Cancer demands an enormous tax on society. It is a long-term morbidity disease. It is considered the second most common reason of human death. According to the World Health Organization (WHO), about 9.6 million people died due to the cancer in 2018 and 70% of those happened in developing countries where cancer diagnosis facilities are still very expensive [1]. Breast cancer is the most common type of cancer in the world [2]. Breast cancer is a disease that is produced due to the irregular growth of breast cells. Generally, to generate the new cell, somatic cells divided in a regular manner. A mass of tissue is considered cancerous when a specific part of the body grows and divides out of control. There are two types of cancer such as benign and malignant. Benign represents the non-cancerous which has no threat to life but malignant represents the most cancerous and it has direct threat to life [5]. In the world, such as the US, in 2022, 30% of cancer is breast cancer among all cancer that is diagnosed in women's bodies [22–24]. In addition, in 2018 in Saudi Arabia, breast cancer was considered the most common type of cancer, with a prevalence rate of 14.8% and a mortality rate of 8.5% [25]. And in South Asia (India, Bangladesh, Nepal, Myanmar, Pakistan, and Tibet) [26] approximately 76,000 women die due to breast cancer. It is estimated that global breast cancer cases will grow from 1.4 million in 2008 to over 2.1 million cases in 2030 [27]. Every year, almost 1.5 million women are diagnosed with breast cancer [3]. Approximately 29.9% of deaths from cancer in women are owing to breast cancer [4]. Hence, it is important to diagnose the breast cancer. However, it requires dozens of medical equipment and staffs to diagnosis a breast cancer patient. Breast cancer can be diagnosed using a variety of procedures including physical syndromes, biopsy and radiographic images [6]. The biopsy method is used to ensure the presence of breast cancer. Mammography is the standard diagnostic method for breast cancer and surgical biopsy [7]. Radiology is the medical way that diagnoses and treats diseases using clinical images. However, the effectiveness of this process depends on radiologists' explanation [8] and radiologists may miss up to 30% of breast cancer based on the density of breasts [9].

To overcome the issues associated with a breast cancer diagnosis, computer scientists have contributed with several automated methods. MF Aslan et al. [12] proposed four different Machine Learning (ML) algorithms to detect breast cancer, such as Artificial Neural Network (ANN), standard Extreme Learning Machine (ELM), Support Vector Machine (SVM), and K-Nearest Neighbor (KNN). Another study, M. Hussain et al. [13] compared different SVM kernels for the

detection of breast cancer and their system achieved around 96% accuracy. On the other hand, Bayrak et al. [19] compare the machine learning model performance, applied two ML (SVM, ANN) models to the Wisconsin Breast Cancer (Original) dataset. For performance measures, they consider accuracy, precision, recall and ROC Area. And Agarap et al. [16] proposed a comparison of six machine learning (ML) algorithms: GRU-SVM [4], Linear Regression, Multilayer Perceptron (MLP), Nearest Neighbor (NN) search, SoftMax Regression, and Support Vector Machine (SVM) on the Wisconsin Diagnosis Breast Cancer (WDBC) datasets. Among them, the MLP algorithm achieved the highest accuracy (99.04%). Moreover, Potdar et al. [20] claim that Artificial Neural Networks (ANN) is better for breast cancer classification than K-NN and Bayesian Classifiers that provides 97.4% accuracy. And Gayathri et al [21] represent another comparison study where Relevance Vector Machine (RVM) provides a low computational cost even though the variables are reduced compared with other machine learning algorithms that are used for breast cancer detection.

In the above research, most of those showed their analysis on a single dataset and used some specific ML algorithms that do not exhibit an overall performance on breast cancer. Because there are many breasts cancer datasets, some of which contain images such as Digital Mammography, Digital Breast Tomosynthesis (3-D Mammography), and Breast Ultrasound and some dataset contain CSV files, analysis of specific datasets may not be accurate to provide information on overall performance. Not only that, different machine learning algorithms perform differently on different datasets. As a result, need a comparative study to compare the performance of different machine learning algorithms on different breast cancer datasets. This comparative study will help medical specialist and future researchers in this field. However, none of those approaches did extensive analysis of ML algorithms.

In this study, we have used Twelve reputed ML algorithms in two different datasets. These Algorithms are both supervised and unsupervised and also Begging and Bosting classifier. The Twelve reputed ML algorithms namely Naive Bayes (NB), Logistic Regression (LR), Decision Tree Classifier (DT), Support Vector Machine (SVM), Linear Discriminant Analysis (LDA), Voting Classifier (VC), KNeighborsClassifier (K-NN), AdaBoost Classifier (AD), Random Forest Classifier (RF), Stochastic Gradient Descent (SGD), Bagging Classifier (BC), Gradient Boosting

Classifier (GB). Hereby, widely verified these two datasets and did numerous research on them. For that, we choose these two datasets. The two datasets namely Wisconsin Breast Cancer (Original) (WBC) and Wisconsin Breast Cancer Diagnosis (WBCD). Finally, we compared the performance of each ML algorithm on two datasets and found out which algorithms provided the best performance.

The rest of the paper is arranged as follows: Section II explains the literature review, section III describes the methodology of the comparative study, section IV shows the comparative results and discussions, and finally, section V draws the conclusion.

## Reference


[1] https://www.who.int/en/news-room/fact-sheets/detail/cancer. Last Access: 06.02.2022.

[2] Sivakami, K., and Nadar Saraswathi.” Mining big data: breast cancer prediction using DT-SVM hybrid model.” International Journal of Scientific Engineering and Applied Science (IJSEAS) 1, no. 5 (2015): 418-429.

[3] Mojrian, Sanaz, Gergo Pinter, Javad Hassannataj Joloudari, Imre Felde, Akos Szabo-Gali, Laszlo Nadai, and Amir Mosavi.” Hybrid machine learning model of extreme learning machine radial basis function for breast cancer detection and diagnosis; a multilayer fuzzy expert system.” In 2020 RIVF International Conference on Computing and Communication Technologies (RIVF), pp. 1-7. IEEE, 2020.

[4] H. You and G. Rumbe,” Comparative study of classification techniques on breast cancer FNA biopsy data,” International Journal of Artificial Intelligence and Interactive Multimedia, vol. 1, no. 3, pp. 6-13, 2010.

[5] Al Bataineh, Ali.” A comparative analysis of nonlinear machine learning algorithms for breast cancer detection.” International Journal of Machine Learning and Computing 9, no. 3 (2019): 248-254.

[6] A. A. Ardakani, A. Gharbali, and A. Mohammadi, “Classification of breast tumors using sonographic texture analysis,” J. Ultrasound Med., vol. 34, no. 2, pp. 225–231, 2015.

[7] Al Bataineh, Ali.” A comparative analysis of nonlinear machine learning algorithms for breast cancer detection.” International Journal of Machine Learning and Computing 9, no. 3 (2019): 248-254.

[8] B. L. Sprague, E. F. Conant, T. Onega, M. P. Garcia, E. F. Beaber, S. D. Herschorn, C. D. Lehman, A. N. A. Tosteson, R. Lacson, M. D. Schnall, D. Kontos, J. S. Haas, D. L. Weaver, and W. E. Barlow, “Variation in Mammographic Breast Density Assessments among Radiologists in Clinical Practice: A Multicenter Observational Study,” Ann. Intern. Med., vol. 165, no. 7, pp. 457–464, 2016.

[9] T. M. Kolb, J. Lichy, and J. H. Newhouse, “Comparison of the Performance of Screening Mammography, Physical Examination, and Breast US and Evaluation of Factors that Influence Them: An Analysis of 27,825 Patient Evaluations,” Radiology, vol. 225, no. 1, pp. 165–175, 2002.

[10] Islam, Md Milon, Hasib Iqbal, Md Rezwanul Haque, and Md Kamrul Hasan. ”Prediction of breast cancer using support vector machine and K-Nearest neighbors.” In 2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC), pp. 226-229. IEEE, 2017.

[11] Sivakami,” Mining Big Data: Breast Cancer Prediction using DT-SVM Hybrid Model”, 2015

[12] Aslan, Muhammet Fatih, Yunus Celik, Kadir Sabancı, and Akif Durdu. ”Breast cancer diagnosis by di↵erent machine learning methods using blood analysis data.” (2018).

[13] M. Hussain, S. K. Wajid, A. Elzaart, and M. Berbar, “A comparison of SVM kernel functions for breast cancer detection.” IEEE Eighth International Conference Computer Graphics, Imaging and Visualization, pp. 145-150, 2011.

[14] Al Bataineh, Ali.” A comparative analysis of nonlinear machine learning algorithms for breast cancer detection.” International Journal of Machine Learning and Computing 9, no. 3 (2019): 248-254.

[15] Mohammed, Siham A., Sadeq Darrab, Salah A. Noaman, and Gunter Saake. ”Analysis of breast cancer detection using di↵erent machine learning techniques.” In International Conference on Data Mining and Big Data, pp. 108-117. Springer, Singapore, 2020.

[16] Agarap, Abien Fred M.” On breast cancer detection: an application of machine learning algorithms on the wisconsin diagnostic dataset.” In Proceedings of the 2nd international conference on machine learning and soft computing, pp. 5-9. 2018.

[17] Sharma, Shubham, Archit Aggarwal, and Tanupriya Choudhury. ”Breast cancer detection using machine learning algorithms.” In 2018 International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS), pp. 114-118. IEEE, 2018.
[18] Assegie, Tsehay Admassu.” An optimized K-Nearest Neighbor based breast cancer detection.” Journal of Robotics and Control (JRC) 2, no. 3 (2021): 115-118.
[19] Bayrak, Ebru Aydındag, Pınar Kırcı, and Tolga Ensari.” Comparison of machine learning methods for breast cancer diagnosis.” In 2019 Scientific meeting on electrical-electronics biomedical engineering and computer science (EBBT), pp. 1-3. Ieee, 2019.
[20] Potdar, Kedar, and Rishab Kinnerkar.” A comparative study of machine learning algorithms applied to predictive breast cancer data.” International Journal of Science and Research 5, no. 9 (2016): 1550-1553.
[21] Gayathri, B. M., and C. P. Sumathi.” Comparative study of relevance vector machine with various machine learning techniques used for detecting breast cancer.” In 2016 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), pp. 1-5. IEEE, 2016
[22] Siegel, R.L.; Miller, K.D.; Fuchs, H.E.; Jemal, A. Cancer statistics. CA Cancer J. Clin. 2022. [CrossRef]

[23] Byrne, D.; Ohalloran, M.; Jones, E.; Glavin, M. A comparison of data-independent microwave beamforming algorithms for the early detection of breast cancer. In Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Minneapolis, MN, USA, 3–6 September 2009. [CrossRef]

[24] Samsuzzaman, M.; Islam, M.T.; Shovon, A.; Faruque, R.I.; Misran, N. A 16-modified antipodal Vivaldi Antenna Array for microwave-based breast tumor imaging applications. Microw. Opt. Technol. Lett. 2019, 61, 2110–2118. [CrossRef].

[25] Alqahtani, W.S.; Almufareh, N.A.; Domiaty, D.M.; Albasher, G.; Alduwish, M.A.; Alkhalaf, H.; Almuzzaini, B.; Al-Marshidy, S.S.; Alfraihi, R.; Elasbali, A.M.; et al. Epidemiology of cancer in Saudi Arabia thru 2010–2019: A systematic review with constrained meta-analysis. AIMS Public Health 2020, 7, 679. [PubMed]

[26] M. Brown, S. Goldie, G. Draisma, J. Harford, and J. Lipscomb, “Health service interventions for cancer control in developing countries,” in Disease Control Priorities in Developing Countries, pp. 569–590, Oxford University Press, New York, NY, USA, 2nd edition, 2006.

[27] World Cancer Research Fund, “Breast cancer worldwide,” http://www.wcrf.org/cancer facts/women-breast-cancer.php/.