

Development of Web Application for Breast Cancer Detection using Machine Learning Classifier

Shital Pawar¹, Prachi Shukla², Priyanka Bagal³, Asawari Dawkhar⁴ and Dr. Rohini Jadhav⁵

¹⁻⁴Computer Engineering Department, Bharati Vidyapeeth's College of Engineering for Women Pune, India
Email: shitalp16@gmail.com, prachishukla646@gmail.com, bagalpiyu@gmail.com, asawaridawkhar@gmail.com

⁵Computer Engineering Department, Bharati Vidyapeeth (Deemed to be University) College of Engineering, Pune

Abstract—Breast cancer among women is second leading cause of death and is one of the dreadful diseases. For manual detection through mammograms, not all general hospitals have the facilities required. So, for early detection, computerised automatic diagnosis system should be developed. We have used machine learning for solving this problem. We have used machine learning classifiers to make a model which will predict whether the person has breast cancer or not using Wisconsin Diagnostic Breast Cancer (WBCD) dataset from UCI repository. In the healthcare domain, accuracy of the system is the main concern. One false prediction can lead to someone's death. So, for reaching the maximum accuracy, we tried comparing efficiency and effectiveness of different classifier algorithms available in machine learning. For the ease of use, we have integrated machine learning model with web application using flask and firebase database where user should enter the values of 30 features and he will get to know whether the person has breast cancer or not.

Index Terms— Machine learning, Breast Cancer, Support vector classifier, Naïve Bayes Classifier, Random forest classifier, K nearest neighbor classifier, XGBoost classifier.

I. INTRODUCTION

Breast cancer is the malignant tumor that is found in the cells of the breast [1]. It can spread to other parts of the body if detected at the last stage. This disease is one of the most death causing disease for women. During the past five years, the survival rates of breast cancer patients are about approximately 60% in India [2]. The mortality rate can be reduced if disease gets detected in initial stage only [3]. Early detection can help to improve the outcome as there is no mechanism for prevention of breast cancer till date [4]. Seeing this requirement, we decided to build a web-based application for breast cancer detection using machine learning algorithms [5] [6]. Machine learning has become an important part of medical research. Machine learning methods have evolved over the years from manual inputs to automatic initializations [7].

The advancements in the machine learning have led to more intelligent and self-reliant systems, as the learning ability of machine learning methods has been constantly improving [8]. Machine learning techniques offer various statistical methods and probabilistic that allow system to learn from past experiences and identify patterns from a dataset [9]. Seeing this, machine learning algorithms can predict accurate breast cancer. Healthcare providers and Physicians can take appropriate decision about treatment of breast cancer patient if the

prediction is precise and timely.

Using machine learning techniques as an effective mechanism for early identification and diagnosis of breast cancer will highly increase the patient's survival rate. So, here our main aim is to develop a machine learning model which will predict whether the person has breast cancer or not.

II. MOTIVATION

Now a day's many women are suffering from breast cancer. Main causes of this disease depend on many human factors and it cannot be discovered easily. The identification method which will decide whether the cancer is malignant or benign additionally requires outstanding efforts from physicians and doctors. The use of standard and extreme machine learning is more superior in terms time whenever there are large number of samples are available. This has tremendously increased the demand for the use of computing and machine learning as diagnostic tools. Using data processing and machine learning as a tool for diagnosis of breast cancer which is one of the critical diseases in the medicine where job of classification plays an important role in diagnosis of this disease. Hence the machine learning techniques will help doctors to appropriately identify the breast cancer and makes the correct classification of malignant and benign tumor.

III. RESEARCH GAP

Many scientists and researchers have applied machine learning and deep learning techniques to diagnose the breast cancer easily. For classification, using DNN model and for the selection of subset of features from all features of Wisconsin Breast Cancer Dataset, using RFE gave accuracy of 98.62% for 80-20% partitioning. But applying neural network on data mining databases has one major limitation of excessive processing [10]. DNN also works good for mammographic images [11]. Applying Black Propagation gave accuracy of 98.28% and Feed Forward and Black Propagation gave an accuracy of 92% [12]. From NB, Support Vector Machine, Random Forest and LR, Random Forest is the best algorithm with outstanding accuracy. It achieved lowest error rate along with highest precision [13]. RVM also gave good classification accuracy for breast cancer dataset [14]. The breast cancer is normally malignant tumor which actuates in the breast cells. This type of tumor has the capability to grow rapidly in the other parts of body too. Healthcare providers and Physicians can take appropriate decision about treatment of breast cancer patient if the prediction is precise and timely. Machine learning would make it easier to detect the cancer without much effort. All the methods implemented so far are just detecting whether the person has breast cancer or not which is a limitation of this research area. More research should be done for detecting the various stages of breast cancer.

IV. PROPOSED METHODOLOGY

For breast cancer detection, we have compared the accuracies of different classification algorithms namely SVM, KNN, Random Forest Classifier, Ada Boost Classifier and XGBoost Classifier to see which algorithm is giving best results for Wisconsin's Breast Cancer Dataset. To perform all the experiments, we used sickit learn python libraries in Anaconda Environment.

The Wisconsin's Breast Cancer Dataset has 699 rows out of which 458 are benign and 241 are malignant. The 'target' is an independent variable which represents whether the person has breast cancer or not. Target value 0 means the person has cancer and the target value 1 means the person don't have cancer. There are no null values in the dataset and all the features are in float64 format.

For the visualization, we plotted count plot of the feature 'mean radius' to get the clear insight about the 'target' attribute and we got to know that the patients that do not have breast cancer have value of 'mean radius' close to 1 and the for the person who has breast cancer the value of 'mean radius' is greater than 1. From the heatmap, the value of 'mean area' was in between 800 and 2400, the value of 'worst area' was in between 800 and 2000 and all other features had values near to 0. We took the correlation of the dataset as well and also plotted correlation barplot of all features with respect to 'target'. 'Smoothness error' had positive correlation with the 'target'. 'mean fractal dimension', 'text error' and 'symmetry error' had very low correlation with 'target' attribute.

From the heatmap that we have plotted, it was clear that the features in our dataset are not in same level of magnitudes. The value of 'mean area' is in between 800 and 2400, the value of 'worst area' is in between 800 and 2000 and all other features have values near to 0. It is necessary to bring all the feature values at same level of magnitude for better performance. We achieved this by scaling using Standard Scalar. We gave this scaled dataset to the algorithms and the accuracy given by the XGBoost Classifier is highest which is 98.24%. We

further performed parameter tuning with Randomized search and grid search using XGBoost classifier. We got same accuracy after parameter tuning which is 98.24%. The results obtained for the algorithms are shown in Table I.

So, lastly, we trained our model with XGBoost Classifier as it was giving best performance. To deploy the machine learning model on local servers we need to save it first. We have used pickel package in python for saving the model. This model can be deployed to any local server or remote cloud.

TABLE I: ACCURACIES OF THE ALGORITHMS

Algorithm	Accuracy Obtained
Support Vector Machine	96.49%
Decision Tree Classifier	94.73%
K-Nearest Neighbour	93.85%
Random Forest Classifier	97.36%
Ada-Boost Classifier	94.73%
XGBoost Classifier	98.24%

In python we have Pickle library for loading the pickle file and Flask library to deploy the model on localhost. Using pickle.load() method we loaded the model into flask and same model can be used for predicting the breast cancer. We created web form for accepting the values of 30 features using Html and Css. Using request. Form () we can get the values entered by user into flask app. By creating data frame from the values entered by the user we can pass it to the predict () method of the pickle which will give the result.

We have used firebase for storing the credentials of the user. Using Pyrebase library of the python connection between firebase and flask is established. The procedure for of breast cancer detection through our web application is represented in fig.1.

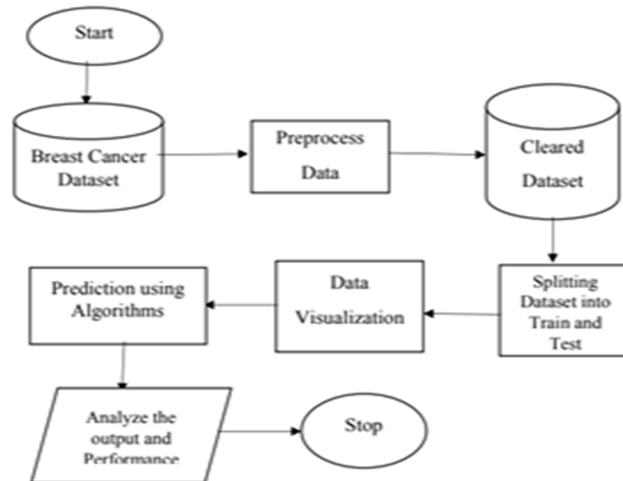


Fig 1: Procedure for the detection of breast cancer

V. RESULTS AND DISCUSSION

From Support Vector Machine, K-Nearest Neighbor, Random Forest Classifier, Ada Boost Classifier and XGBoost Classifier, XGBoost classifier gave the highest accuracy for Wisconsin Breast Cancer Dataset. SVM and Random Forest Classifier also gave good results 96.49% and 97.36% respectively. We further performed parameter tuning with Randomized search and grid search using XGBoost classifier. We got same accuracy after parameter tuning which is 98.24%. Precision for malignant tumor is 1 and recall we got 0.95 and that for benign tumor, precision recall we got are 0.96 and 1 respectively. Two patients who did not had breast cancer detected

incorrectly by the model but we got zero type II error for XGBoost classifier. So, the patient who has breast cancer will never be detected incorrectly by the model and model is safe to use in real world.

User can register into the system by providing Email ID, Name, Mobile Number and Name of the hospital. After successful registration he can login into the system with Email and Password. If the login is successful, he will be redirected to the prediction page where he is supposed to enter the values of 30 features. After submitting this data immediately, he will see whether the person has breast cancer or not. The type II error of the system is zero so the person who has breast cancer will never be detected incorrectly. Fig.2 shows the prediction page of the website. 30 text fields are provided for the values of the features. After submitting these details user will be redirected to the prediction page.

Breast Cancer Prediction

Mean Radius	Mean Texture
Mean Perimeter	Mean Area
Mean Smoothness	Mean Compactness
Mean Concavity	Mean Concave Points
Mean Symmetry	Mean Fractal Dimension
Radius Error	Texture Error
Perimeter Error	Area Error
Smoothness Error	Compactness Error

Smoothness Error	Compactness Error
Concavity Error	Concave Points Error
Symmetry Error	Fractal Dimension Error
Worst Radius	Worst Texture
Worst Perimeter	Worst Area
Worst Smoothness	Worst Compactness
Worst Concavity	Worst Concave Points
Worst Symmetry	Worst Fractal Dimension

SUBMIT

LOGOUT

Fig. 2: Prediction Page of the website

Fig. 3 shows the result page where user can clearly see the result whether person has breast cancer or not. On clicking OK button, he will be redirected to the prediction page again.

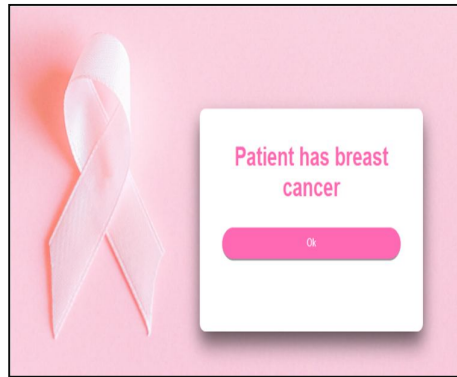


Fig. 3: Result Page of the website

VI. CONCLUSION

In the healthcare domain, accuracy of the system is the main concern. One false prediction can lead to someone's death. So, for reaching the maximum accuracy, we tried comparing efficiency and effectiveness of different classifier algorithms and found that XGBoost classifier is giving maximum accuracy of 98.24% for our dataset. Besides this, creating good user interface is also a challenge. The system should be user friendly. Using flask, firebase and pickle, we developed a web system where user can register himself and can easily see the result by entering the values of features.

In future, machine learning is set to change the medical industry. Using these techniques, even the stages of the breast cancer can be detected in coming decades.

REFERENCES

- [1] Bharat Gupta and Madhuri Gupta, "A Comparative Study of Breast Cancer Diagnosis using Supervised Machine Learning Techniques", *IEEE*, pp. 997-1002, 2018
- [2] Pinar Kirci, Ebru Aydindag Bayrak, and Tolga Ensari, "Comparison of Machine Learning Methods for Breast Cancer Diagnosis", *IEEE*, pp. 1-3, 2019
- [3] Naresh Khuriwal and Nidhi Mishra, "Breast Cancer Diagnosis Using Deep Learning Algorithm", *IEEE*, pp. 98-103, 2019.
- [4] Sneha Memane, Mansi Hiwase, Trupti Rajule, Aparna Chopade and Shital Pawar, "Prediction of Ophthalmic Disease using Image Processing and Machine Learning", *International Journal of Emerging Technologies and Innovative Research*, vol. 6, pp. 649-653, 2019.
- [5] N. Rane, J. Sunny, R. Kanade and P. Devi, "Breast Cancer Classification and Prediction using Machine Learning," *International Journal of Engineering Research and Technology*, vol. 9 , pp. 576-580,2020
- [6] Anji Reddy Vaka, Badal Soni and Sudheer Reddy K., "Breast Cancer Detection by leveraging Machine Learning", *ICT Express*, vol. 6, pp. 320-324, 2020.
- [7] Shital Pawar, Sneha Memane, Mansi Hiwase, Trupti Rajule and Aparna Chopade, "Analysis of Ocular Disease Using Multiple Information Domains", *International Journal of Innovative Research in Engineering & Multidisciplinary Physical Sciences*, Vol. 6, Issue 6,pp. 172-175, 2018.
- [8] S. A. Mohammed, S. Darrab, S. A. Noaman and G. Saake, "Analysis of Breast Cancer Detection using Different Machine Learning Techniques," *Communications in Computer and Information Science*, vol. 1234, pp. 108-117,2020.
- [9] Nur Syahmi Ismail, Cheab Sovuthy, "Breast Cancer Detection Based on Deep Learning Technique", *IEEE*, pp. 89-92, 2019
- [10] Karthik Sekaran, Srinivasa Perumal Ramalingam, Chandra Mouli, "Breast Cancer Classification Using Deep Neural Networks", *ResearchGate*, pp. 227-241, 2019.
- [11] R.Chtihakkannan, P.Kaviitha, T.Mangayarkarasi, R.Karthikeyan, "Breast Cancer Detection using Machine Learning", *International Journal of Innovative Technology and Exploring Engineering*, vol. 8, pp. 3123-3126, 2019
- [12] Sunil Kumar, Cherry Bhargava, "A review on Breast Cancer Analysis using Machine Learning", *International Journal of Emerging Technologies and Innovative Research*, vol. 6, pp. 326-329, 2019
- [13] Sivapriya J, Arvind Kumar V, Siddarth Sai S and Sriram S, "Breast Cancer Prediction using Machine Learning," *International Journal of Recent Technology and Engineering*, vol. 8,pp. 4879-4881, 2019.
- [14] C.P.Sumathi., B.M.Gayathri and T.Santhanam, "Breast Cancer Diagnosis Using Machine Learning Algorithms- A Survey," *International Journal of Distributed and Parallel Systems*, vol. 4,pp. 105-112, 2013.