

## A Review on Breast Cancer Detection

Anija J Anilkumar<sup>1</sup>, Dr. Deepesh Edwin.<sup>2</sup>

<sup>1</sup>PG Student, Dept. of Electronics & Communication Engineering, LBSITW, Kerala, India

<sup>2</sup>Assistant Professor, Dept. of Electronics & Communication Engineering, LBSITW, Kerala, India

\*\*\*

**Abstract** -Breast Cancer is a disease in which cells in the breast grow out of control. It is the most common type of cancer found in women around the world and it is among the leading causes of deaths in women. For the early detection process of cancer, machine learning techniques make a huge contribution. This paper presents a comparative analysis of machine learning and data mining techniques used to predict breast cancer. Many researchers have made their own efforts in the diagnosis and prediction of breast cancer, and each technique has a different accuracy rate, which varies in different situations. The main focus of this paper is to make a comparative analysis of existing machine learning and data mining techniques to find the most suitable method that supports with good accuracy of prediction. The main purpose of this review is to highlight some of the previous studies of machine learning used to predict breast cancer.

**Key Words:** Machine learning, Breast Cancer prediction, data mining.

### 1. INTRODUCTION

The second most common cancer diagnosed in women in the world is Breast Cancer. Breast cancer occurs when some breast cells begin to grow abnormally. Due to early detection, the survival rates have increased and the number of deaths associated with this disease is steadily declining. It is important to understand that most breast lumps are benign (non-cancerous) and malignant (cancerous). Benign tumors grow slowly and do not spread, where malignant tumors grow expeditiously, occupy and demolish nearby normal tissues and unfurl throughout the body. There are multiple tests for diagnosing breast cancer, including mammogram, ultrasound, MRI and biopsy. Mammogram is an X-ray of the breast. If an abnormality is detected on a screening mammogram, the doctor may recommend a diagnostic mammogram to further evaluate that abnormality.

Ultrasound is used after a mammogram. It is used to determine whether a new breast lump is a solid mass or a fluid-filled cyst. Biopsy is the only sure way to diagnose most cancers. It is the main symptomatic system that can decide whether the suspicious region is carcinogenic. There are different types of breast cancer. The most common types are invasive ductal carcinoma and invasive lobular carcinoma.

Data mining algorithms used in the healthcare industry plays an important role due to its advanced performance in prediction, diagnosis and making real time decisions to save people's lives. Classification and prediction are the most common data mining modeling goals, which uses several algorithms for the prediction of breast cancer. Data mining is an important part of machine learning, and it is used to find valuable patterns and trends hidden within vast volumes of data. Data mining and Machine Learning both employ advanced algorithms to uncover relevant data patterns. Machine learning process is based on three main strategies that consist of preprocessing, feature extraction and classification. This paper presents a machine learning strategy for breast cancer detection. Various study strategies are suggested for detection of breast cancer.

### 2. REVIEW ON DIFFERENT PAPERS

Jayanthi et al. breast cancer detection using KNN algorithm. The initial step of the KNN is to calculate the distance of the input test sample with all the dataset entries of training samples. KNN gives an accuracy of 96.49%.[1]

Sengar et al. Breast cancer prediction using 2 machine learning algorithms namely Decision Tree classifier and Logistic Regression. The best algorithm for prediction is the decision tree. It has pinpoint prediction accuracy. Decision Tree gives an accuracy of 95.10%. [2]

Lin et al. Breast cancer prediction using a hybrid algorithm of K-means and SOM neural network model. Compared with the K-means model and SOM neural network model, hybrid algorithm can accurately cluster the datasets. Hybrid algorithm model gives better accuracy and computing time.[3]

Kathale et al. proposed a method for detecting the cancer region and classifying cancerous or normal patients. Here, Random Forest algorithm is being used for detection and classification of breast cancer. It gives an accuracy of 95.3%. [4]

Bayrak et al. proposed two most popular machine learning techniques including SVM and ANN. The classification performance of SVM and ANN are done through performance metrics such as accuracy, precision,

recall and ROC area. SVM gives best performance with an accuracy of 96.9.[5]

Mahila et al. Breast cancer prediction using Naïve Bayes, KNN and J48 algorithm. For predicting cancer disease, 10-fold cross-validation has been used. KNN works more accurate than other two algorithms.[6]

Sinha et al. Breast cancer prediction using frequent item sets mining. Sinha has so far been compared with 4 classifiers, including Decision Tree, Naïve Bayes, Support Vector Machine (SVM) and K-Nearest Neighbors (KNN). Classifier with SVM gives the highest accuracy of 96.61%. [7]

Omondiagbe et al. proposes a hybrid approach for breast cancer diagnosis. Here, the dataset has been analyzed by using dimensionality reduction technique and 3 machine learning algorithms like SVM, ANN, and Naïve Bayes. Authors comparatively analyzed each technique with different types of feature selection such as coloration-based feature selection (CRS), linear discriminant analysis and Recursive feature elimination (RFE). SVM-LDA is been chosen because it has an accuracy of 98.8% and less computational time.[8]

**Table-1:** Comparison of the Review Paper

Year & Reference	Algorithm	Dataset	Description	Limitation	Accuracy
2021 [1]	KNN	Wisconsin Breast Cancer	Performance analysis of KNN using performance metric elements like accuracy and sensitivity.	The KNN algorithm takes more time in feature selection.	96.49%
2020 [2]	Decision Tree Logistic Regression	Wisconsin Breast Cancer	Comparing the accuracies of the algorithms namely Decision Tree and Logistic Regression for predicting breast cancer.	Decision trees are generally unstable and often relatively inaccurate.	95.10% 94.40%
2020 [3]	K-means SOM neural network Hybrid neural network	Wisconsin Breast Cancer	Comprehensive evaluation of K-means, SOM neural network and hybrid neural network for predicting the cancer.	Running time of the hybrid algorithm is higher than the K-means algorithm.	(precision) 85.7% 88.8% 92.1%

**Table-1.2:** (Continued) Comparison of the Review Paper

Year & Reference	Algorithm	Dataset	Description	Limitation	Accuracy
2020 [4]	Random Forest	Mammogram	Classification of Breast cancer using Random Forest Classifier and examination of image is done by using the features GLCM, Entropy and mean.	Features are not extracted properly from the database.	95.3%
2019 [5]	ANN SVM	Wisconsin Breast Cancer	Performance metrics such as accuracy, precision, recall and ROC area were used to evaluate SVM and ANN.	K fold cross validation is used to calculate the expected probabilities of occurrence and nonoccurrence.	95.4% 96.9%
2019 [6]	Naïve Bayes KNN J48	Collected from Doctor and Cancer Experts	Dataset was divided into two parts one is training data and another one is testing data, 10-fold cross validation was applied for the evaluation algorithm.	The testing phase is slow and takes a lot of time. Difficult to select K value.	98.2% 98.8% 98.5%

2019 [7]	Decision Tree Naïve Bayes SVM KNN	Wisconsin Breast Cancer	Comparative analysis of algorithms like KNN, SVM, decision tree and Naïve Bayes was performed.	For implementation R programming is used, this language consists of lots of packages, processing is lesser than other languages.	90.27 96.46 96.61 91.74
2019 [8]	SVM ANN Naïve Bayes SVM-LDA	Wisconsin Breast Cancer	Comparison of Machine Learning techniques done through dimensionality reduction techniques.	Due to the usage of CFS, LDA, PCA method, the evaluation phase took too much time.	96.47% 97.06% 91.18% 98.82%

### 3. CONCLUSION

The detection of Breast Cancer is a challenging problem because it is the most popular and injurious disease. When the initial symptoms of breast cancer are ignored, the patient can have serious consequences for their health and lead to death. This paper presents different machine learning techniques to detect breast cancer. Here different techniques like SVM, KNN, ANN etc. are being used to detect cancer. In future work, we can also compare the machine learning technique with deep learning to analyze which technique gives more accuracy and efficiency.

### REFERENCES

- [1] Jayanthi, N., and Gitanjali Wadhwa. "Classification of Breast Cancer Detection Using K-Nearest Neighbor Algorithm Trained with Wisconsin Dataset." *Annals of the Romanian Society for Cell Biology* (2021): 4440-4448
- [2] Sengar, Prateek P., Mihir J. Gaikwad, and Ashlesha S. Nagdive. "Comparative study of machine learning algorithms for breast cancer prediction." *2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT)*. IEEE, 2020.
- [3] Lin, Haoquan, and Zhenzhou Ji. "Breast cancer prediction based on K-Means and SOM Hybrid Algorithm." *Journal of Physics:m Conference Series*. Vol. 1624. No. 4. IOP Publishing, 2020.
- [4] Kathale, Poonam, and Snehal Thorat. "Breast cancer detection and classification." *2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE)*. IEEE, 2020.
- [5] Bayrak, Ebru Aydinag, Pinar Kirci, and Tolga Ensari. "Comparison of machine learning methods for breast cancer diagnosis." *2019 Scientific meeting on electrical-electronics & biomedical engineering and computer science (EBBT)*. IEEE, 2019.
- [6] Maliha, Shanjida Khan, et al. "Cancer disease prediction using Naive Bayes, K-nearest neighbor and J48 algorithm." *2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*. IEEE, 2019.
- [7] Sinha, Ankita, et al. "An Optimized Model for Breast Cancer Prediction Using Frequent Itemsets Mining." *International Journal of Information Engineering and Electronic Business (IJIEEB)* 11.5 (2019): 11-18.
- [8] Omondiagbe, David A., Shanmugam Veeramani, and Amandeep S. Sidhu. "Machine learning classification techniques for breast cancer diagnosis." *IOP Conference Series: Materials Science and Engineering*. Vol. 495. No. 1. IOP Publishing, 2019.

### BIOGRAPHIES



Anija J Anilkumar is a student. Currently pursuing M. Tech (Signal Processing) at LBS Institute of Technology for Women, Poojappura, APJ Abdul Kalam Technological University, Trivandrum, Kerala.



Dr. Deepesh Edwin, Ph.D. (Biomedical Instrumentation), Assistant Professor at LBS Institute of Technology for Women, Poojappura, APJ Abdul Kalam Technological University, Trivandrum, Kerala.