

¹ Dawood Salim
Mohammed

² Firas Saaduldeen
Ahmed

³ Havall Muhssin
Mohammad

⁴ Zozan Saadallah
Hussain

Enhancing Breast Cancer Detection with Ensemble Machine Learning Models



Abstract: - Breast cancer remains a significant public health issue worldwide, underlining the need for accurate and efficient diagnostic methods. In this paper, we propose a new technique to enhance breast cancer diagnosis through the integration of multiple machine-learning models. Our strategy employs a combination of the Naive Bayes classifier, Stochastic Gradient Descent (SGD), Bagging, and the ZeroR classifier, alongside Bayes Network learning. The cornerstone of our approach is Bayes Network learning, a probabilistic graphical model designed to map out the intricate interconnections among various diagnostic factors. This is significant in the way that it can help to uncover complex relationships in the data for the sake of leading to more accurate predictions. Added to the above, we use the Naïve Bayes classifier, a classifier showing good validity in classification tasks and based on probabilistic reasoning, for the screening of breast cancer. Further, a refined model's parameter is included using the SGD and leads to enhancement of the generalization and overall performance of the model. In addition, as part of controlling overfitting, one can also use Bagging. It is an ensemble method in the sense that it considers several models. ZeroR classifier is a very basic classifier and is just used to compare its performance with our composite approach. We are comparing complex ensemble results to its simplicity. We will validate the ability of our proposed methodology to compare the performance of our integrated models against ZeroR.

Keywords: Data mining, Machine Learning, Breast Cancer, Classification algorithms.

I. INTRODUCTION

Breast cancer is the health issue of the day, and hereby accentuates to the fact that better ways need to be found in helping patients achieve better health. The paper introduces an innovative approach to breast cancer diagnosis using a suite of machine-learning models aimed at achieving higher detection accuracy [1][2]. For this purpose, our approach marries the Naïve Bayes classifier, though with bagging, Stochastic Gradient Descent (SGD) with the ZeroR classifier, and perhaps most importantly, Bayes network learning, into one holistic diagnostic tool [3]. Our approach uses Bayes Network Learning to be able to map, in an effective way, the complex interactions that different diagnostic indicators hold in the Breast Cancer Dataset, revealing subtle but important correlations [4][5][6]. Further, to improve the method, this classifier has been taken. The Naive Bayes classifier brought the advantage of probabilistic reasoning in very successful ways in classification tasks and has proven to be a very successful classifier. More importantly, we applied Stochastic Gradient Descent (SGD) to make the system more stable and adaptive; with such a system, better performance is optimized, and generalization is guaranteed across several datasets. On the other hand, techniques such as Bagging, an advanced ensemble method, help in the amalgamation of different model predictions so as to reduce overfitting and boost predictive accuracy. Example: Mammography, Breast Imaging—a traditional approach, M [9]. It is a diagnostic practice checking the health of female breast tissue by use of X-rays, and it helps a lot in the early detection of the condition, which can be elusive through an external examination. The technique is further revolutionized through the incorporation of dynamic magnetic resonance imaging (MRI). The technique is so important not only in providing detailed images of the breast abnormalities but also in further understanding the dynamics of growing tumor and angiogenesis rates, which contribute by all means to the earliest diagnosis of metastatic contrasts reduction for the patient. Their use of the multi-model approach, with the integration of state-of-the-art machine learning models and the cutting-edge

² *Corresponding author: firas_saad@ntu.edu.iq

^{1,2,4} Department of Electrical Technology, Technical Institute, Northern Technical University, Mosul, Iraq, davemohamed@ntu.edu.iq, zozan.technic@ntu.edu.iq.

³Department of Computer Information System, Mergasor Technical Institute, Erbil Polytechnic University, Erbil, Iraq, havallmohssin@gmail.com

imaging technologies, only heralds a big leap for the detection of breast cancer. It could be in new detection methods or new patient care methods. In one of the recently invented ways, ultrasound [11], sound waves are used in the penetration of the body in search of any form of abnormalities. The transducer sends sound waves that are picked up by the reflections of tissue in the breast [12]. It is an important non-invasive procedure that yields useful diagnostic data. The second tool, the electrography [13], whose researchers have developed only recently, is applicable in cases when breast cancer tissue exceeds that one surrounding it in size. To discriminate between benign and malignant tissue types, compression color mapping is used. The technique produces echoes that are transformed into grayscale or numerical representations for computer analysis. Using F-fluorodeoxyglucose tracers in Positron Emission Tomography (PET) [14] pictures, doctors can gain a thorough understanding of the function of the tumor within the human body. This method uses radiation tracers to examine the metabolic activity of the tumor and determine its effect on the patient's overall health.

Machine learning [13, 15, 16, 17], bio-inspired computation approaches [17, 18, 19, 20, 21, 22, 23], deep learning [24], and other diverse techniques have recently found use in the field of medical prediction. Despite the validation of a variety of strategies, none have consistently produced exact and trustworthy results. The pursuit of accuracy faces difficulties, particularly in the field of mammography where medical practitioners must evaluate vast amounts of picture data. This method not only takes time, but it can occasionally result in inaccurate disease diagnosis. To overcome this problem, this article compares various machine-learning methods for data-driven disease diagnosis. Six supervised machine learning techniques were used to successfully identify medical problems. This article is organized as follows: The introduction is in Section 1, and a thorough literature analysis is in Section 2. The preprocessing steps required for machine learning are covered in Section 3. The software and simulation approaches, as well as thorough justifications of the outcomes, are presented in section four. The final section, Section 5, summarizes the findings of this study.

II. RELATED WORK

In [25], the implementation and comparison of various machine learning classification techniques like Decision Trees, Bayesian methods, and Neural Networks in predicting breast cancer outcomes. The study would detail the dataset used, data preprocessing steps, model training, validation processes, and a comparative analysis based on sensitivity and accuracy metrics, concluding with the superior performance of Neural Networks in this application. A machine has been created in [26] that introduces a diagnostic tool combining Random Forest classifiers with feature selection techniques to distinguish between malignant and benign breast tumors. It would detail the feature selection process, Random Forest model specifications, training and testing phases, and evaluation metrics, highlighting the model's diagnostic performance and prognostic value, with emphasis on the reduction in feature dimensionality and its impact on accuracy.

In [27], explores the application of SVM, ANN, and Decision Trees in predicting breast cancer recurrence, utilizing a dataset from the Iranian Center. The paper will discuss each algorithm's methodology, performance evaluation, and a comparison based on accuracy, sensitivity, and specificity, concluding with findings on the most effective model for recurrence prediction [28]. investigates the efficacy of an ensemble approach combining Support Vector Machines and Decision Trees, focusing on breast cancer identification with a reduced feature set. The study would describe the ensemble methodology, feature reduction process, model validation, and results, emphasizing improvements in accuracy and model interpretability.

In [29], the role of machine learning in cancer prognosis and prediction, analyzing various models' applicability. The following paper intends to discuss, in detail, the role of machine learning in oncology. It provides an analysis of several statistical models and an extensive review of relevant literature focusing on the accuracy and promise of machine learning in appropriately forecasting cancer outcomes. It is further elaborating the highlighted comparative studies described in [30], describing the applied algorithms under testing, applied datasets in the experiments, methodologies, and results—hence, further justifying utility and clinical applicability of these machine learning approaches. Section [31] focuses basically on the comparison of the performance of different variants of the decision tree algorithm within the domain of breast cancer classification. This section outlines further comparative studies presented with respect to approaches in data preprocessing, feature selection, and evaluation metrics such as accuracy, sensitivity, and specificity. It identified the best model from the decision tree

purposed and discussed its utility in the clinical setup and for the betterment of automated medical systems. Further, in [32], a methodology based on the neural network of diagnosing breast cancer is proposed.

The following section gives an idea of the architecture of the neural network used, explains steps taken in training and validation, and goes on to compare its performance and accuracy with conventional classification methods. The section will also take into consideration both the disadvantages and advantages of using neural networks for medical diagnostics. The research works on the use of a DT-SVM hybrid model in prediction of breast cancer with much focus laid on big data analytics. The study will involve integrating Decision Trees and Support Vector Machines, data sets, model training, model testing, and the effect of big data on enhancing predictive accuracy and model scalability [33]. applies a Mamdani fuzzy inference system to detect breast cancer risk, detailing the setup of the fuzzy inference system, the rules used for classification, and how this method compares to traditional risk detection methods. The paper will highlight the advantages of fuzzy logic in handling uncertainty and variability in medical data [34].

In [35], explore cancer classification using an SVM approach, detailing the methodology, dataset, and comparative analysis with other machine learning models. The findings would discuss the relevance of vector machine's performance in terms of accuracy, precision, and speed in cancer classification tasks.

focuses on enhancing breast cancer detection using a weighted naive Bayes classifier. The paper will discuss the weighting approach, dataset, feature selection, and the impact of this method on the classifier's performance, particularly in handling imbalanced datasets and improving detection rates. [36]. Weighted ideas are implemented to expand and improve the performance of standard Naive Bayes models. Breast cancer dataset-based domain-awareness weight assignment utilizing the machine learning library at UCI. The experiments show that an approach to heavy naive berries is preferable to the naive method. the application of convolutional neural networks (CNNs) to mammogram imaging systems for breast cancer detection, detailing the CNN architecture, data augmentation techniques, training process, and performance metrics, with insights on the advantages of CNNs over traditional image analysis methods. [37]. The study would detail the feature selection process, model development, and comparative analysis, highlighting how correlation-based feature selection impacts diagnostic accuracy and model efficiency [38].

In [39] Focuses on classifying mammogram images using textural features and artificial neural networks. The paper will detail the textural feature extraction process, ANN model development, and findings on the model's effectiveness in differentiating between malignant and benign mammogram images.

In [40] develops an SVM-based ensemble algorithm for breast cancer diagnosis, detailing the ensemble strategy, SVM models, performance evaluation, and how this approach improves upon single-model diagnostic systems. In [41], the use of particle swarm optimization for feature selection in predicting breast cancer recurrence. The study would detail the PSO algorithm, feature selection process, model performance, and the benefits of this approach in enhancing predictive accuracy.

In [42], r explores normalized neural networks for breast cancer classification, detailing the normalization process, neural network model, dataset, and performance comparison with traditional neural network approaches. In [43], Focuses on using a deep neural network with multiple activation functions for breast cancer diagnosis, detailing the network architecture, activation functions, dataset, and comparative performance analysis, highlighting the benefits of this approach over standard DNN models.

In [44] discusses machine learning techniques for classifying types of breast cancer, detailing the dataset, machine learning models used, feature selection techniques, and findings on model performance, particularly focusing on the effectiveness of SVMs in classifying breast cancer subtypes. The recommended methods were employed across various gene expression datasets to assess and confirm their effectiveness. The Support Vector Machine algorithm successfully differentiated between triple-negative and non-triple-negative breast cancer subtypes, resulting in fewer misclassification mistakes compared to the other three algorithms under review.

Table 1. Compare different algorithms used in the diagnosis of breast cancer.

Ref	Method	Objective	Achievements
[31]	Decision Tree Algorithms	Breast cancer classification	Analyzed the performance of various decision tree models
[32]	Neural Network	Breast cancer diagnosis classification	Proposed a neural network approach for diagnosis
[33]	DT-SVM Hybrid Model	Big data: breast cancer prediction	Integrated DT and SVM for improved prediction
[34]	Mamdani Fuzzy Inference System	Breast cancer risk detection	Utilized fuzzy logic for risk assessment
[35]	Relevance Vector Machine	Cancer classification	Applied RVM for enhanced cancer classification
[36]	Weighted Naive Bayes Classifier	Breast cancer detection predictive model	Introduced weighted approach for better detection
[37]	Convolutional Neural Networks	Mammogram imaging system for breast cancer detection	Employed CNNs for improved mammogram analysis
[38]	ANN and SVM	Correlation-based feature selection for breast cancer diagnosis	Investigated the effect of feature selection on diagnosis
[39]	Artificial Neural Network (ANN)	Textural feature-based classification of mammogram images	Utilized ANN for textural analysis in mammograms
[40]	SVM-based Ensemble Algorithm	Breast cancer diagnosis	Developed an ensemble algorithm for better accuracy
[41]	Particle Swarm Optimization Feature Selection	Breast cancer recurrence prediction	Applied PSO for feature selection and prediction
[42]	Normalized Neural Networks	Breast cancer classification	Introduced normalized neural networks for classification
[43]	Deep Neural Network	Breast cancer diagnosis	Used multiple activation functions for improved diagnosis
[44]	SVM, KNN, Naive Bayes, and DT	Breast cancer type classification using machine learning	Evaluated different models for type classification
[45]	Preprocessing Methods	Impact on Healthcare Predictions	Assessed the influence of preprocessing on predictions

III. PREPROCESS OF DATA MINING

Machine learning approaches can be categorized into two main types: unsupervised and supervised learning. In supervised learning, data instances are labeled and utilized to train the system, aiming for the highest possible accuracy. This method relies on existing datasets with predefined classifications to teach the model how to predict outcomes based on new inputs. In contrast, unsupervised learning does not use pre-labeled data. Instead, it attempts to identify patterns and relationships directly from the dataset without prior knowledge, making specific goal achievement more challenging due to the absence of predefined outcomes. Classification, a common type of supervised learning, leverages historical data to set standards for future predictions, effectively using past instances to forecast upcoming results. This method is particularly prevalent in the medical field, where large databases of patient histories, symptoms, and diagnoses are maintained by clinics and hospitals. Researchers use this extensive data to build classification models that reflect past occurrences, simplifying medical inference

through computational support. This has become increasingly feasible with the growing availability of extensive medical datasets [Reference 45].

A. *Bayes Network learning*

Applying search algorithms and quality criteria to Bayes Network learning Bayes Network's building block. Provides data structures (network structure, conditional probability distributions, etc.) and Bayes Network learning methods like K2 facilities. numDecimalPlaces provides the number of digits following the decimal point for the model's numerical output.

batchSize is the recommended number of instances for batch prediction. Implementations can determine their appropriate batch size if more or fewer instances are provided. Estimator Find conditional probability tables using Bayes Network's Select Estimator. The classifier may produce more information if debugging is enabled. searchAlgorithm is a network-searching choice method.

If doNotCheckCapabilities is set, classifier capabilities aren't checked before construction (Use with caution to reduce runtime). useADTree When ADTree, a data structure for improving count speed, is used, learning time is lowered. ADTrees users may suffer memory issues due to its RAM requirements. Disabling this option slows down structure learning algorithms and utilizes less memory. Figure 1 shows how ADTrees are typically used.

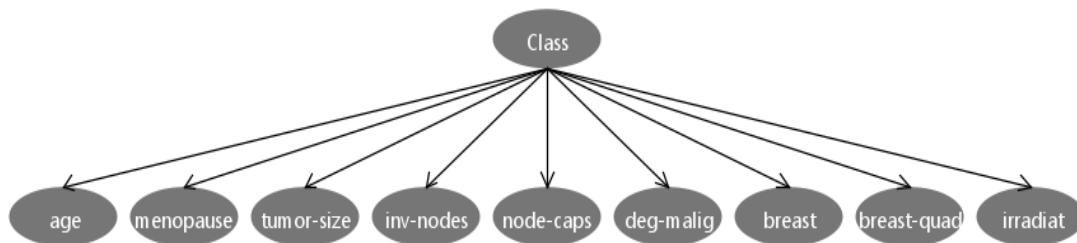


Figure 1: Bayesian network structure for Breast Cancer

B. *Naive Bayes classifier*

Class for an estimator that uses Naive Bayes's classification algorithm. The precision values of the numerical estimator are determined by analyzing the training data. Because of this, the classifier is not an UpdateableClassifier (which in common usage are initialized with zero training examples) — if you require the capability of an UpdateableClassifier, the NaiveBayesUpdateable classifier should be used instead. When buildClassifier is run with zero training instances, the NaiveBayesUpdateable classifier will use a default precision of 0.1 for numeric attributes. This precision will be used for all numeric attributes [18].

C. *Stochastic Gradient Descent*

Stochastic gradient descent is implemented to learn various linear models (binary class SVM, binary class logistic regression, squared loss, Huber loss, and epsilon-insensitive loss linear regression). Replaces, on a global scale, any values that are absent and converts nominal properties into binary ones. In addition to this, all of the qualities are normalized, and as a result, the output coefficients are derived from the normalized data.

It is necessary to make use of the squared, Huber, or epsilon-insensitive loss function when dealing with numeric class properties. Epsilon-insensitivity and Huber loss may call for a significantly faster learning rate.

D. *Bagging:*

Bagging, short for "bootstrap aggregation," is a strategy developed by Leo Breiman to enhance the accuracy of classification algorithms in machine learning [Reference 19]. The process begins with an original dataset D from which a classification algorithm generates a basic classifier H , intended to categorize data points into one of two possible categories. By leveraging the unique properties of the training set, bagging produces multiple classifiers

H_m (where m ranges from 1 to M) [20]. These individual classifiers, each trained on a different subset of the data, are then aggregated to form a final, composite classifier. This method increases the overall accuracy and stability of the classification model by reducing variance and avoiding overfitting.

E. ZeroR

Create and implement a 0-R classifier with this course! estimates either the median or mean (for a given set of numbers) (for a nominal class). Decimal places to use in model output (numDecimalPlaces). If you're performing batch prediction, you can specify the optimal batch size with the batchSize parameter. It allows implementations to choose their batch size, which may be larger or smaller depending on demand. if debug is enabled, the classifier will print debug information to the console.

IV. PERFORMANCE EVALUATION

The dataset utilized in this research is the Wisconsin Breast Cancer Dataset, accessible from the UCI Machine Learning Repository [Reference 55]. This multivariate dataset comprises over 569 instances, where cell nuclei characteristics from digital images are used to classify them as either malignant or benign, based on more than 10 distinct features [Reference 56]. These features include Area, Compactness (defined as $(\text{perimeter}^2/\text{area}) - 1$), number of concave points, concavity, and other morphological attributes. In this study, Anaconda, an open-source Python distribution, was employed as a primary tool for machine learning. Released under the New BSD License in 2012, Anaconda supports numerous machine learning algorithms and is widely used for scientific computing, data science, and predictive analytics due to its comprehensive collection of data management and analysis tools [48]. The research evaluates the effectiveness of various data mining models, presenting a comparison of their parameters and outcomes. The analysis highlighted that Support Vector Machine (SVM) and K* star algorithms exhibited the highest accuracy, reaching 98.6%, with Bagging, Backpropagation (BP), and Additive Regression following closely. This study examines *equal* performance measures of machine learning algorithms in the prediction of breast cancer. Accuracy is very important and has been used; it is defined as the number of correct predictions (both True Positives and True Negatives) divided by the total predictions made. Further, the paper discusses how precision and recall are relevant to the field of medical diagnosis as key indicators. Precision, being one of the measures, represents the number of outcomes that are *truly* positive. In other words, this is the measure of how exact the classifier is in making positive predictions. Another name for sensitivity, recall, gives *the* ability of a model to tell apart correctly between positive instances. *These* parameters are part of the clinical utility that machine learning models use toward early detection of breast cancer, making sure these models are proficient in spotting relevant instances within the dataset. Above, it is seen that the research concludes with a comparison of recall and precision side by side for five different ML models implemented in the Wisconsin breast cancer dataset. This shows recall and precision metrics with comparatively equal views seen for the Wisconsin dataset against the same five different machine learning strategies, as detailed in Table 2. This plots the points to show the prediction margin, defined as the difference in the predicted probability for the correct class and the highest predicted probability across other classes. It is posited that the success of boosting algorithms is partially attributable to their capacity to enlarge these margins within the training dataset, which subsequently leads to improved outcomes on the testing dataset. This notion posits that boosting algorithms enhance model generalization not merely by fitting to training data but also by ensuring a broader demarcation between classes, thereby facilitating more accurate classification of novel, unseen data. Figures 2, 3, 4, 5, and 6 illustrate the results of the margin curves for each respective method.

Table 2: Evaluation of algorithms for classifying Wisconsin breast cancer

<u>Method</u>	<u>Accuracy</u>	<u>Precision</u>	<u>Recall</u>
Bayesian Network	97.7%	97.5%	96.9%
Naïve Bayesian	92.6%	93.3%	95%
SGD	97.4%	97.1%	96.8%
Bagging	97%	97.2%	95.9%
ZeroR	93.5%	93.2%	93.1%

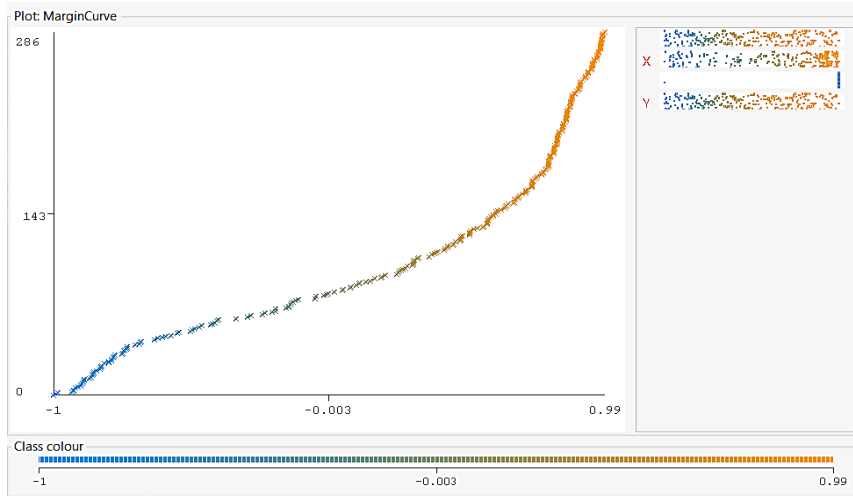


Figure 2: Margin Curve for Bayesian network.

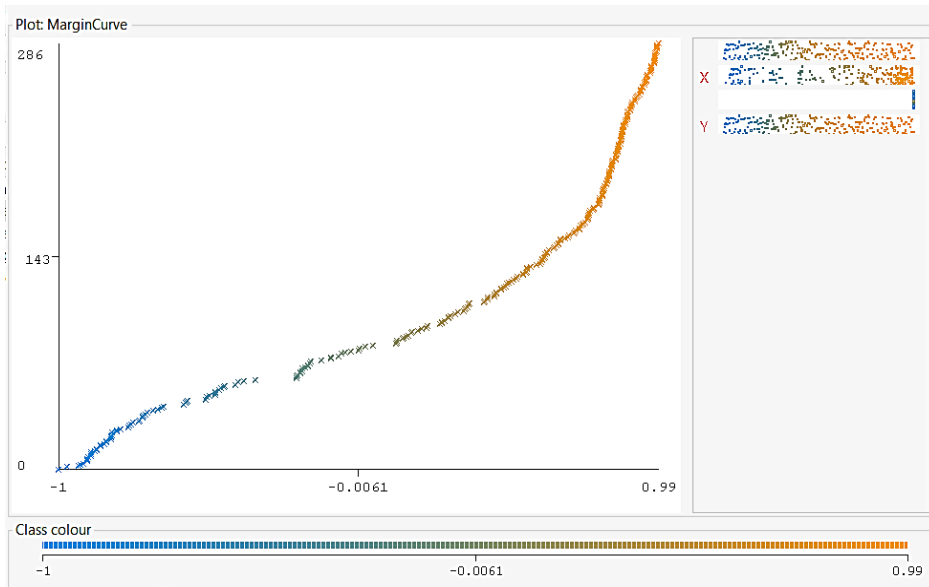


Figure 3: Margin Curve for Naïve Bayesian.

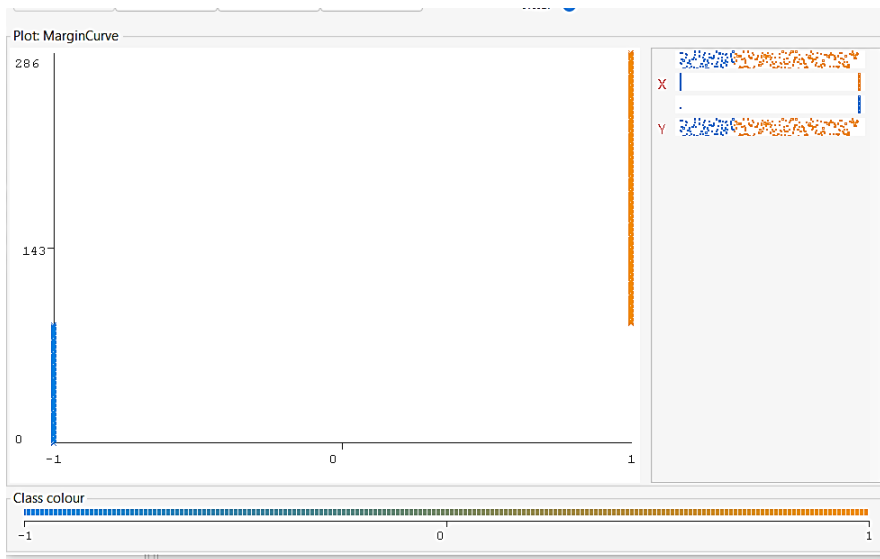


Figure 4: Margin Curve for SGD.

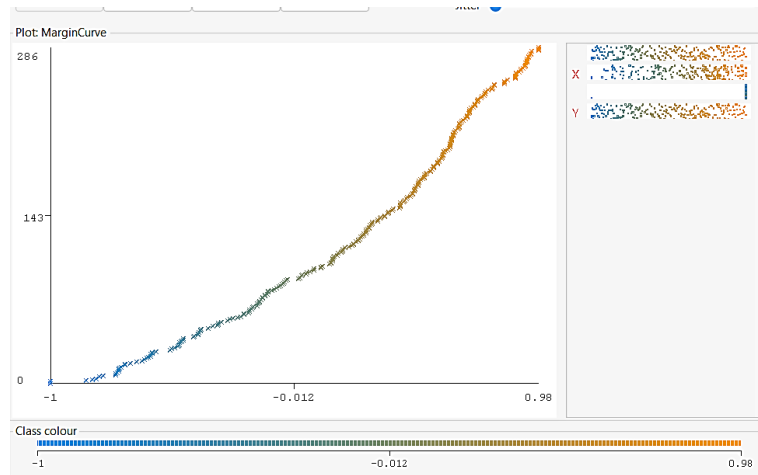


Figure 5: Margin Curve for Bayesian network.

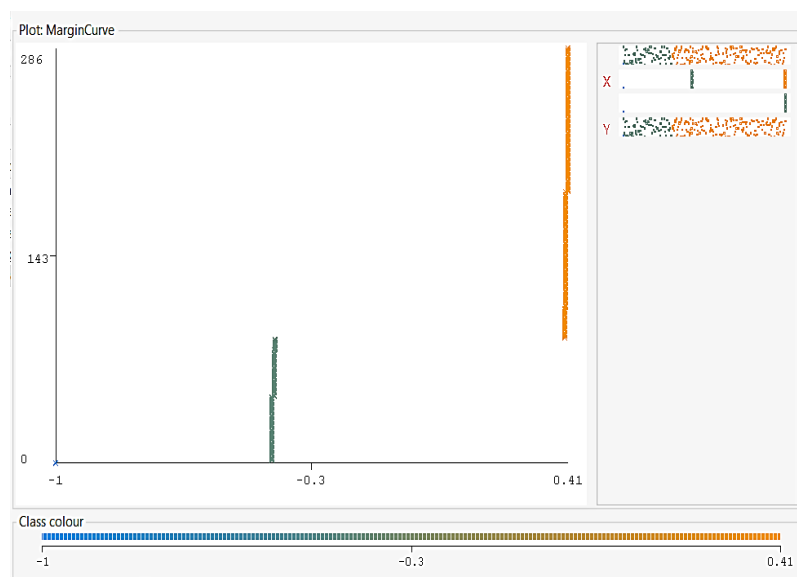


Figure 6 Margin Curve for ZeroR.

V. CONCLUSION

Breast cancer remains a significant health challenge globally, underscoring the need for accurate and effective detection methods. This paper proposes an innovative approach for enhancing breast cancer diagnosis through a suite of machine learning models, incorporating techniques such as the Naive Bayes classifier, Stochastic Gradient Descent (SGD), Bagging, and the ZeroR classifier, alongside the pivotal Bayes Network learning. Our methodology was able to differentiate the intricate correlations of diagnostic factors of different classes effectively learned by Bayes Network which would result in better accuracy of prediction. The contribution of the Naive Bayes leverages the use of probabilistic reasoning in a bid to enhance the accuracy of diagnosis. Equally important is the use of Stochastic Gradient Descent (SGD), which might improve the flexibility of learning while further enhancing the performance of the model. On the other extreme, bagging works by reducing variance and overfitting towards improving the resulting accuracy through prediction integration from several models. The ZeroR classifier further serves to set a very important threshold in comparing our ensemble method with simple classification standards. In brief, this study further extends the diagnostic accuracy of breast cancer by incorporating several methodologies from the domain of machine learning, especially probabilistic graphical modeling given by Bayes Network learning. This combination of Naive Bayes, SGD, and Bagging certainly goes well beyond, by a wide margin, the baseline set with ZeroR in diagnostic precision. Such an integrative approach would enhance the predictive accuracy and reliability of breast cancer categorization from the comprehensive, but by nature coarse, data sets that are commonly

available in healthcare. These improvements, in all respects, would give a strong contribution to patient outcomes by early detection and diagnosis, thus enabling early treatment of the disease.

REFERENCES

- [1] S. Kharya, D. Dubey, and S. Soni, "Predictive machine learning techniques for breast cancer detection," *IJCSIT International Journal of Computer Science and Information Technologies*, vol. 4, pp. 1023-1028, 2013.
- [2] S. H. Ismael, S. W. Kareem, and F. H. Almkhtar, "Medical Image Classification Using Different Machine Learning Algorithms," *AL-Rafidain Journal of Computer Sciences and Mathematics*, vol. 14, pp. 135-147, 2020.
- [3] P. Suryachandra and P. V. S. Reddy, "Comparison of machine learning algorithms for breast cancer," in *2016 International Conference on Inventive Computation Technologies (ICICT)*, 2016, pp. 1-6.
- [4] S. Kareem and M. C. Okur, "Evaluation Of Bayesian Network Structure Learning," in *2nd International Mediterranean Science and Engineering Congress (IMSEC 2017), Adana, TURKEY*, 2017.
- [5] S. W. Kareem and M. C. Okur, "Pigeon-inspired optimization of Bayesian network structure learning and a comparative evaluation," *Journal of Cognitive Science*, vol. 20, pp. 535-552, 2019.
- [6] N. Mallios, E. Papageorgiou, and M. Samarinas, "Comparison of machine learning techniques using the WEKA environment for prostate cancer therapy plan," in *2011 IEEE 20th International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises*, 2011, pp. 151-155.
- [7] A. Karplus, "Machine learning algorithms for cancer diagnosis," *Santa Cruz County Science Fair*, 2012.
- [8] A. S. Mohammed, S. W. Kareem, A. Al Azzawi, and M. Sivaram, "Time series prediction using SRE-NAR and SRE-ADALINE," *Journal of Advanced Research in Dynamical and Control Systems*, Pages, pp. 1716-1726, 2018.
- [9] M. Mori, S. Akashi-Tanaka, S. Suzuki, M. I. Daniels, C. Watanabe, M. Hirose, *et al.*, "Diagnostic accuracy of contrast-enhanced spectral mammography in comparison to conventional full-field digital mammography in a population of women with dense breasts," *Breast Cancer*, vol. 24, pp. 104-110, 2017.
- [10] T. Nagashima, M. Suzuki, H. Yagata, H. Hashimoto, T. Shishikura, N. Imanaka, *et al.*, "Dynamic-enhanced MRI predicts the metastatic potential of invasive ductal breast cancer," *Breast Cancer*, vol. 9, p. 226, 2002.
- [11] Azheen Ghafour Mohammed, "A Study of Scheduling Algorithms in LTE -Advanced HetNet", *QZJ*, vol. 6, no. 3, pp. 945-968, Sep. 2021.
- [12] S. W. Kareem, "Hybrid Public Key Encryption Algorithms For E-Commerce," *Erbil: University of Salahaddin-Hawler*, 2009.
- [13] C. S. Park, S. H. Kim, N. Y. Jung, J. J. Choi, B. J. Kang, and H. S. Jung, "Interobserver variability of ultrasound elastography and the ultrasound BI-RADS lexicon of breast lesions," *Breast Cancer*, vol. 22, pp. 153-160, 2015.
- [14] A. T. Azar and S. A. El-Said, "Probabilistic neural network for breast cancer classification," *Neural Computing and Applications*, vol. 23, pp. 1737-1751, 2013.
- [15] M. R. Haque, M. M. Islam, H. Iqbal, M. S. Reza, and M. K. Hasan, "Performance evaluation of random forests and artificial neural networks for the classification of liver disorder," in *2018 international conference on computer, communication, chemical, material and electronic engineering (IC4ME2)*, 2018, pp. 1-5.
- [16] M. M. Islam, M. R. Haque, H. Iqbal, M. M. Hasan, M. Hasan, and M. N. Kabir, "Breast cancer prediction: a comparative study using machine learning techniques," *SN Computer Science*, vol. 1, pp. 1-14, 2020.
- [17] Kh., T. & Hamarash, I. (2020). *Model-Based Quality Assessment of Internet of Things Software Applications: A Systematic Mapping Study*. International Association of Online Engineering. Retrieved March 15, 2022
- [18] Ashish Sharma, Ara Miran, Zanyar Rzgar Ahmed, The 3D Facemask Recognition: Minimization for Spreading COVID-19 and Enhance Security, *ICT Analysis and Applications*, Springer Singapore 2022 P 619-637
- [19] Al-Alkawi, H. J. M., Hanfesh, A. O., & Mohammed Rauof, S. M. N. (2019). Development of High Performance (Mechanical and Wear Properties) of AA 6061-Hybrid Nano Composites Via Liquid Metallurgy Route. *Al-Nahrain Journal for Engineering Sciences*, 22(2), 143-150.

- [20] Qasim, A.J., Din, R.E., Alyousuf, F.Q.A.: Review on techniques and file formats of image compression. *Bull. Electr. Eng. Inf.* 9(2), 602–610 (2020)
- [21] M. K. Hasan, M. M. Islam, and M. Hashem, "Mathematical model development to detect breast cancer using multigene genetic programming," in *2016 5th International Conference on Informatics, Electronics, and Vision (ICIEV)*, 2016, pp. 574-579.
- [22] M.V. Prakash, V. Porkodi, S. Rajanarayanan, M. Khan, B.F. Ibrahim, M. Sivaram, Improved conservation of energy in fog IoT services using machine learning model, in *2020 International Conference on Computing and Information Technology (ICCIIT-1441)* (2020), pp. 1–4
- [23] W. Kareem, R. Z. Yousif, and S. M. J. Abdalwahid, "An approach for enhancing data confidentiality in Hadoop," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 20, pp. 1547-1555, 2020.
- [24] S. I. Ayon and M. Islam, "Diabetes Prediction: A Deep Learning Approach," *International Journal of Information Engineering & Electronic Business*, vol. 11, 2019.
- [25] S. Chaurasia, P. Chakrabarti, and N. Chourasia, "An application of classification techniques on breast cancer prognosis," *International Journal of Computer Applications*, vol. 59, 2012.
- [26] C. Nguyen, Y. Wang, and H. N. Nguyen, "Random forest classifier combined with feature selection for breast cancer diagnosis and prognostic," 2013.
- [27] L. G. Ahmad, A. Eshlaghy, A. Poorebrahimi, M. Ebrahimi, and A. Razavi, "Using three machine learning techniques for predicting breast cancer recurrence," *J Health Med Inform*, vol. 4, p. 3, 2013.
- [28] H. Hota, "Identification of breast cancer using an ensemble of support vector machine and decision tree with reduced feature subset," *International Journal of Innovative Technology and Exploring Engineering*, vol. 3, pp. 99-102, 2014.
- [29] K. Kourou, T. P. Exarchos, K. P. Exarchos, M. V. Karamouzis, and D. I. Fotiadis, "Machine learning applications in cancer prognosis and prediction," *Computational and structural biotechnology journal*, vol. 13, pp. 8-17, 2015.
- [30] M. Rana, P. Chandorkar, A. Dsouza, and N. Kazi, "Breast cancer diagnosis and recurrence prediction using machine learning techniques," *IJRET: International Journal of Research in Engineering and Technology eISSN*, pp. 2319-1163, 2015.
- [31] E. Venkatesan and T. Velmurugan, "Performance analysis of decision tree algorithms for breast cancer classification," *Indian Journal of Science and Technology*, vol. 8, pp. 1-8, 2015.
- [32] H. T. T. Thein and K. M. M. Tun, "An approach for breast cancer diagnosis classification using neural network," *Advanced Computing*, vol. 6, p. 1, 2015.
- [33] K. Sivakami and N. Saraswathi, "Mining big data: breast cancer prediction using DT-SVM hybrid model," *International Journal of Scientific Engineering and Applied Science (IJSEAS)*, vol. 1, pp. 418-429, 2015.
- [34] B. Gayathri and C. Sumathi, "Mamdani fuzzy inference system for breast cancer risk detection," in *2015 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)*, 2015, pp. 1-6.
- [35] A. Bharathi and K. Anandakumar, "Cancer classification using relevance vector machine learning approach," *Journal of Medical Imaging and Health Informatics*, vol. 5, pp. 630-634, 2015.
- [36] S. Kharya and S. Soni, "Weighted naive Bayes classifier: a predictive model for breast cancer detection," *International Journal of Computer Applications*, vol. 133, pp. 32-37, 2016.
- [37] Y. Tan, K. Sim, and F. Ting, "Breast cancer detection using convolutional neural networks for a mammogram imaging system," in *2017 International Conference on Robotics, Automation, and Sciences (ICORAS)*, 2017, pp. 1-5.
- [38] R. Alyami, J. Alhajjaj, B. Alnajrani, I. Elaalami, A. Alqahtani, N. Aldhafferi, *et al.*, "Investigating the effect of correlation-based feature selection on breast cancer diagnosis using artificial neural network and support vector machines," in *2017 International Conference on Informatics, Health & Technology (ICIHT)*, 2017, pp. 1-7.
- [39] I. Routray and N. P. Rath, "Textural feature based classification of mammogram images using ANN," in *2018 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, 2018, pp. 1-6.

- [40] H. Wang, B. Zheng, S. W. Yoon, and H. S. Ko, "A support vector machine-based ensemble algorithm for breast cancer diagnosis," *European Journal of Operational Research*, vol. 267, pp. 687-699, 2018.
- [41] S. B. Sakri, N. B. A. Rashid, and Z. M. Zain, "Particle swarm optimization feature selection for breast cancer recurrence prediction," *IEEE Access*, vol. 6, pp. 29637-29647, 2018.
- [42] E. Alickovic and A. Subasi, "Normalized neural networks for breast cancer classification," in *International Conference on Medical and Biological Engineering*, 2019, pp. 519-524.
- [43] K. Vijayakumar, V. J. Kadam, and S. K. Sharma, "Breast cancer diagnosis using multiple activation deep neural network," *Concurrent Engineering*, vol. 29, pp. 275-284, 2021.
- [44] J. Wu and C. Hicks, "Breast Cancer Type Classification Using Machine Learning," *Journal of personalized medicine*, vol. 11, p. 61, 2021.
- [45] P. Misra and A. S. Yadav, "Impact of Preprocessing Methods on Healthcare Predictions," in Proceedings of 2nd International Conference on Advanced Computing and Software Engineering (ICACSE), 2019.
- [46] S. W. Kareem, F. H. Almkhtar, A. T. Guron and H. M. Salman, "Medical Image Categorization Combining Image Segmentation and Machine Learning," *2023 9th International Engineering Conference on Sustainable Technology and Development (IEC)*, Erbil, Iraq, 2023, pp. 38-44, doi: 10.1109/IEC57380.2023.10438714..
- [47] I. Kononenko, "Machine learning for medical diagnosis: history, state of the art and perspective," *Artificial Intelligence in medicine*, vol. 23, pp. 89-109, 2001.
- [48] Y. Yasui and X. Wang, "Statistical Learning from a Regression Perspective by BERK, RA," ed: Wiley Online Library, 2009.
- [49] W. Yue, Z. Wang, H. Chen, A. Payne, and X. Liu, "Machine learning with applications in breast cancer diagnosis and prognosis," *Designs*, vol. 2, p. 13, 2018.
- [50] F. Cugnata, R. S. Kenett, and S. Salini, "Bayesian networks in survey data: Robustness and sensitivity issues," *Journal of Quality Technology*, vol. 48, pp. 253-264, 2016.
- [51] S. W. Kareem and M. C. Okur, "Evaluation of Bayesian Network Structure Learning Using Elephant Swarm Water Search Algorithm," in *Handbook of Research on Advancements of Swarm Intelligence Algorithms for Solving Real-World Problems*, ed: IGI Global, 2020, pp. 139-159.
- [52] S. W. KAREEM, "NOVEL SWARM INTELLIGENCE ALGORITHMS FOR STRUCTURE LEARNING OF BAYESIAN NETWORKS AND A COMPARATIVE EVALUATION," YAŞAR UNIVERSITY, 2020.
- [53] S. K. Sarkar and A. Nag, "Identifying patients at risk of breast cancer through decision trees," *International Journal of Advanced Research in Computer Science*, vol. 8, pp. 88-91, 2017.
- [54] S. Kareem and M. C. Okur, "Bayesian Network Structure Learning Using Hybrid Bee Optimization and Greedy Search," *Adana, Turkey: Çukurova University*, 2018.
- [55] N. Louridi, M. Amar, and B. El Ouahidi, "Identification of Cardiovascular Diseases Using Machine Learning," in *2019 7th Mediterranean Congress of Telecommunications (CMT)*, 2019, pp. 1-6.
- [56] R. S. Ismael, R. S. Youail, and S. W. Kareem, "Image encryption by using RC4 algorithm," *European Academic Research, II*, vol. 4, pp. 5833-5839, 2014.
- [57] I. Staff, *2019 IEEE 15th International Colloquium on Signal Processing and Its Applications (CSPA)*: IEEE, 2019.
- [58] D. Bazazeh and R. Shubair, "Comparative study of machine learning algorithms for breast cancer detection and diagnosis," in *2016 5th International Conference on Electronic Devices, systems and Applications (ICEDSA)*,