Research Paper

Implementation and Detection of Breast Cancer Using Different Machine Learning Algorithm

Tanmay Padhi[1], Aditi Kumari[1]

[1]Department of Electrical And Electronics Engineering, SRM Institute Of Science And Technology,
Kattankulathur, Chennai, 603203

***ABSTRACT:** This paper deals with breast cancer detection using different Machine Learning algorithm and having a comparative analysis among all of them. Here, we have taken a dataset and used different approaches namely Random Forest Classifiers, SVM and Bayesian Network to find out which one detects in a better way. Though ML is being used long back before to detect breast cancer. It is expected that the involvement of ML in diagnosis could provide the doctors a second opinion and help them to make a better diagnosis. Regarded the assisted cancer diagnosis, Google has reached to an accuracy level in identifying skin cancers, suggesting that the cancer accessibility diagnosis could potentially be extended for aside from medical clinics. In this paper we have put forward many different approaches which are used to detect breast cancer using Machine Learning and each of them give different efficiencies.*
***KEYWORDS:** SVM, KNN, Bayesian Network, ML, Prediction*

I. INTRODUCTION

Throughout the whole world, breast cancer is one of the most widespread disease among women. In such conditions, it is quite important to detect the early diagnosis during treatment. According to [1], India ranks on 2nd position when we talk about cancer in women among the world. The emergence of new revolution in medical field and advance technology has motivated the people to develop new strategy in prediction and detection of cancer [2]. Anyhow, there are certain data which are collected from the user or patient and on that basis only detection can be done. There are several tools available which help us to facilitate diagnosis. According to [3] breast cancer is caused by the mutation of one or more cells which can be cause reckless cell division. Machine Learning is a kind of the subfield of Artificial Intelligence in which machines are allowed to functioned by exposing them to specific datasets and enable them to perform certain tasks. Although here the datasets which are taken by the patients and physicians who contribute in such fields.

ML is a tool which is required for evaluation, comparison of algorithms based on the data available and test done on the model. So, in ML process, the machine takes the decision and predicts or detects the outcome [4]. Though ML is being used long back before to detect breast cancer [5]. It is expected that the involvement of ML in diagnosis could provide the doctors a second opinion and help them to make a better diagnosis. Regarded the assisted cancer diagnosis, Google has reached to an accuracy level in identifying skin cancers, suggesting that the cancer accessibility diagnosis could potentially be extended for aside from medical clinics [6]. In [7] they have proposed hybrid model which comprises of normalization, feature weighting, and classifier has been combined to classify the breast cancer dataset which can be used as alternative method. In this paper we have put forward many different approaches which are used to detect breast cancer using Machine Learning and each of them give different efficiencies. The rest of the paper explains about the fundamental concepts of each model which is followed by simulation setup and comparative study.

Machine Learning Techniques

ML techniques are basically categorized into two main types, supervised and unsupervised. In the former type, the datasets are the used to train the model and are expected to give correct results. Whereas if we see in unsupervised learning, there are no predefined data and hence there are no prediction or detection on the outcome. A term Classification comes in use when supervised learning is concerned, it is used to develop model and predict future results [8]. In this paper, all the methods used comes under classification techniques.

1.	KNN (K-Nearest Neighbor)- In this technique, prediction is done using the dataset directly. Prediction is made generally for a new instance after searching the whole dataset 'K' and predicting the output variable for K instance. To find the similarity in the dataset to a new input, distance method is used. Generally, Euclidian method is used to find the distance by calculating the distance between new and existing point in the dataset.

2.	Navies Bayes – These are collection of different classification of Bayes' theorem. This has many algorithms which can be used in machine learning for prediction of outcome but all the algorithm shares a common principle. This method makes easier for binary and categorial input value data to predict.

3.	Supervised Vector Machine – One of the most renowned method which can be used for classification and regression of data. In this method, the data is plotted in open space of n dimension with each value having a different coordinate. After performing the classification, the two class are differentiated very well. The marginal distance between two point that are closer to boundary is maximized. The obtained result now can be used for classification of new sample.

## II.	METHODOLOGY

1.	**Collection of Data**
To predict the breast cancer, we used Wisconsin dataset which was already available on Kaggle [9], containing 11 features and around 650 different instances. Of the 11 available features 10 were input feature while the remaining was the output feature. Around 80% of the datasets were used as a training dataset while the rest 20% were computed as testing datasets.

2.	**Pre-Processing of Data**
The available data has some missing values too, to overcome this problem data pre-processing was done The missing values were replaced by the mean value of the data and few attributes that were irrelevant were removed from the current dataset.

## III.	EXPERIMENT RESULTS

1.	**System Specification**
The model is trained on Google Collab and uses GPU of NVIDIA MX250 and RAM of 8GB and 4 hours of continuous use.

2.	**Results and Simulation**
Many different algorithms like SVM, KNN, NB were used for prediction of result and were compared to precision, accuracy and recall.

Table 1

Techniques	TP	FP	TN	FN	P	R	Accuracy(%)
SVM	82	1	54	3	0.987	0.964	97.14%
K-NN	83	3	52	2	0.965	0.976	96.52%
NB	81	3	53	4	0.975	0.953	95.71%

The above table shows the comparative study of different algorithms used and the accuracy in each method. Here we can see that there is very less margin among all of them for a given instance.
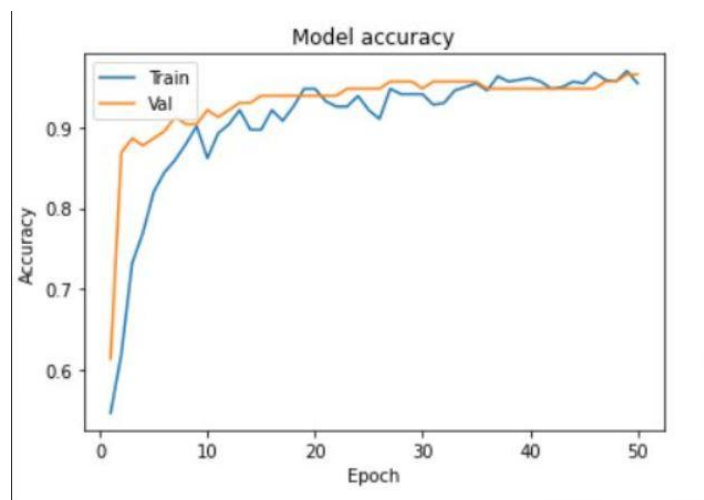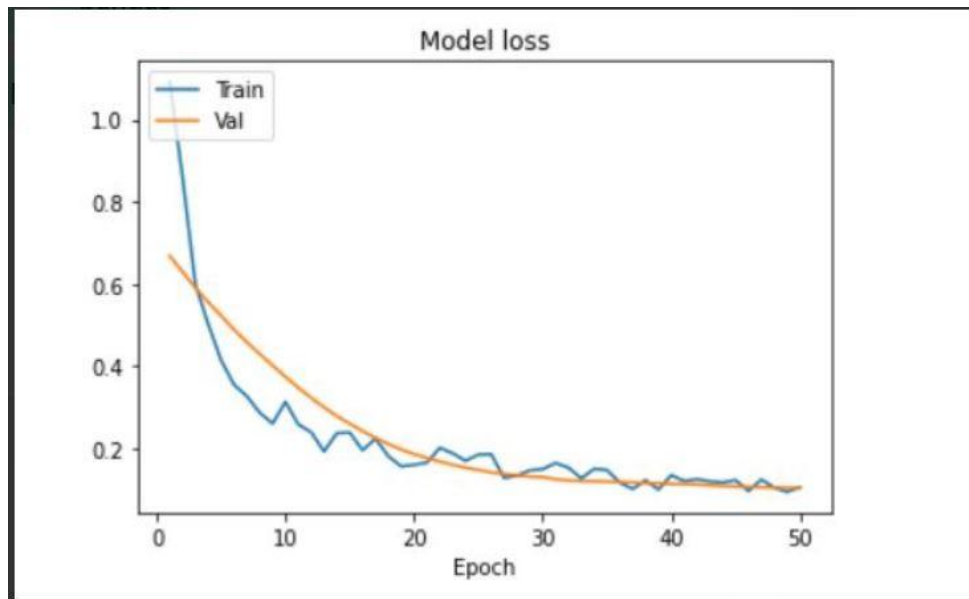

Figure 1

The above figure shows the accuracy of the model in SVM technique and the comparison of test value and trained value.

Figure 2

This figure 2 shows the loss of the model in SVM technique and compares the graph of trained value and tested value.

IV. CONCLUSION

Breast Cancer is a disease which should be detected in early stages in order to be cured. Despite the ongoing pandemic in the current situation many people are struggling for cure against breast cancer and many have already died in past few years. After applying different algorithms on the dataset, we have concluded that SVM gives a better precision and accuracy than any other method and can be used for medical service. This data will improve in upcoming years with the change in data size and other details.

REFERENCES

[1].   N. Kumar, G. Sharma and L. Bhargava, "The Machine Learning based Optimized Prediction Method for Breast Cancer Detection," 2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, India, 2020, pp. 1594-1598, doi: 10.1109/ICECA49313.2020.9297479.

[2].   D. Bazazeh and R. Shubair, "Comparative study of machine learning algorithms for breast cancer detection and diagnosis," 2016 5th International Conference on Electronic Devices, Systems and Applications (ICEDSA), Ras Al Khaimah, United Arab Emirates, 2016, pp. 1-4, doi: 10.1109/ICEDSA.2016.7818560.

[3].   M. Amrane, S. Oukid, I. Gagaoua and T. Ensarİ, "Breast cancer classification using machine learning," 2018 Electric Electronics, Computer Science, Biomedical Engineerings' Meeting (EBBT), Istanbul, Turkey, 2018, pp. 1-4, doi: 10.1109/EBBT.2018.8391453.

[4].   S. Ray, "A Quick Review of Machine Learning Algorithms," 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon), Faridabad, India, 2019, pp. 35-39, doi: 10.1109/COMITCon.2019.8862451.

[5].   M. Gupta and B. Gupta, "An Ensemble Model for Breast Cancer Prediction Using Sequential Least Squares Programming Method (SLSQP)," 2018 Eleventh International Conference on Contemporary Computing (IC3), Noida, India, 2018, pp. 1-3, doi: 10.1109/IC3.2018.8530572.

[6].   Teixeira, Fabiano; Montenegro, Joao Luis Zeni; da Costa, Cristiano Andre; da Rosa Righi, Rodrigo (2019). [IEEE 2019 XLV Latin American Computing Conference (CLEI) - Panama, Panama (2019.9.30-2019.10.4)] 2019 XLV Latin American Computing Conference (CLEI) - An Analysis of Machine Learning Classifiers in Breast Cancer Diagnosis. , (), 1–10. doi:10.1109/CLEI47609.2019.235094

[7].   Polat, Kemal; Senturk, Umit (2018). [IEEE 2018 2nd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT) - Ankara, Turkey (2018.10.19-2018.10.21)] 2018 2nd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT) - A Novel ML Approach to Prediction of Breast Cancer: Combining of mad normalization, KMC based feature weighting and AdaBoostM1 classifier. , (), 1–4. doi:10.1109/ismsit.2018.8567245

[8].   A. Bharat, N. Pooja and R. A. Reddy, "Using Machine Learning algorithms for breast cancer risk prediction and diagnosis," 2018 3rd International Conference on Circuits, Control, Communication and Computing (I4C), 2018, pp. 1-4, doi: 10.1109/CIMCA.2018.8739696.

[9].   https://www.kaggle.com/uciml/breast-cancer-wisconsin-data