International Journal of Advanced Technology in Engineering and Science
Vol. No. 11, Issue No. 04, April 2023
www.ijates.com
ijates
ISSN 2348 - 7550

Breast Cancer Detection Using Machine Learning

K.Vindhya, M.Neelima, P.Sandeep

Under the guidance of Mr. ***M.Chennakesavarao****, Assosiate Professor, Department of Computer science and Engineering, Tirumala Engineering College*

ABSTRACT

Breast cancer is the second most common cancer in the world and most prevalent in the female population.The early detection and diagnosis is the optimal solution to prevent tumor progression and allow successful medical intervention, saves lives and reduce cost.The aim of project is to present the comparision of popular machine learning algorithms, techniques and the performance of various machine learning techniques commonly used for breast cancer detection.

Keywords— Cancer, recognition, Logistic Regression, KNN, SVM,Decision Tree

I. INTRODUCTION

"Breast Cancer Detection" Breast cancer is the uncontrolled growth of breast cells. This causes as a result of mutations, or abnormal changes, in the genes responsible fo r regulating the growth of cells and keeping them healthy[1]. The genes are in each cell's nucleus, which acts as the "control room" of each cell. Normally, the cells in our bodies replace themselves through an orderly process of cell growth: healthy new cells take over as old ones die out. But over time, mutations can "turn on" certain genes and "turn off" others in a cell. That changed cell gains the ability to keep dividing without control or order, producing more cells just like it and forming a tumour. A tumour can be benign (not dangerous to health) or malignant (has the potential to be dangerous). Benign tumours are not considered cancerous: their cells are close to normal in appearance, they grow slowly, and they do not invade nearby tissues or spread to other parts of the body Malignant tumours are cancerous. Left unchecked, malignant cells eventually can spread beyond the original tumour to other parts of the body.

The term "breast cancer" refers to a malignant tumour that has developed from cells in the breast. Usually breast cancer either begins in the cells of the lobules, which are the milkproducing glands, or the ducts, the passages that drain milk from the lobules to the nipple. Less commonly, breast cancer can begin in the stromal tissues, which include the fatty and fibrous connective tissues of the breast. Over time, cancer cells can invade nearby healthy breast tissue and make their way into the underarm lymph nodes, small organs that filter out foreign substances in the body.

If cancer cells get into the lymph nodes, they then have a pathway into other parts of the body. The breast cancer's stage refers to how far the cancer cells have spread beyond the original tumour. Breast cancer is always caused by a genetic abnormality (a "mistake" in the genetic material). However, only 5-10% of cancers are due to an abnormality inherited from your mother or father. Instead, 85-90% of breast cancers are due to genetic

abnormalities that happen as a result of the aging process and the "wear and tear" of life in general.

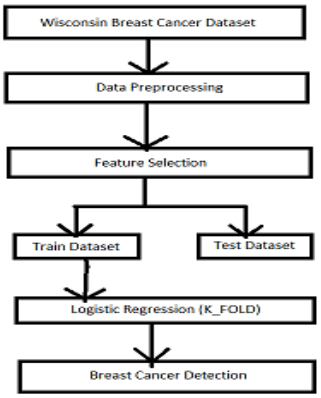
II. LITERATURE SURVEY

The existing system consists the data Ultrasound characterisation of breast masses by S. Gokhale written by proposed a system where they found that doctors have known and experienced that breast cancer occurs when some breast cells begin to grow abnormally. These cells divide more briskly and disperse faster than healthy cells do and continue to accumulate, form- ing a lump or mass that the may start causing pain. Cells may spread rapidly through your breast to your lymph nodes or to other parts of your body. Some women can be at a higher risk for breast cancer because of their family history, lifestyle, obesity, radiation, and reproductive factors. In the case of cancer, if the diagnosis occurs quickly, the patient can be saved as there have been advances in cancer treatment. In this study we use four machine learning classifiers which are Naive Bayesian Classifier, k-Nearest Neighbour, Support Vector Machine, Artificial Neural Network and random forest. Harmonic imaging and real-time compounding has been shown to enhance image resolution and lesion characterisation. More recently, USG elastography seems to be quite ncouraging. Initial results show that it can improve the specificity and positive predictive value of USG within the characterisation of breast masses. The reason why any lesion is visible on mammography or USG is that the relative difference within the density and acoustic resistance of the lesion, respectively, as compared to the encompassing breast tissue. **[3]**On Breast Cancer Detection: An Application of Machine Learn ing Algorithms on the Wisconsin Diagnostic Dataset by the Abien Fred M. Agarap. In this paper, six machine learning algorithms are used for detection of cancer . GRUSVMmodel is used for the diagnosis of breast cancer GRUSVM, Linear Regression, Multilayer Perceptron (MLP), Nearest Neighbour (NN) search, Softmax Regression, and Support Vector Ma- chine (SVM) on the Wisconsin Diagnostic Breast Cancer (WDBC) dataset by measuring their classification test accuracy, and their sensitivity and specificity values. The said dataset consists of features which were computed from digitised images of FNA tests on a breast mass.

For the implementation of the ML algorithms, the dataset was partitioned in the following fashion 70 percent for training phase, and 30 percent for the testing phase.

algorithms exhibited high performance on the binary classification of carcinoma, i.e. determining whether benign tumour or malignant tumour. Therefore, the statistical measures on the classification problem were also satisfactory. To further corroborate the results of this study, a CV technique such as k-fold cross-validation should be used. The appliance of such a way won't only provide a more accurate measure of model prediction performance, but it'll also assist in determining the foremost optimal hyper-parameters for the ML algorithms.

III. PROPOSED SYSTEM



We proposed to develop a system which will help practitioners to predict breast cancer disease based on some attributes like clump thickness, uniformity cell size, mitoses and so on. So, there is a need for developing a decision system which will help practitioners to predict the breast cancer in an easier way, which can offer prediction about the breast condition of patient so that further treatment can be made effectively. This proposed system not only accurately predicts breast cancer but also reduces time for prediction. The machine learning algorithms like decision tree, random forest, Naive Bayes, K Nearest Neighbours haveproven to be most accurate & reliable and hence, used in this proposed system.

IV. RESULTS

Among all those algorithms, the best performance is selected.

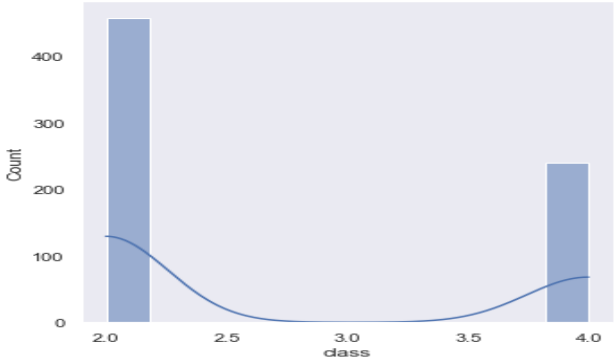
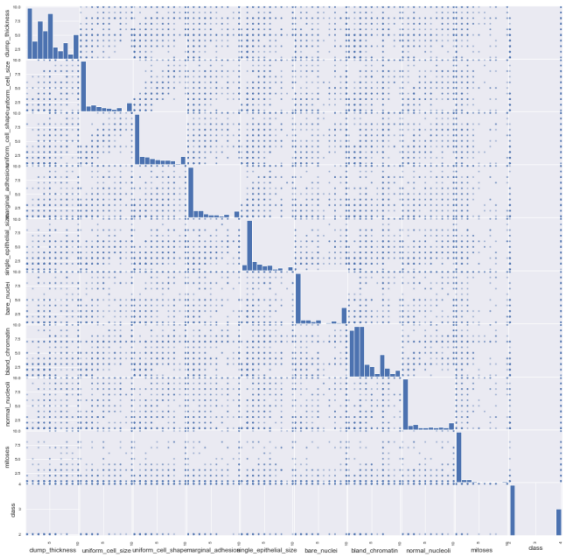


Fig: Bivariate Analysis



The web page is created. Then opting for the cancers is present on the left side. By choosing the symptoms, the Breast cancer can be predicted.

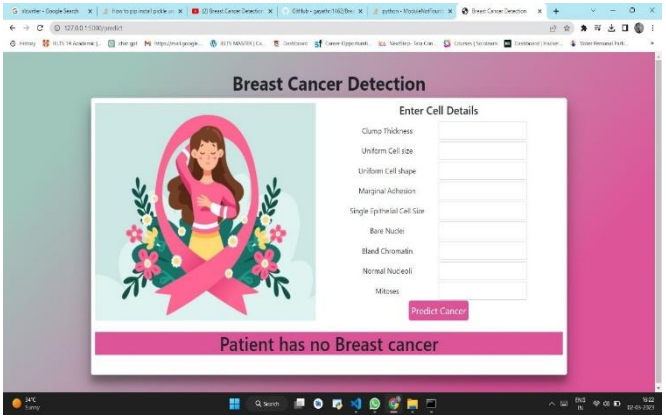


Fig: Breast Cancer web page

The Breast cancer is predicted by opting the symptoms.
Then result is given.
The website link for Cancer prediction is: Local URL : http://127.0.0.1:5000

CONCLUSION

We have used 4 algorithms like Classification and regression algorithm Naive Bayes, SVM, KNN and in-order to predict presence or absence of breast cancer Technique. The accuracy varies for different algorithms. The

accuracy for classification and regression algorithm is 93.33%. The accuracy for Naive Bayes algorithm is 93.80%.

REFERENCES

[1] K Kourou, Konstantinos P.Exarchos, Michalis V.Karamouzis, "Machine learning applications in cancer prognosis and prediction",2015.

[2] U.S. Cancer Statistics Working Group. United States Cancer Statistics: 1999–2008 Incidence and Mortality Web-based Report. Atlanta (GA): Department of Health and Human Services, Centers for Disease Control.

[3]. Madhu Kumaria, Vijendra Singhb Department of Computer Science and Engineering, The NorthCap University, Sector 23A,Gurugram, Haryana, 2017, India.

[4] Hiba Asria, Hajar mousannifb, Hassaan Al Moata ssimec, Thomas Noeld. The 6th International Symposium on Frontiers in Ambient and Mobile System (FAMS 2016).

[5] "UCI Machine Learning Repository: Breast Cancer Wisconsin (Original)DataSet." [Online].Available:https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+%28Original%29.

[6] Step 7 - Science Fair 2012: Project *Sites.google.com*, 2020. [Online].

[7] Data Pre-processing- an overview | Science Direct Topics, Sciencedirect.com