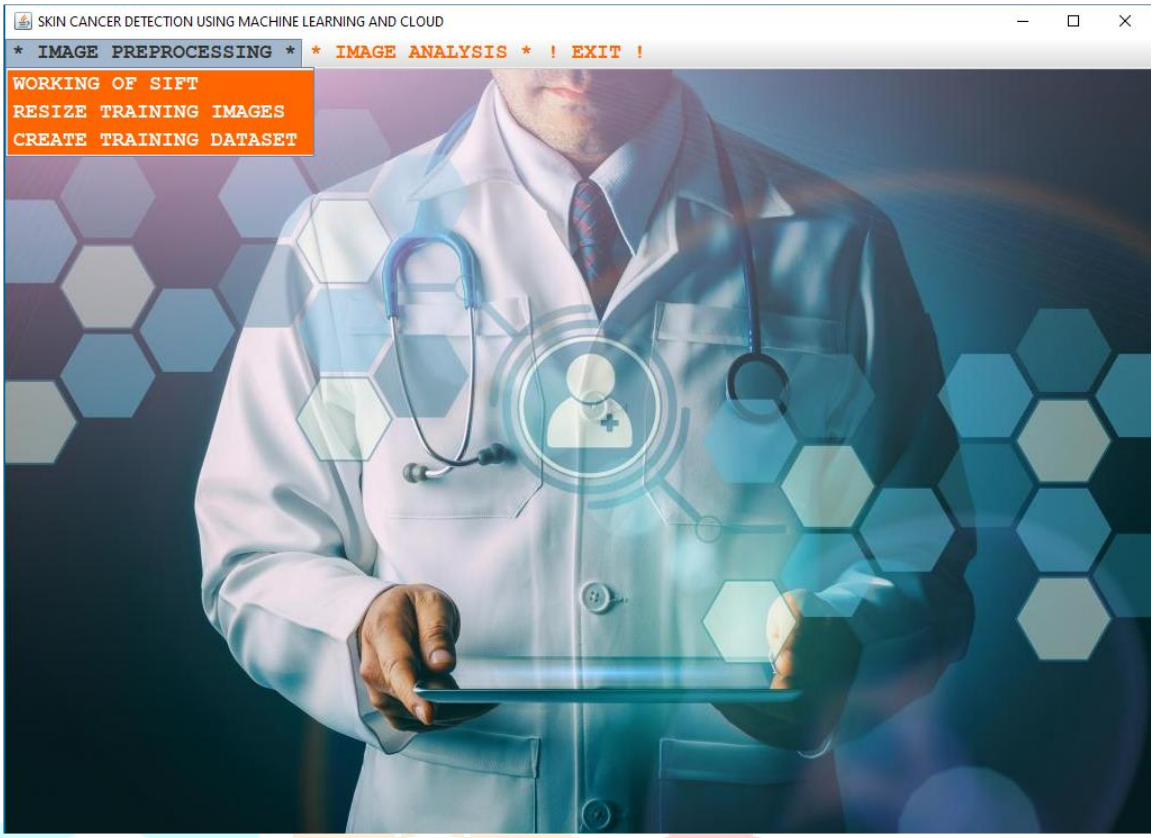


Fig.2.Menu screen

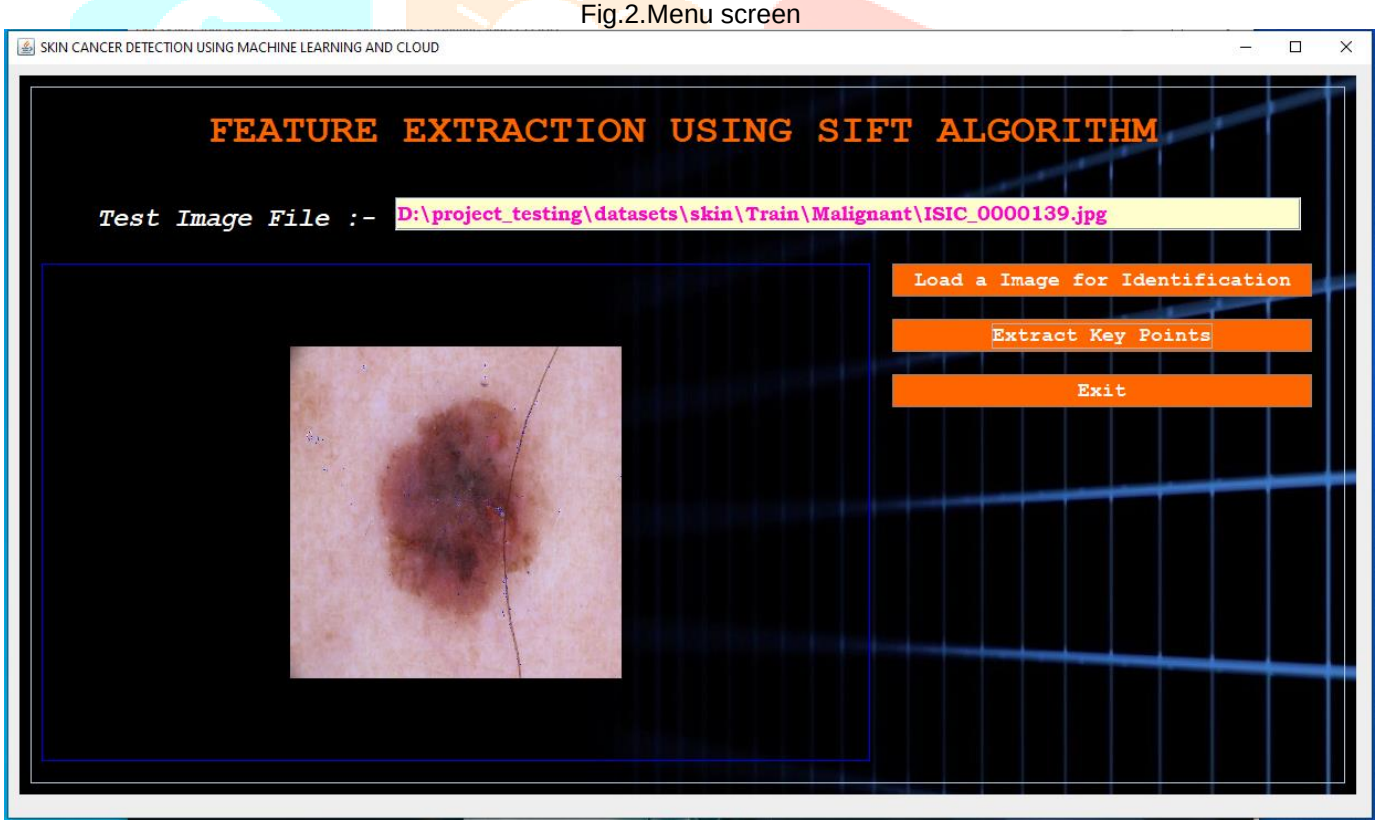


Fig 3.Demonstrate SIFT working screen

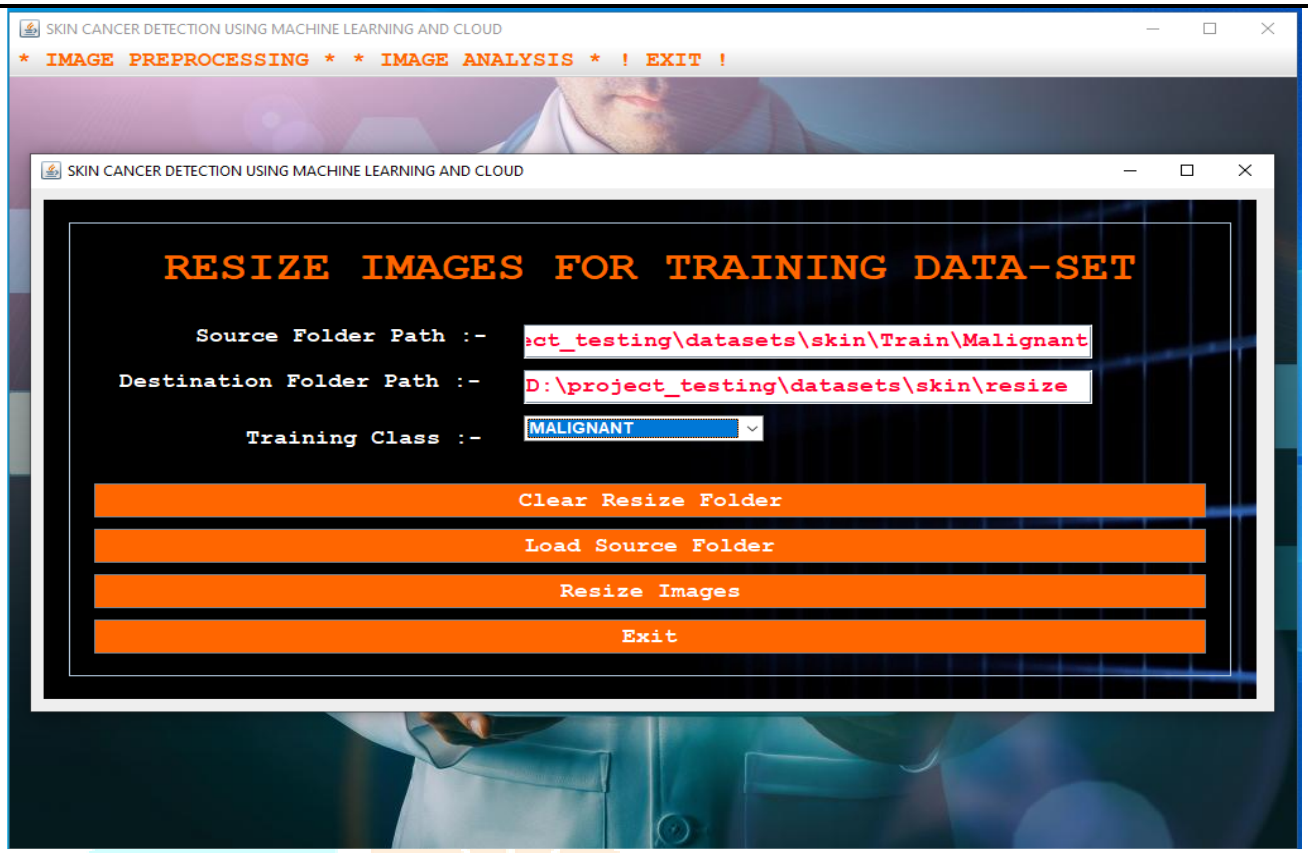


Fig 4.Resize training images screen

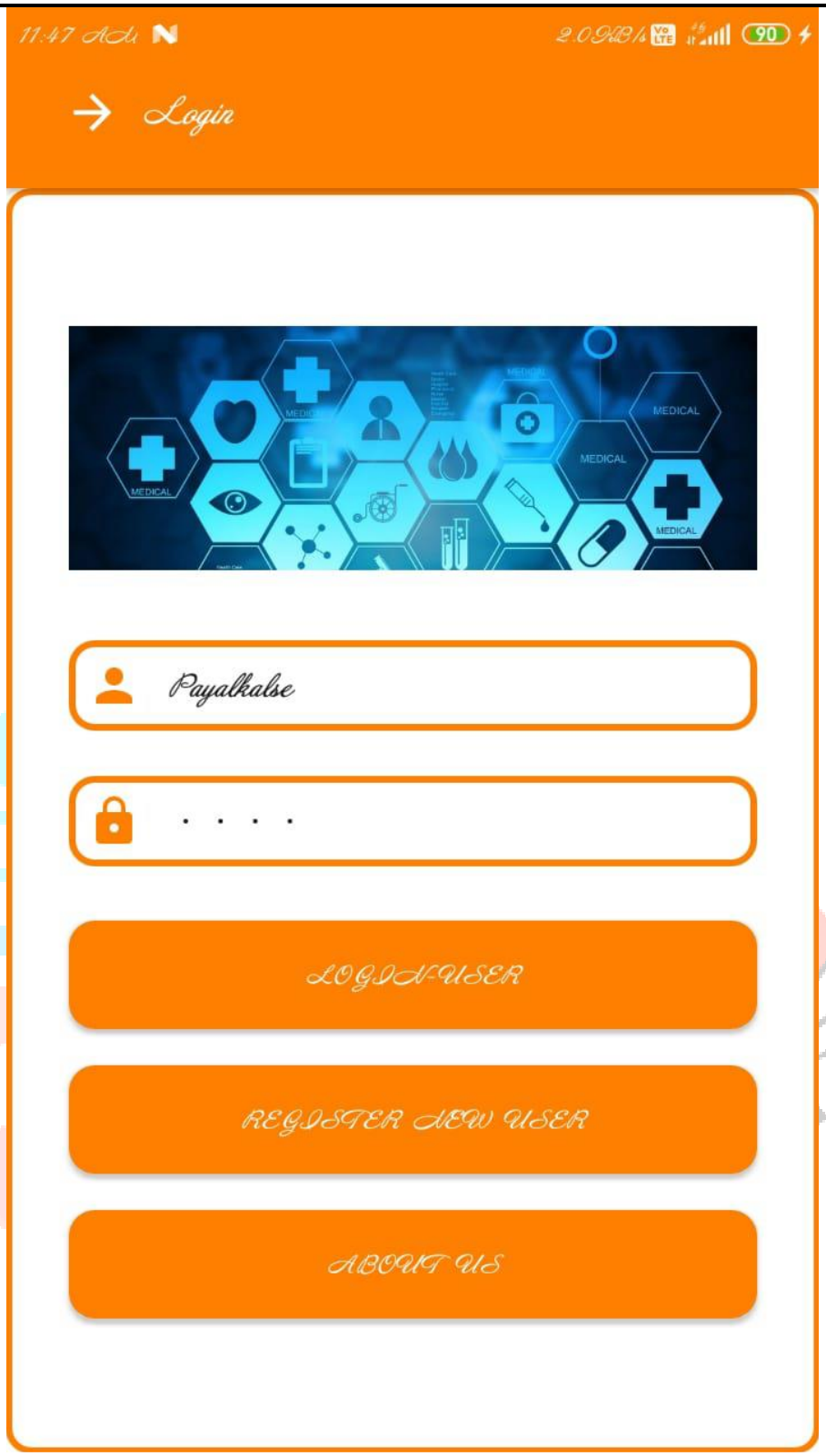


Fig 5.Mobile login screen

VI. Conclusions

In This project going to develop a novel approach of Skin Cancer Diagnosisusing image processing and machine learning together. To get the result more accurate we are going to use large skin cancer image database with machine learning. We will try to combine and make use of Java as our programming language. Thus, our project will ease work of a medical practitioner and give him more accurate diagnosis of skin cancer. Thus, we conclude that our application will be helpful in saving alot of lives by early diagnosis of a skin cancer.

REFERENCES

[1] Fulgencio Navarro, Marcos Escudero-Vinolo and Jes ˜ us Besc ´ os., et al. ” Accurate ´ segmentation and registration of skin lesion images to evaluate lesion change.”IEEE-2019

[2] Feng Guo, Jie Yang, Yilei Chen, Bao Yao., et al. Research on Image Detection and Matching Based on SIFT Features .” IEEE-2018.

[3] Tu Yawen and Guo Jinxu., et al. Research on Vehicle Detection Technology Based on SIFT Feature.” IEEE-2018.

[4] H. Cheng, et al. “Computer-aided detection and classification of microcalcifications in mammograms: a survey”, Pattern recognition, 36(12), 2967-2991, 2003.

[5] M. E. Celebi, et al. “Lesion border detection in dermoscopy images”, Comput. Med. Imag. Graph., 33(2), 148-153, 2009.

[6] C.A.Z. Barcelos, V.B. Pires. “An automatic based nonlinear diffusion equations scheme for skin lesion segmentation” Applied Mathematics and Computation, 215, 251–261, 2009 .

[7] D. H. Chung and G. Sapiro, “Segmenting skin lesions with partialdifferentialequations-based image processing algorithms,” Trans. on Medical Imaging., 19, 763–767, 2000.

[8] R. Kaur et al. “Thresholding methods for lesion segmentation of basal cell carcinoma in dermoscopy images”, Skin Research and Technology, 23(3), 416-428, 2017.

[9] R. Garvani, et al. “Border detection in dermoscopy images using hybrid thresholding on optimized color channels”, Comput. Med. Imag. Graph., 35(2), 105- 115, 2011.

[10] M. Silveira, et al., “Comparison of segmentation methods for melanoma diagnosis in dermoscopy images” Journal of Selected Topics in Signal Processing