IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

SMART DOCTOR ANDROID APPLICATION FOR BREAST CANCER RISK PREDICTION AND DIAGNOSIS

Samruddhi Deshmukh[1], Nikita Umredkar[2], Esha Sharma[3], Ranjita Chalke[4]

[1-3]BE Student, Department of Information Technology, MES Pillai College of Engineering, Navi Mumbai, India,

[4]Assistant Professor, Department of Computer Engineering, MES Pillai College of Engineering, Navi Mumbai, India.

Abstract: Medical consultation and prediction has become an interactive decision-making process for patients. E-healthcare and telemedicine are now widely accepted as a method for remote consultation between patient and doctor. This paper proposes an android mobile health application model for breast cancer diagnosis and daily heath prediction with the help of which we can improve the quality of treatment for patients. Breast cancer is one of the primary reasons for the death of women in recent times, being the second most common cause of cancer deaths of women worldwide. A lot of research has been done on early detection of breast cancer so as to allow starting of treatment early to increase the chance of survival. Breast Cancer Diagnosis is distinguishing benign from malignant breast lumps. Machine Learning Techniques are used as an approach for diagnosis of this disease. The user interface will be implemented using android/java programming, and the disease diagnosis part is designed with the help of symptoms-disease data and ML prediction techniques.

Index Terms **- E-healthcare, Breast Cancer, diagnosis, Early Detection of Breast Cancer, Machine Learning, ML Prediction Techniques, Android app, Smart Doctor.**

I. INTRODUCTION

The deaths caused by Breast Cancer can be reduced by early detection of the cancerous cells. This application helps predict cancer helping patients and doctors to conduct remote analysis. Cancerous cells are detected by performing various tests like MRI, mammogram, ultrasound and biopsy. Diagnosis of breast cancer is dependent on the type of tumour present. Tumours are of two types i.e., benign and malignant. Malignant tumours are more harmful than the benign. Unfortunately, not all physicians are expert in distinguishing between the benign and malignant tumours and the classification of tumour cells may take up to several days. Machine learning algorithms are used to predict the type of cancerous cells efficiently and accurately. Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. The objective of using Machine Learning techniques is to give the most accurate results for the given inputs. The project is to create an android mobile health application model for breast cancer diagnosis and daily heath prediction. The users are doctors who can easily diagnose using this application and may also be used by patients skilled enough to provide required medical inputs.

II. LITERATURE SURVEY

In [2] the authors have classified the cells either as Benign(B) or Malignant(M) to diagnose cancer. Out of the many algorithms for cancer prediction and classification they have used Support Vector Machine (SVM) , Decision Tree (CART) , Naive Bayes (NB) and k Nearest Neighbors (kNN). For the training the Wisconsin Breast Cancer Dataset and SVM is used on it and it is also trained on the other algorithms. Each algorithm's accuracy is then compared.

In [3] Authors have proposed a Support Vector Machines (SVMs) that is, Supervised Machine Learning Algorithm which acts as a classification technique to predict whether breast cancer has occurred or not. SVM maps the input vector into a higher dimensional feature space and helps identify hyperplane which separates the data points into classes. This classifier is effective and reliable but the only disadvantage is that it is mainly effective in cases where the number of dimensions is more than the samples. They have created a mobile based health application called Consult-AI which helps to predict diseases and health issues depending on what the patient's symptoms are. For this they have used Artificial Intelligence, Natural Language Processing and Android.

In [8] Naive Bayes Classifier has been used for the construction of classifiers. The Machine Learning Techniques used in this are Support Vector Machine, Logistic Regression, KNN and Naive Bayes. The Naive Bayes probabilistic classifier assumes that the values of features are independent of each other given the class variable. The advantage of using this classifier is that it requires

a lesser amount of training data and lesser parameters but requires prior calculation of probability. In addition to this they have used the KNN (K- Nearest Neighbour) which is a supervised learning technique which is used for diagnosing breast cancer.

III. METHODOLOGY

3.1. Existing System Architecture

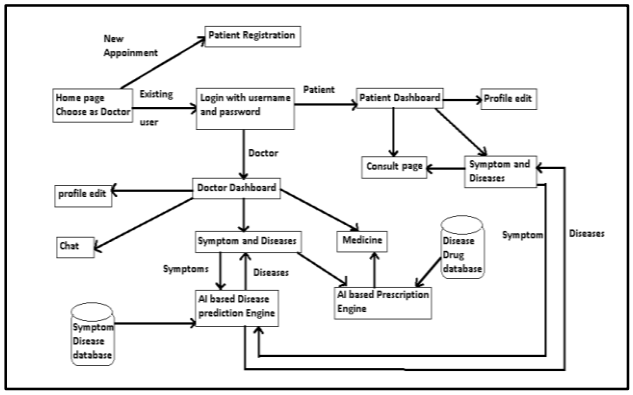

Figure 3.1.1 existing system architecture

The above figure 3.1.1 is an existing system which is an Android App for Medical Consultation system based on Artificial Intelligence techniques. The system is built using AI technique . For these users need to enter the issue or symptoms .The data is fetched from symptoms disease database and for users response , application displays output as a set of diseases with their probability in percentage .

Login and Registration

In these, users and doctors both need to enter their credentials for login purposes . If not a valid user ,they can register themselves through a new registration page .

Patients Dashboard

After successful login, patients can self evaluate for disease prediction by entering the symptoms from which they are going and AI based technique will give accurate results within a short duration of time . This gives an idea to patients how serious the condition is and this makes it easier for them to decide to visit a doctor or not .

Doctor's Dashboard

The doctor's list available in the application are specialists in respective domains and these are registered in the app for consultation to patients . Doctors can see medical history of patients also they can chat with patients .They have a facility to access AI based medicine prediction systems which makes it easier for them to give accurate treatment for each disease .

Medicine prediction

The access for these is permissible only to doctor's .It predicts medicine lists according to the disease as symptom disease database recommends medicine accurately based on the symptoms of the particular illness .

Consultation Page

Patients can take appointments of doctors online in advance with time and date . They can take appointments offline as well as for online interaction with specialists in respective fields .

3.2 Proposed System Architecture

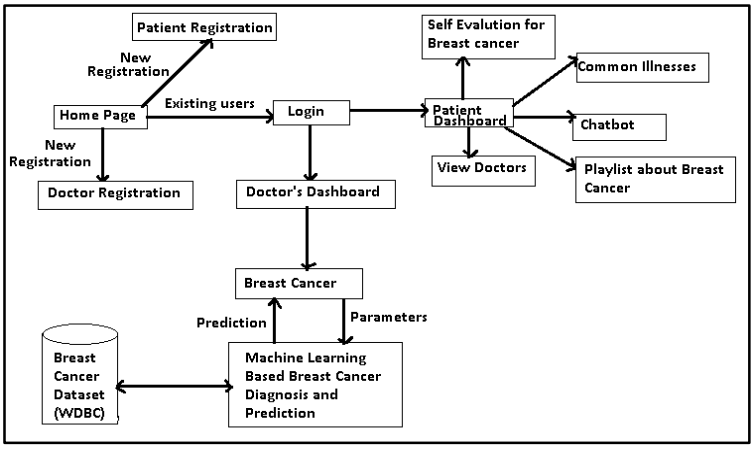

Figure 3.2.1 proposed system architecture

Splash screen:

Splash screen consists of the application logo which is the landing page of the application.

Home screen:

After Splash screen, it switches to the home screen which consists of two options- Users and Doctor. The Users options consists of login and registration for the users-patients. The Doctor option consists of the login and registration for the doctors.

Login and Registration Screen (User):
The User after registration can login to access the patient dashboard.
Login and Registration Screen (Doctor):
The Doctor after registration can login to access the doctor dashboard.

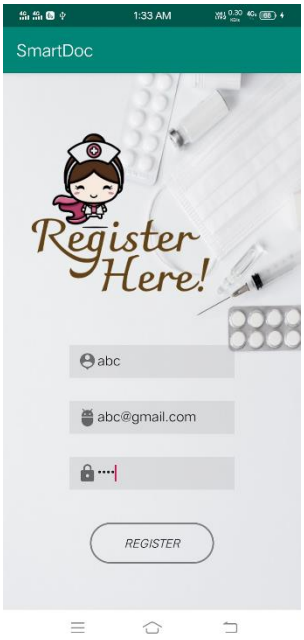

Figure 3.2.2 home screen Figure 3.2.3 registration page

Dashboard screen:
It is of two types:
(a) Doctor's dashboard: It consists of the Breast Cancer Prediction
(b)Patient Dashboard: Provides the options to View Doctors, Information and self-evaluation of common illnesses, Playlist containing information about breast cancer, Self-evaluation check for breast cancer and Chatbot.

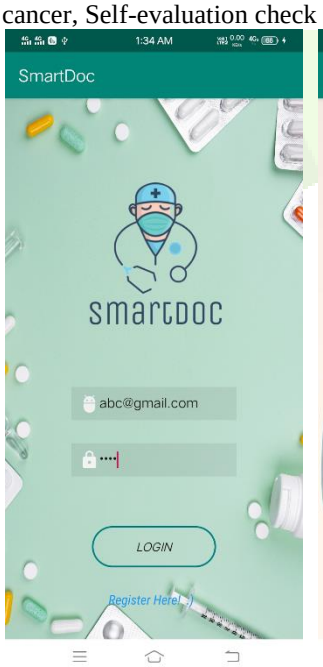

Figure 3.2.4 Dashboard of Patient

Symptoms and Diseases (Common diseases) screen:
The symptoms and diseases screen contains the information of the common diseases.
It contains a questionnaire for self-evaluation which predicts that the patient may have the disease or not.
It also contains treatment (some home remedies) on the same.
Related Videos:
A playlist having videos that contains information about breast cancer. Videos regarding what is breast cancer, self-evaluation for breast cancer, Stages of breast cancer etc. are added.

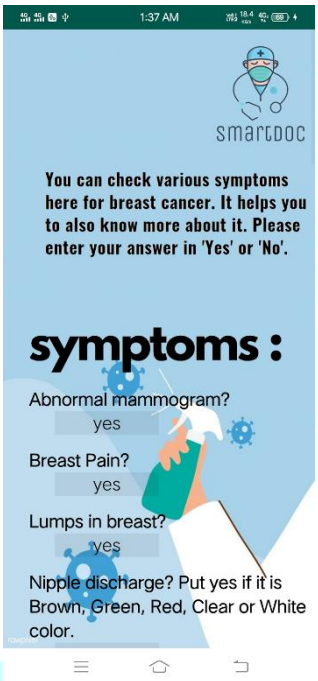

Figure 3.2.5 self evaluation

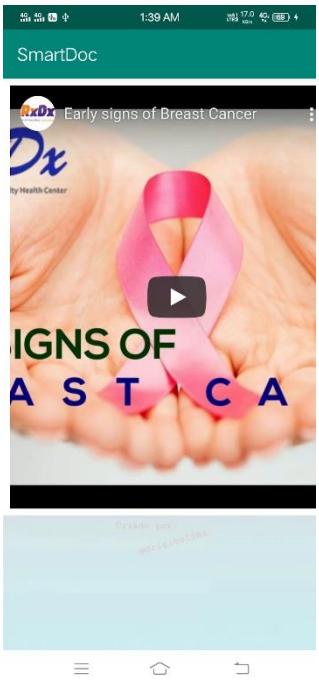

Figure 3.2.6 YouTube playlist

View Doctors:
It contains static information about the doctors and surgeons.
Chatbot:
Chatbot gives basic information about the common illness. Patients can interact with chatbot for quick information about the disease, symptoms of the disease.
Self-evaluation for breast cancer:
It contains a questionnaire which is used to self-evaluate for the possibility of breast cancer. Based on the result the patient can decide the course of treatment and consult the doctor regarding the same.

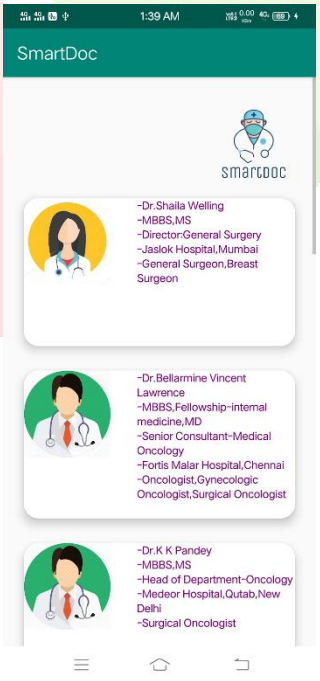

Figure 3.2.7 View Doctor

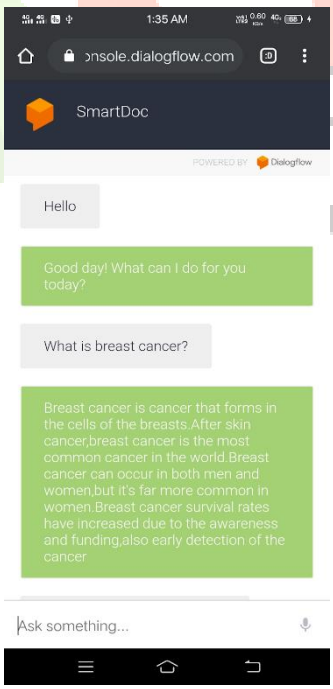

Figure 3.2.8 Chatbot

Machine learning based Breast cancer prediction and diagnosis:
We have used XGBoost classifier to predict whether the tumor is malignant or benign which in turn implies whether the patient has Breast cancer or not. The dataset is divided into a training set and testing set. 80% of the data is used to train the system and the remaining 20% is used for testing. The machine learning algorithm are trained on the trained data and tested on the testing data. The accuracy of XGBoost classifier used in the model for WDBC dataset is 98.24%.

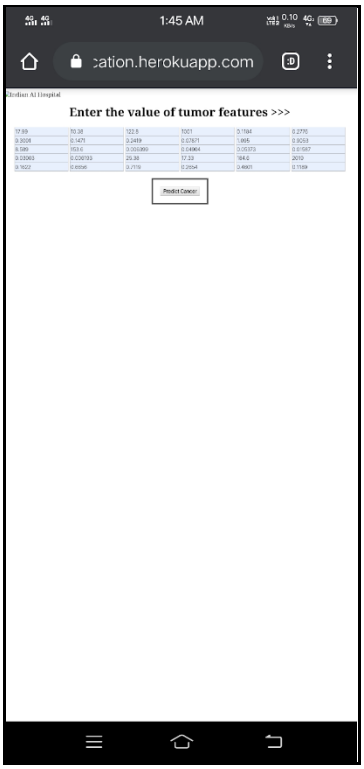

Figure 3.2.9 Prediction page

3.3 Dataset Used

Wisconsin Diagnostic Breast Cancer dataset is obtained from UCI depository. This dataset is available at the website https://archive.ics.uci.edu/ml/machine-learning-databases/breast-cancer/. .The dataset contains around 30 parameters and 569 instances .The dataset does not obtain any missing values . There are minimum and maximum values for every parameter for detection of cancerous cells. The dataset is Wisconsin Diagnostic Breast Cancer (WDBC) dataset. It consists of 569 instances. It has 32 attributes which includes two class attribute labels for diagnosis which is B= benign, M = Malignant, ID number and 30 real value attributes.

The goal is to accurately distinguish between benign and malignant tumors to aid in breast cancer diagnosis from the train dataset by applying the correct Machine Learning Technique algorithm.

Data set characteristics:

Number of Records: 569

Number of Attributes: 30 numeric, predictive attributes and the class

Attribute Information:

- radius (mean of distances from center to points on the perimeter)
- texture (standard deviation of gray-scale values)
- perimeter
- area
- smoothness (local variation in radius lengths)
- compactness (perimeter^2 / area - 1.0)
- concavity (severity of concave portions of the contour)
- concave points (number of concave portions of the contour)
- symmetry
- fractal dimension ("coastline approximation" - 1)

Class:

- WDBC-Malignant
- WDBC-Benign

3.4 Algorithm and methods used

We have used XGBoost classifier to predict whether the tumour is malignant or benign which inturn implies whether the patient has Breast cancer or not.The dataset is divided into a training set and testing set. 80% of the data is used to train the system and the remaining 20% is used for testing.The machine learning algorithms are trained on the trained data and tested on the testing data .

XGBoost classifier:

XGBoost stands for eXtreme Gradient Boosting. XGBoost is an algorithm that has recently been dominating applied machine learning and competitions for structured or tabular data. XGBoost is an optimized distributed gradient boosting library designed to be highly efficient, flexible and portable. The implementation of the algorithm was engineered for efficiency of compute time and memory resources.

Extreme Gradient Boosting or XGBoost is a library of gradient boosting algorithms optimized for modern data science problems and tools. Some of the major benefits of XGBoost are that its highly scalable/parallelizable, quick to execute, and typically outperforms other algorithms and uses a more regularized model formalization, to control over-fitting, which gives it better performance.

The two reasons to use XGBoost classifier are also the two goals of the project:

1.Execution Speed

2.Model Performance

VI. RESULT

To build the best model, we had train and test the dataset with multiple Machine Learning algorithms so that we could find the best ML model.To get more accuracy, we trained all supervised classification algorithms .After training all algorithms, we found that Logistic Regression, Random Forest and XGBoost classifiers are given higher accuracy than remaining but we have chosen XGBoost.The most accurate prediction according to WDBC dataset is given by XGBoost Classifier with accuracy of 98.245%. It determines whether the tumor is benign or malignant which in turn determines whether the patient has breast cancer or not.

V. ACKNOWLEDGEMENT

We would like to express our heartfelt gratitude to our Principal Dr. Sandeep M. Joshi for providing a platform for us students to showcase our skills and knowledge.

We are thankful to the Department of Information Technology engineering and our Head of the Department Dr. Satish L. Varma for providing us with the necessary resources and help due to which we could work on this project.

Also we would like to thank our project coordinator Prof. Gayatri Hegde and Dr. Shushopti Gawade who helped us a lot in finalizing this project within the limited time frame and constantly supporting us.

We extend our thanks to our guide Prof. Ranjita Chalke who guided us by providing us with her valuable suggestions in numerous consultations on this project. Without her constant supervision, guidance and motivation our project would not have been at par

REFERENCES

[1] Sindhumol S, Athira Ardhet , " Artificial Intelligence based Android App for Medical Consultation " , International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-9, July 2019.

[2] David A. Omondiagbe ,Shanmugam Veeramani ,Amandeep S. Sidhu(Curtin University, Malaysia) , "Machine Learning Classification Techniques for Breast Cancer Diagnosis " ,IOP Conf. Series: Materials Science and Engineering , 2019.

[3] Anusha Bharat , Pooja N , R Anishka Reddy (2018) , " Using Machine Learning algorithms for breast cancer risk prediction and diagnosis " , IEEE Third International Conference on Circuits, Control, Communication and Computing.

[4] Ambien Fred M. Agarap ,"On Breast Cancer Detection :An Application of Machine Learning Algorithms on Wisconsin Diagnostic Dataset " ,Association for Computing Machinery , ACM ISBN 978-1-4503-6336 , February 2018.

[5] M. Tahmooresi , A. Afshar ,B. Bashari Rad ,K.B. Nowshath, M.A.Bamiah , "Early Detection of Breast Cancer Using Machine Learning Techniques " Journal of Telecommunication, Electronic and Computer Engineering ,eISSN :2289-8131 ,Vol .10 , September 2018.

[6] Rida Sara Khan, Asad Ali Zardar, Zeeshan Bhatti , " Artificial Intelligence based Smart Doctor using Decision Tree Algorithm " , Journal of Information & Communication Technology - JICT ISSN 1816-613X , Vol. 11 Issue. 2 , December 2017.

[7] Hiba Asria ,Hajar Mousannifb ,Hassan El Moatassime ,Thomas Noeld,"Using Machine Learning Algorithms for Breast Cancer Risk Prediction and Diagnosis",The 6th International Symposium on Frontiers in Ambient and Mobile Systems , Procedia Computer Science 83 (2016) ,November 2016.

[8] Mandeep Rana , Pooja Chandorkar , Alishiba Dsouza , Nikahat Kazi , " BREAST CANCER DIAGNOSIS AND RECURRENCE PREDICTION USING MACHINE LEARNING TECHNIQUES " ,International Journal of Research in Engineering and Technology ,eISSN: 2319-1163 | pISSN: 2321-7308 , Volume: 04 Issue: 04 , April 2015.

[9] SK Kharya , D. Dubey , S. Soni ," Predictive Machine Learning technique for Breast Cancer Detection " , International Journal of Computer Science and Information Technologies ,Volume 4(6), ISSN :0975-9646 , 2013.