IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

MACHINE LEARNING-BASED BREAST CANCER PREDICTION: A COMPARATIVE REVIEW OF PUBLIC DATASETS, CLASSIFICATION TECHNIQUES AND THEIR ANALYSIS

Author 1:

V. Vinija,
Research Scholar,
Government Arts College,
Coimbatore, Tamil Nadu.
vinija2002@gmail.com

Author 2:

Dr. R. A. Roseline,
Associate Professor and Head,
Government Arts College,
Coimbatore, Tamil Nadu.
roselinera@yahoo.com

Abstract: A Tumor that develops in the breast tissue is Breast Cancer. One of the main reasons women are dying more frequently around the world is Breast Cancer. A sickness known as cancer manifests in the human body due to genetic changes. Cells become Malignant and Proliferate rapidly, harming healthy cells, due to several factors. Many cases of Breast Cancer that are not detected early might result in death. To increase the survival rate, early identification of Breast Cancer is crucial. These approaches can significantly aid Breast Cancer prediction and early diagnosis. However, each method has variable accuracy depending on the circumstances, instruments, and datasets used. In this review, we examine the potential role of Machine Learning in the prompt and precise identification of Breast Cancer. For this reason, the most popular datasets, including Wisconsin Prognostic Breast Cancer (WPBC), Wisconsin Original Breast Cancer (WOBC), and the SEER Breast Cancer dataset (SEEDBCD), were analysed. For academics and medical professionals seeking to develop dependable, comprehensible, and practically applicable Machine Learning models for Breast Cancer Classification, this review offers insights by summarising recent developments and highlighting current research gaps.

Keywords: Breast Cancer, Machine Learning, Prediction, Diagnosis, Accuracy, Machine Learning Classifiers.

1. INTRODUCTION

Breast Cancer continues to be one of the most common cancers diagnosed in women, a global public health challenge. According to the National Cancer Registry Programme (NCRP), India is expected to report more than 2,32,832 new breast Cancer cases among women in 2025. This accounts for approximately 14.8% of all cancer cases, making breast cancer the most prevalent malignancy among Indian women [1]. In the United States (US), projections for 2025 estimate around 3,16,950 new cases of Invasive Breast Cancer and 59,080 new cases of Ductal Carcinoma In Situ (DCIS), a Non-Invasive form of the disease [2]. Likewise, the European Union (EU) is projected to record nearly 3,60,000 new diagnoses of Breast Cancer in 2025. The annual estimates range from 4,70,000 to 5,50,000 new cases, highlighting Breast Cancer as the most prevalent cancer among women in Europe [3]. These statistics emphasise the increasing global burden of Breast Cancer and stress the significance of enhancing early diagnosis, accurate diagnosis and timely treatment to improve patient outcomes.

A significant amount of Breast Cancer may result from the abnormal growth of fatty and fibrous tissue in the breast. These different stages of the disease are caused by cancer cells spreading throughout the tumors and affecting more cells and tissues throughout the body. Breast Cancer can take many different forms [4]. Ductal Carcinoma In Situ (DCIS) is also known as Non-Invasive Cancer and occurs when cells proliferate outside the ducts [5]—unlike Invasive Ductal Carcinoma (IDC) [6]. Often referred to as Infiltrative Ductal Carcinoma [7], it is the second form. IDC is a kind of cancer that usually affects men and is caused by breast cells spreading throughout all breast tissues [8]. Invasive Mammary Breast Cancer (IMBC) [9] is another name for Mixed Tumor Breast Cancer (MTBC), the third kind of Breast Cancer. This Cancer is brought on by Lobular and abnormal duct cells [10]. Lobular Breast Cancer (LBC) is the fourth type [11], and it can raise the risk of other Invasive tumors such as Mucinous Ductal Cells (MDC), also called Colloid Breast Cancer (CBC). It happens when the tissue encircles the duct [13]. The final kind of Breast

Cancer that results in breast swelling and redness is called Inflammatory Breast Cancer (IBC). This fast-growing kind of Breast Cancer starts when cancer cells obstruct the lymphatic capillaries in the breast [14].

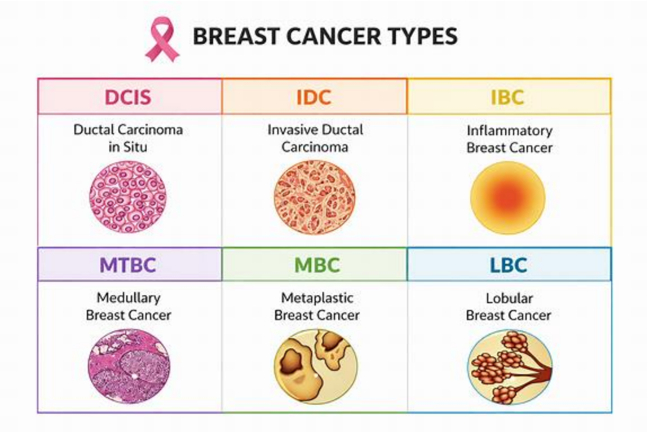

Figure 1: Types of Breast Cancer

This article's research on Breast Cancer is organised as follows: Related Work in Section II; Breast Cancer Classification in Section III; Datasets in Section IV; Review of the Literature in Section V; and, lastly, Conclusion in Section VI.

2. RELATED WORKS

Numerous studies on Breast Cancer published in the literature have used various Machine Learning approaches. To provide novice Machine Learning researchers with more relevant information, several studies have been compiled in a Comparative Review that highlights the application of Data Mining, Deep Learning, and Machine Learning methods for forecasting Breast Cancer recurrences. To predict Malignant and Benign Breast Cancer, Kumar et al. [15] evaluated 12 categorisation techniques using the UCI repository's Wisconsin Breast Cancer data. Several techniques were used in the study by Morles-Ortega et al. [16] to identify Breast Cancer recurrence in patients following surgery. To obtain notable improvements, the authors compared Decision Tree (DT), Naïve Bayes (NB), and Support Vector Machine (SVM) and then combined the best outcomes using the simple K-Means technique. To predict Breast Cancer and its recurrence, Aishwarya et al. examined K-Nearest Neighbours (KNN), SVM, NB, and Random Forest (RF) in [17]. SVM is the best at predicting Breast Cancer recurrences, while KNN is the best at predicting Breast Cancer. Temesgen Abera Asfew [18] examined the efficacy of K-Nearest Neighbours (KNN), NB, Logistic Regression (LR) and DT in detecting Breast Cancer. To greatly increase accuracy, they employed a weighting factor that is a function of feature quality and type during the optimisation process, based on the sensitivity and specificity of existing logistic functions. Lu et al. [19] used predictive models to predict recurrences after 10 years following breast surgery; Artificial Neural Network (ANN), K-NN, SVM, NB, and Cox Proportional hazards regression were employed. ANN achieved the best predictive performance in this study. Breast Cancer has been identified through a search for data visualisation and Machine Learning approaches, such as LR, KNN, SVM, NB, DT, and Rotational Forests. Machine Learning can influence cancer diagnosis in the decision-making process, as evidenced by the highest classification accuracy (98.1%) achieved with Logistic Regression. When Borges [20] applied Machine Learning techniques to the Wisconsin Breast Cancer dataset, he found that Bayesian networks achieved the highest accuracy (97.80%). Lastly, 3 Machine Learning models – Decision Tree (DT), Naïve Bayes (NB), and Sequential Minimal Optimisation (SMO)- were examined by Mohammeds et al. [21]. SMO had the best classifier performance on the Wisconsin Breast Cancer dataset, particularly after using an oversampling filter during preprocessing.

Similarly, utilising prognostic data that included time, Radvin and Clark [22] used a Neural network to predict a patient's chance of survival. The prediction made by the Neural Network was also contrasted with that of a Regression model using data from 1373 patients. Additionally, a linear diagnostic model was created by Wahlberg and others [23] to predict Malignant risk in Non-recurrent and Recurrent cases based on duration of illness. This model achieved 97.6% accuracy via cross-validation on a dataset of 569 patients. Quinlan [24] constructed the following model: Medical diagnosis and prediction with minimal addition C4.5 Decision Tree Description Length (DTDL) Penalty. This method yielded an accuracy of 94.74%. Furthermore, much information can be obtained from the literature; for instance, it was discovered that about 2,00,000 patients' records were utilised. They constructed models with multiple Neural and Linear networks, including the C4.5 Regression model. They say that Decision Tree techniques like C4.5 are superior to others for big datasets. Second, it achieves an accuracy of over 93.6 percent.

3. BREAST CANCER CLASSIFICATION

For each patient, identifying the specific cancer type is essential for therapy and outlook. Precisely identifying the type of cancer allows doctors to develop the most effective treatment strategy and forecast the patient's likelihood of survival. Numerous factors can be evaluated when identifying the kind of cancer, but the most prevalent ones are the type and stage [26].

3.1 BASED ON TYPE

Breast Cancer types: generally, there are two main forms of Breast Cancer: Invasive and Non-Invasive. Furthermore, treatment is influenced by subcategorisation based on the state of hormone receptors, proteins or genes in cancer cell strategy [27].

3.1.1 INVASIVE BREAST CANCER

When a cancer spreads and penetrates healthy breast tissue, it is referred to as an Invasive Cancer [27,28]. Aggressive Malignancies account for the bulk of Breast Cancer cases in women; among the several forms of Breast Cancer, Invasive Ductal Carcinoma (IDC) is the most frequently diagnosed variety, constituting nearly 80% of Breast Cancer diagnoses. In IDC, the Tumor may spread to the breast or other tissues after beginning in the milk ducts. Infiltrating Lobular Carcinoma (ILC) is the next most common form. Cancer spreads throughout the body after starting in the lobes or lobules. Infiltrating Lobular Carcinoma (ILC) typically shows a positive response to hormone treatment. There exist various other kinds of Breast Cancer. These are observed less frequently. Breast Cancer in men is extremely rare, and it is always Ductal Carcinoma. A rare but extremely aggressive subtype of

Breast Cancer, inflammatory Breast Cancer usually manifests as redness, oedema, and inflammation rather than a distinct lump. In addition, Breast Cancer may occur as metastatic Breast Cancer, which shows that the disease has spread to distant organs including the brain, liver, lungs or bones. Recurrent Breast Cancer is defined as cancer that reappears after initial treatment.

3.1.2 NON-INVASIVE BREAST CANCER

The term “Non-invasive Breast Cancer”, which is also known as “Carcinoma In Situ” or “pre-cancer”, denotes that the Malignant cells originate from their original place. Ductal Carcinoma In Situ (DCIS) is the first kind. It describes a situation where normal breast tissue is uninjured and in good health, but cancer cells are confined inside the milk ducts [28]. As the name implies, the second kind is Lobular Carcinoma In Situ (LCIS). The Lobule contains the Malignancy. Many people consider both of these cancer types to be pre-cancerous conditions that pose no significant threat, but as indicators that could progress into aggressive cancers.

3.2 BASED ON STAGE

Breast Cancer types are categorised into 5 levels, ranging from 0 to 4 [30]. The cancer stage provides insight into the disease's prognosis. There are numerous indicators and parameters to determine the exact cancer stage. In conclusion, stage 0 happens when the cancer is limited to the glands or milk ducts and has not spread to other areas. This indicates that the cancer is found in its early stages. Stage *I* indicates the presence of Invasive cancer, meaning Malignant cells are present within the breast tissue. For Stage *II*, the cancer might be increasing or extending to nearby lymph nodes. Stage *III* is characterised as a progressive phase. It is usually used for situations where cancer has not yet spread to any organs, but is found in the lymph nodes or the chest wall. Stage *IV* is considered the final stage and is used for scenarios in which tumors are present in organs distant from the breast, such as the brain, bones, liver, or lungs.

4. DATASETS

Breast Cancer datasets are crucial resources for developing and testing Machine Learning models. These data collections include various types of data, such as Clinical records, Genomic profiles, Histopathology images, Mammograms, Ultrasound images, Magnetic Resonance Imaging (MRI) and survival information. Publicly available datasets play a vital role in supporting reproducible research, enabling fair and consistent comparisons of computational methods, and driving advancements in Early disease detection, prognosis, treatment planning, and precision medicine.

4.1 The Wisconsin Original Breast Cancer (WOBC)

The Wisconsin Original Breast Cancer (WOBC) dataset [31] is among the first and most widely used benchmark datasets for Breast Cancer Classification. It contains Cytological characteristics collected from Fine-Needle Aspiration (FNA) specimens of breast lesions, which are typically used to categorise tumors as either Benign or Malignant.

4.2 The Wisconsin Diagnostic Breast Cancer (WDBC)

The Wisconsin Diagnostic Breast Cancer Dataset (WDBC) comprises digitised images of Fine-Needle Aspiration (FNA) specimens. It has been widely adopted to create and evaluate Machine Learning models for the Binary Classification of Breast Cancer [32].

4.3 The Wisconsin Prognosis Breast Cancer (WPBC)

The Wisconsin Breast Cancer Prognosis (WPBC) dataset [33] contains Clinical and Pathological features along with patient follow-up information, making it valuable for Prognosis-related studies to predict disease recurrence and evaluate patient outcomes following Breast Cancer treatment**.**

4.4 METABRIC Dataset

The Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) dataset [34] provides a

comprehensive clinical, genomic and transcriptomic information from patients with Breast Cancer. Owing to its rich Molecular and clinical data. It is extensively used for survival prediction, molecular subtype identification and biomarker discovery.

4.5 Breast Cancer SEER Dataset

The Surveillance, Epidemiology, and End Results (SEER) dataset is a large, population-based cancer registry containing Demographic, Clinical treatment and Survival information. It is widely used in Breast Cancer research for survival analysis, prognosis prediction, epidemiological investigation and the evaluation of long-term clinical outcomes. [35].

Overall, this review presents a comparative review of Machine Learning techniques used for Breast Cancer Prediction. Table 1 provides a Comparative analysis of Machine Learning algorithms across different studies.

5. REVIEW OF LITERATURE

A systematic review of the literature was conducted to identify studies on Machine Learning-based Breast Cancer prediction. Four widely used digital databases – IEEE Xplore, SpringerLink, ResearchGate and ScienceDirect (Elsevier) were searched for peer-reviewed journal articles published from 2020-2025. This search strategy included terms such as “Breast Cancer Prediction”, “Machine Learning”, “Breast cancer Classification” and Early Breast Cancer Detection”. The studies retrieved from these databases were initially screened by examining the titles, abstracts, and publication records; duplicates, review papers, book chapters, conference abstracts, or studies not relevant to Machine Learning-based Breast Cancer Prediction were removed. The remaining studies were evaluated in detail through full-text screening. Articles that met the established inclusion criteria were subsequently selected. The stages involved in the literature search and study selection are summarised in Figure 2. The continuous growth in publications can be attributed to several factors, including the increasing availability of publicly accessible Breast Cancer datasets, advancements in Machine Learning and Deep Learning algorithms, improved computational resources, and the growing need for accurate and early cancer detection. These developments have enabled researchers to design predictive models with higher accuracy and better clinical applicability.

The number of studies increased progressively during this period, indicating a growing interest in applying Machine Learning techniques to prediction. In each year, journal publications outnumbered conference papers, suggesting a stronger preference for detailed, peer-reviewed research. The highest publication count was recorded in 2025, with 39 papers, whereas 2020 had the lowest, with 16 papers. Overall, the upward trend highlights the continued. Advancements in Machine Learning techniques and their growing significance in aiding the early identification and forecasting of Breast

Table 1: Comparative Review of Machine Learning techniques used for Breast Cancer Prediction.

Author	Year	Research Focus	Technique / Algorithm Used	Key Findings	Accuracy	Limitations
[36]	2020	Prediction of Breast Cancer using diagnostic features from the WDBC dataset	SVM, Random Forest, Decision Tree, KNN	SVM outperformed other classifiers and demonstrated excellent classification capability for early Breast Cancer diagnosis.	98.24% (SVM)	Small dataset, limited generalisation, no external validation.
[37]	2020	Breast Cancer prediction using Clinical and biochemical attributes	Logistic Regression, SVM, Random Forest	Random Forest achieved the highest predictive performance among the evaluated algorithms.	96.55% (Random Forest)	Coimbra dataset is small (116 samples), increasing the risk of overfitting.
[38]	2021	Improving Breast Cancer prediction using Ensemble Learning methods	Random Forest, XGBoost, AdaBoost	XGBoost achieved the best classification performance and greater robustness than traditional models.	99.12% (XGBoost)	Evaluated only on WDBC dataset; external clinical validation was not performed.
[39]	2021	Clinical data-based Breast Cancer prediction	SVM, Artificial Neural Network (ANN), Decision Tree	ANN produced superior prediction accuracy using biochemical markers.	97.41% (ANN)	Limited sample size and lack of multi-centre clinical data.
[40]	2022	Hybrid Machine Learning model for Breast Cancer diagnosis using genomic and clinical data	Random Forest, XGBoost, LightGBM	LightGBM effectively handled high-dimensional genomic data and achieved the best performance.	95.80% (LightGBM)	Computational complexity is high and requires large-scale genomic datasets.
[41]	2022	Breast Cancer risk prediction using multiple clinical datasets	Neural Network, Logistic Regression, Random Forest	Machine Learning models improved risk prediction, with Neural Networks showing better discrimination.	AUC = 0.73	Moderate predictive performance and heterogeneous datasets affected model consistency.
[42]	2023	Breast Cancer prediction using publicly available clinical data	Logistic Regression, Random Forest, SVM, KNN	Logistic Regression achieved the highest classification accuracy despite its simplicity.	98.00% (Logistic Regression)	Dataset diversity was limited; explainability analysis was not included.
[43]	2023	Explainable Machine Learning for Breast Cancer survival prediction	XGBoost, CatBoost, Random Forest	XGBoost combined with explainable AI improved prediction accuracy and model interpretability.	96.40% (XGBoost)	Interpretability methods increase computational complexity and require expert analysis.
[44]	2025	Machine Learning framework for Breast Cancer treatment Classification	Gradient Boosting, XGBoost, AdaBoost	Gradient Boosting performed best for treatment classification but with lower accuracy than diagnostic studies.	77.18% (Gradient Boosting Machine)	Treatment prediction is more complex; model performance is affected by data imbalance and heterogeneous clinical records.

Author	Year	Research Focus	Technique / Algorithm Used	Key Findings	Accuracy	Limitations
[45]	2020	Breast Cancer prediction using clinical and biochemical features	SVM, Logistic Regression, Random Forest	Random Forest demonstrated superior classification performance using clinical attributes.	96.55% (Random Forest)	Small dataset with limited patient diversity and no external validation.
[46]	2021	Classification of Breast Cancer using diagnostic features	SVM, Decision Tree, KNN, Random Forest	SVM achieved the highest classification accuracy and outperformed other conventional machine learning models.	98.24% (SVM)	Performance evaluated only on the WDBC dataset; limited real-world clinical applicability.
[47]	2022	Breast Cancer prediction using clinical and genomic information	XGBoost, Random Forest, Logistic Regression	XGBoost effectively handled heterogeneous clinical and genomic features, providing the best predictive performance.	95.60% (XGBoost)	Requires extensive preprocessing and validation on larger multi-centre datasets.
[48]	2023	Comparative evaluation of multiple Machine Learning techniques for Breast Cancer prediction	Naïve Bayes, Logistic Regression, SVM, KNN, Decision Tree, Random Forest, AdaBoost, XGBoost	Decision Tree and XGBoost produced the highest predictive performance among the evaluated algorithms.	97.00% (Decision Tree & XGBoost)	The study relied on a single benchmark dataset and lacked external validation.
[49]	2023	Breast Cancer prediction using publicly available clinical data	Logistic Regression, SVM, Decision Tree, Random Forest	Logistic Regression achieved high predictive accuracy while maintaining model simplicity and interpretability.	98.00% (Logistic Regression)	Limited dataset diversity and absence of explainable AI techniques.
[50]	2023	Performance comparison of Machine Learning algorithms for Breast Cancer prediction	SVM, Random Forest, ANN, KNN, Decision Tree	SVM consistently outperformed other classifiers and demonstrated robust diagnostic performance.	98.24% (SVM)	The comparative study was restricted to the WDBC dataset and did not include external clinical validation.
[51]	2024	Breast Cancer Classification for Bangladeshi patients using explainable AI	Decision Tree, Random Forest, Logistic Regression, Naïve Bayes, XGBoost	XGBoost combined with explainable AI improved both prediction accuracy and model interpretability.	97.00% (XGBoost)	Dataset collected from a single hospital, limiting generalizability across populations.
[52]	2025	Breast Cancer Classification using Supervised Machine Learning algorithms	Logistic Regression, SVM, Decision Tree, Random Forest, LightGBM	Logistic Regression achieved the highest accuracy among the evaluated classifiers on the selected clinical dataset.	91.67% (Logistic Regression)	Moderate dataset size and limited feature diversity reduced overall predictive performance.

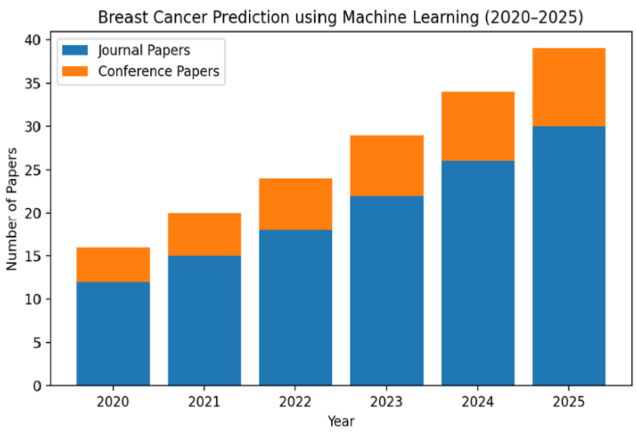

Figure 2: Analysis of papers from 2020 – 2025

Cancer. The upward trend in publications from 2020- 2025 highlights the growing research interest in applying Machine Learning to Breast Cancer prediction.

6. CONCLUSION

This article explores various Machine Learning methods and algorithms for predicting Breast Cancer. Our primary aim is to determine the best option. An algorithm capable of forecasting Breast Cancer cases more effectively. The main aim of this review is to highlight previous studies on Machine Learning algorithms. Utilised for predicting Breast Cancer, this article offers essential details for newcomers who aim to examine Machine Learning algorithms to build a foundation for curing cancer. Researchers can address the problem of limited dataset availability by applying specific data enhancement methods. Researchers must take into account the disparity between positive and negative data, as it may lead to bias toward favourable or unfavourable forecasts. Another significant problem that required attention was the uneven quantity of Breast Cancer images in comparison with impacted areas for accurate diagnosis and forecasting of breast Malignancy. A substantial region was examined in this study of early-stage breast cancer identification. Interest in breast cancer detection is evident in the rapid rise in the number of research papers published annually. Examining current studies will enable healthcare professionals to provide more precise diagnoses, resulting in better patient outcomes.

REFERENCES

[1] S. Gupta, A. Batra, S. Prinja, H. Malhotra, *et al.*, *"Task force report on breast cancer in India," Annals of the National Academy of Medical Sciences (India)*, vol. 61, no. 2, Apr.–Jun. 2025.
[2] American Cancer Society, *Cancer Facts & Figures 2025*. Atlanta, GA, USA: American Cancer Society, 2025.
[3] R. Cardoso, I. Ola, L. Jansen, N. Van Damme, et al., *"Breast cancer incidence by stage at diagnosis and mortality in 21 European countries in the era of mammography screening: An international population-based study,"* The Lancet Regional Health – Europe, 2025, Art. no. 101574, doi: 10.1016/j.lanepe.2025.101574.
[4] F. K. Ahmad and N. Yusoff, *"Classifying breast cancer types based on fine needle aspiration biopsy data using random forest classifier," in Proc. 13th Int. Conf. Intelligent Syst. Design Appl.*, Dec. 2013, pp. 121–125.
[5] R. Hou, M. A. Mazurowski, L. J. Grimm, J. R. Marks, L. M. King, C. C. Maley, E.-S.-S. Hwang, and J. Y. Lo, *"Prediction of upstaged ductal carcinoma in situ using forced labelling and domain adaptation," IEEE Trans. Biomed. Eng.*, vol. 67, no. 6, pp. 1565–1572, Jun. 2020.
[6] A. R. Chaudhury, R. Iyer, K. K. Iychettira, and A. Sreedevi, *"Diagnosis of invasive ductal carcinoma using image processing techniques,"* in Proc. Int. Conf. Image Inf. Process., Nov. 2011, pp. 1–6.
[7] S. Pervez and H. Khan, *"Infiltrating ductal carcinoma breast with central necrosis closely mimicking ductal carcinoma in situ (comedo type): A case series,"* J. Med. Case Rep., vol. 1, no. 1, p. 83, Dec. 2007.
[8] D. L. Page, W. D. Dupont, L. W. Rogers, and M. Landenberger, *"Intraductal carcinoma of the breast: Follow-up after biopsy only,"* Cancer, vol. 49, no. 4, pp. 751–758, 1982.
[9] A. B. Tuck, F. P. O'Malley, H. Singhal, and K. S. Tonkin, *"Osteopontin and p53 expression are associated with tumour progression in a case of synchronous, bilateral, invasive mammary carcinomas,"* Arch. Pathol. Lab. Med., vol. 121, no. 6, p. 578, 1997.
[10] B. Lee, K. Kim, J. Y. Choi, D. H. Suh, J. H. No, H.-Y. Lee, K.-Y. Eom, H. Kim, S. I. Hwang, H. J. Lee, and Y. B. Kim, *"Efficacy of the multidisciplinary tumour board conference in gynecologic oncology: A prospective study,"* Medicine, vol. 96, no. 48, p. e8089, Dec. 2017.
[11] S. Masciari, N. Larsson, J. Senz, N. Boyd, P. Kaurah, M. J. Kandel, L. N. Harris, H. C. Pinheiro, A. Troussard, P. Miron, N. Tung, C. Oliveira, L. Collins, S. Schnitt, J. E. Garber, and D. Huntsman, *"Germline E-cadherin mutations in familial lobular breast cancer,"* J. Med. Genet., vol. 44, no. 11, pp. 726–731, Aug. 2007.
[12] A. Memis, N. Ozdemir, M. Parildar, E. E. Ustun, and Y. Erhan, *"Mucinous (colloid) breast cancer: Mammographic and US features with histologic correlation,"* Eur. J. Radiol., vol. 35, no. 1, pp. 39–43, Jul. 2000.
[13] A. Gradilone, G. Naso, C. Raimondi, E. Cortesi, O. Gandini, B. Vincenzi, R. Saltarelli, E. Chiapparino, F. Spremberg, M. Cristofanilli, L. Frati, A. M. Aglianò, and P. Gazzaniga, *"Circulating tumour cells (CTCs) in metastatic breast cancer (MBC): Prognosis, drug resistance and phenotypic characterisation,"* Ann. Oncol., vol. 22, no. 1, pp. 86–92, Jan. 2011.
[14] F. M. Robertson, M. Bondy, W. Yang, H. Yamauchi, S. Wiggins, S. Kamrudin, S. Krishnamurthy, H. Le-Petross, L. Bidaut, A. N. Player, S. H. Barsky, W. A. Woodward, T. Buchholz, A. Lucci, N. Ueno, and M. Cristofanilli, *"Inflammatory breast cancer: The disease, the biology, the treatment,"* CA, Cancer J. Clin., vol. 60, no. 6, pp. 351–375, 2010.
[15] R. C. O. Roberto César, L. B. Germán, P. P. A. C. Paola Patricia, E. A. R. Eugenia, E. C. O. M. Elisa Clementina, J. C. O. José, M. A. P. M. Marlon Alberto, F. E. M. P. Fabio

Enrique, and M. R. V. Margarita, *"Method based on data mining techniques for breast cancer recurrence analysis,"* in Advances in Swarm Intelligence: 11th Int. Conf., ICSI 2020, Belgrade, Serbia, Jul. 14–20, 2020, vol. 12145, 584–596. https://doi.org/10.1007/978-3-030-53956-6_54
[16] Aishwarja A.I., Eva N.J., Mushtary S., Tasnim Z., Khan N.I., Islam M.N. (2021) *"Exploring the Machine Learning Algorithms to Find the Best Features for Predicting the Breast Cancer and Its Recurrence"*. In: Vasant P., Zelinka I., Weber GW. (eds) Intelligent Computing and Optimisation. ICO 2020. Advances in Intelligent Systems and Computing, vol 1324. Springer, Cham. https://doi.org/10.1007/978-3-030-68154-8_48
[17] Temesgen Abera Asfaw. (2019) *"Comparative Analysis Of Classification Approaches for Breast Cancer"*

International Journal of Computer Engineering and Technology (IJCET) - Scope Database Indexed.Volume:10,Issue:4,Pages:10-16. http://www.iaeme.com/IJCET/issues.asp?JType=IJCET&VType=10&IType=4

[18] Laila Khairunnahar, Mohammad Abdul Hasib, Razib Hasan Bin Rezanur, Mohammad Rakibul Islam, Md Kamal Hosain. (2019) *"Classification of malignant and benign tissue with Logistic Regression"* Informatics in Medicine Unlocked, Volume 16, 100189, ISSN 2352-9148. https://doi.org/10.1016/j.imu.2019.100189.

[19] Lou, S. J., Hou, M. F., Chang, H. T., Chiu, C. C., Lee, H. H., Yeh, S. J., & Shi, H. Y. (2020). *"Machine Learning Algorithms to Predict Recurrence within 10 Years after Breast Cancer Surgery: A Prospective Cohort Study. Cancers"*12(12),3817. https://doi.org/10.3390/cancers12123817

[20] Ak MF. (2020) *"A Comparative Analysis of Breast Cancer Detection and Diagnosis Using Data Visualisation and Machine Learning Applications"* Healthcare (Basel). 8(2):111. https://doi.org/10.3390/healthcare8020111.

[21] Borges, L. (2015). *"Analysis of the Wisconsin Breast Cancer Dataset and Machine Learning for Breast Cancer Detection",* Workshop de Visão Computational.

[22] Mohammed, S. A., Darrab, S., Noaman, S. A., & Saake, G. (2020). *"Analysis of Breast Cancer Detection Using Different Machine Learning Techniques. Data Mining and Big Data":* 5th International Conference, DMBD 2020, Belgrade, Serbia, July 14–20, 2020, Proceedings, 1234, 108–117. https://doi.org/10.1007/978-981-15-7205-0_10.

[23] P. M. Ravdin and G. M. Clark, *"A practical application of neural network analysis for predicting outcome of individual breast cancer patients,"* Breast Cancer Research and Treatment, vol. 22, no. 3, pp. 285–293, 1992.

[24] U. Zahid, I. Ashraf, M. A. Khan et al., *"BrainNet: optimal deep learning feature fusion for brain tumor classification,"* Computational Intelligence and Neuroscience, vol. 2022, pp. 1–13, 2022.

[25] J. R. Quinlan, *"Improved use of continuous attributes in C4. 5,"* Journal of Artificial Intelligence Research, vol. 4, pp. 77– 90, 1996.

[26] D. Delen, G. Walker, and A. Kadam, *"Predicting breast cancer survivability: a comparison of three data mining methods,"* Artificial Intelligence in Medicine, vol. 34, no. 2, pp. 113–127, 2005.

[26] Machine Learning, *https://www.ibm.com/cloud/learn/machine-learning, 2023.*

[27] G. Chugh, S. Kumar, N. Singh, *"Survey on machine learning and deep learning applications in breast cancer diagnosis"*, Cogn. Comput. (2021) 1–20.

[28] M. Akram, M. Iqbal, M. Daniyal, A.U. Khan, *"Awareness and current knowledge of breast cancer"*, Biol. Res. 50 (2017) 1–23

[29] A. Prat, E. Pineda, B. Adamo, P. Galván, A. Fernández, L. Gaba, et al., "Clinical implications of the intrinsic molecular subtypes of breast cancer", Breast 24 (2015) S26–S35

[30] K.L. Maughan, M.A. Lutterbie, P.S. Ham, *"Treatment of breast cancer"*, Am. Fam. Phys. 81 (11) (2010) 1339–1346

[31]https://archive.ics.uci.edu/dataset/15/breast+cancer+wisconsin+original

[32]https://archive.ics.uci.edu/dataset/17/breast+cancer+wisconsin+diagnostic

[33]https://archive.ics.uci.edu/dataset/16/breast+cancer+wisconsin+prognostic

[34]https://www.kaggle.com/datasets/raghadalharbi/breast-cancer-gene-expression-profiles-metabric

[35]https://ieee-dataport.org/open-access/seer-breast-cancer-data

[36] A. A. Nahid, M. S. Hossain, and M. A. Rahman, *"Breast Cancer Prediction Using Machine Learning Techniques," SAGE Open Medicine*, vol. 8, 2020.

[**37**] D. R. Dua, M. K. Sharma, and P. K. Singh, *"Machine Learning Approach for Breast Cancer Prediction," SAGE Open Medicine*, vol. 8, 2020.

[38] Y. Zhang, H. Wang, and X. Li, *"Breast Cancer Prediction Using Ensemble Machine Learning Methods,"* BMC Medical Informatics and Decision Making, vol. 21, no. 1, 2021.

[39]M. Alshamlan, G. Badr, and Y. Alohali, *"Machine Learning-Based Breast Cancer Prediction Using Clinical Data*," BMC Medical Informatics and Decision Making, vol. 21, 2021

[40] H. Huang, J. Wang, and X. Zhang, *"Breast Cancer Prediction Using Hybrid Machine Learning Models,"* JMIR Medical Informatics, vol. 10, no. 8, 2022.

[41] A. Yala, C. Lehman, T. Schuster, T. Portnoi, and R. Barzilay, *"Machine Learning-Based Breast Cancer Risk Prediction Models: A Systematic Review,"* JMIR Public Health and Surveillance, vol. 8, 2022.

[42] A. Sharma and R. Kumar, *"Prediction of Breast Cancer with 98% Accuracy Using Machine Learning Algorithms,"* preprint arXiv: 2307, 2023

[45] Y. Liu, J. Chen, and X. Zhao, "*Breast Cancer Prediction Using Explainable Machine Learning,"* Scientific Reports, vol. 13, 2023

[43] M. Alkhateeb, A. Alzubi, and M. Alsmadi, *"A Machine Learning Framework for Breast Cancer Treatment Classification Using a Novel Dataset,"* preprint arXiv:2503, 2025.

[44] M. A. A. Almansour et al., *"Machine Learning-Based Breast Cancer Prediction Using Clinical Features,"* Diagnostics, vol. 10, 2020.

[45]"*Breast Cancer Prediction Using Machine Learning Classification Algorithms,"* Diagnostics, 2021.

[46] *"Machine Learning Models for Breast Cancer Prediction Using Clinical Data,"* arXiv preprint, 2022. *"Machine Learning Techniques for Breast Cancer Prediction,"*

[47] *"Expert Systems with Applications"* (ScienceDirect), 2023.

[48] *"Prediction of Breast Cancer with 98% Accuracy,"* arXiv preprint, 2023.

[49] *"Accuracy Assessment of Machine Learning Algorithms Used to Predict Breast Cancer,"* Diagnostics, 2023.

[50] Predictive Modelling for Breast Cancer Classification in the Context of Bangladeshi Patients: *"A Supervised Machine Learning Approach with Explainable AI,"* arXiv preprint, 2024.

[51] *"Breast Cancer Prediction Using Machine Learning Classification Algorithms"* Journal of Biomedical Informatics / ScienceDirect, 2025.

[52] N.I. Yassin, S. Omran, E.M. El Houby, H. Allam, *"Machine learning techniques for breast cancer computer-aided diagnosis using different image modalities: a systematic review"*, Computer Methods Programs Biomed. 156 (2018) 25–45.

V. Vinija is a Research Scholar, PG and Research Department of Computer Applications. She did her MCA at Bharathiar University. She is pursuing her PhD in Machine Learning, with a focus on Data Mining and Warehousing.

Dr R. A. Roseline is Associate Professor and Head of the PG and Research Department of Computer Applications. She has 32 years of experience in Higher education. She completed her PhD in Computer Science with a focus on Wireless Sensor Networks at Bharathiar University. She has published about 30 papers in International Journals. She has completed a UGC Minor Project in Pollution Monitoring using Sensor Networks with funding of 1-7 Lacs. She has presented papers at International Conferences in Australia and Malaysia.