Computer and Decision Making

An International Journal

www.comdem.org
eISSN: xxxx-xxxx

Comprehensive Evaluation of Machine Learning and ANN Models for Breast Cancer Detection

Yigitcan Cakmak[1], Sinem Safak[1], Muhammed Ali Bayram[1], Ishak Pacal[1,*]

[1] Computer Engineering, Faculty of Engineering, Igdir University, 76000, Igdir, Turkey

ARTICLE INFO

Article history:
Received 23 July 2024
Received in revised form 19 August 2024
Accepted 21 August 2024
Available online 21 August 2024

Keywords:

Breast Cancer Detection; Machine Learning; Artificial Neural Networks; Logistic Regression; Naïve Bayes; Support Vector Machines

ABSTRACT

Breast cancer is one of the most common types of cancer among women, making early detection critically important. In recent years, machine learning algorithms have demonstrated high accuracy in identifying breast cancer, contributing significantly to earlier diagnoses. Various machine learning models can analyze tumor characteristics, aiding in cancer identification and treatment decision-making. In this study, we comprehensively evaluated the performance of 14 different machine learning algorithms and an improved Artificial Neural Network (ANN) model using the Wisconsin dataset for breast cancer detection. Experimental results revealed that the developed ANN model outperformed other machine learning algorithms, achieving an impressive accuracy of 98.25%. This study highlights the potential impact of machine learning algorithms in clinical applications, enabling researchers and healthcare providers to enhance the accuracy and efficiency of breast cancer diagnosis, ultimately improving patient outcomes and advancing this critical field of oncology.

1. Introduction

The most common cancer in women worldwide, breast cancer, is a disease where cells from the breast tissue grow without control, resulting in the formation of tumors. [1].Cancer usually appears as a palpable lump in the breast, yet there might be no symptoms in the early phase. Different risk factors also contribute to the development of breast cancer, including genetic tendencies, family history, age, hormonal treatments as well as lifestyle-related behavior. Besides this, mutations in genes like BRCA1 and BRCA2 cause a very high risk [2]Early detection is crucial in the proper fight against cancer. Therefore, regular mammography screenings and self-examinations are crucial, and low-fat foods are the best choice. A good diet and reduction of excess body weight help prevent cancer. [3]The treatments are differentiated by their stages and kinds, and they may include surgical intervention, chemotherapy, radiotherapy, hormone therapy, and targeted therapies. [4].

* *Corresponding author.*
E-mail address: ishak.pacal@igdir.edu.tr

Progress in research and awareness and treatment campaigns implies that scientists can now detect the disease earlier and provide good treatment. [3]. Furthermore, healthy living styles like good nutrition, regular exercise, and less alcohol use are directly associated with reducing the possibility of the disease [6]. The growing focus on research and public health education has led to a deeper understanding and improved management of breast cancer. In recent years, there has been a significant increase in cancer cases, with various types of cancer emerging as prominent threats to human health. Cancer, a devastating disease, remains a global challenge, affecting approximately one in every six people. Among these, lung cancer is the most prevalent type worldwide, while liver cancer ranks as the leading cause of cancer-related deaths among women. These trends underscore the critical importance of continued efforts in research and public health initiatives to combat this pervasive and life-threatening disease. [5], [6] . The quicker the diagnosis is, the greater the chances of a cure [7] Due to the asymmetry of symptoms among patients, complete characterization and timely detection are necessary for an early diagnosis.

Detecting breast cancer using traditional methods can be inefficient and costly, complicating the treatment process further. However, with the advent of big data and ANN-based tolls, we can now focus on quality rather than quantity. [8]. AI algorithms are extensively utilized in the healthcare sector due to their versatility and ability to perform various tasks, such as disease prediction and diagnosis, reducing medication costs, and enabling real-time decision-making. [9]. A comparison of thirteen classification systems (Logistic Regression (LR), Gaussian Naive Bayes (GNB), Bernoulli Naive Bayes (BNB), Multinomial Naive Bayes (MNB), Support Vector Machines (SVM), Decision Trees (DT), Random Forest (RF), K-Nearest Neighbors (KNN), Gradient Boosting (GB), Extreme Gradient Boosting (XGBoost), AdaBoost, LightGBM (LGBM), Linear Discriminant Analysis (LDA) underscores the importance of supporting cancer research and treatment funding. To enhance overall healthcare and cancer treatment, ANN-based systems must be employed to boost the clinical performance of information and communication technologies. [10]. Artificial intelligence's capability to analyze vast amounts of patient data for patterns that human physicians might miss can significantly improve cancer detection and treatment. This could lead to better, more personalized care for patients. Various machine-learning techniques can be employed to diagnose and predict breast cancer. [11]Early detection is essential for successful treatment in this particular science sector, as various potential options have been demonstrated to be successful. The necessity of boosting breast cancer tests' efficiency (effectiveness) is non-negotiable.

Artificial intelligence and machine learning are technologies used to detect and determine the nature of cancer. [12]ANNs have generally been appreciated for their ability to analyze complicated datasets and search for new patterns. [13], [14], [15]. These techniques offer great potential in the diagnosis of diseases like cancer. [16]. An important aspect that is automatically identified is that these models are enabled to learn from large datasets and thus spot small changes that traditional approaches cannot identify, and this leads to a diagnosis more rapidly and possible complications being eliminated. [17]. These latest innovations in diagnosing cancer allow the disease sufferers to be treated early on, thus curing the disease and resulting in lesser utility medical intervention. [18]. Artificial intelligence and machine learning approach much more patient-centered therapies and better solutions in cancer treatment. Thus, they are providing a fundamental solution to this great health issue. [19].

This article explores the use of various machine learning algorithms and an improved ANN model to detect breast cancer. Recognizing the crucial role of early diagnosis in improving patient outcomes, the study evaluates 14 machine-learning models using a breast cancer dataset. The process includes thorough data preprocessing and training and testing phases to determine the most accurate model.

The research highlights how machine learning can enhance the precision and efficiency of breast cancer diagnosis. By adopting these advanced computational techniques, healthcare providers can detect cancer earlier, avoid unnecessary medical procedures, and advance the field of oncology. The results demonstrate the significant potential of integrating artificial intelligence and machine learning into clinical practices, leading to more personalized and effective patient care.

1.1 Literature Review

Attallah et al. introduce Histo-CADx, a new system for automatically diagnosing breast cancer. Unlike past approaches that relied on single deep learning models, Histo-CADx combines features from multiple CNNs and handcrafted methods to improve accuracy. The system uses a two-stage fusion process, first blending deep learning techniques with handcrafted features to find the best combination for enhancing performance. [20]. Ragab et al. developed a CAD system using deep learning to help radiologists classify breast cancer in mammograms. They tested different approaches, including fine-tuning pre-trained neural networks and using their deep features with SVM classifiers. They found that combining features from multiple networks improved classification accuracy. [21]. Mushtaq et al. explored how well KNN performs using different distance measures and K values. They tested this on two breast cancer datasets from UC Irvine. The experiments were done in three rounds: first without feature selection, then using L1-norm and a linear support vector classifier, and finally with Chi-square feature selection. They evaluated the results using accuracy, AUC, and sensitivity metrics to find the best approach. [22]. Alshayeji et al. developed a simple artificial neural network (ANN) model with just one hidden layer to diagnose and predict breast cancer. For this purpose, they used two well-known datasets, the Wisconsin Breast Cancer Dataset (WBCD) and the Wisconsin Diagnostic Breast Cancer (WDBC) dataset. Notably, they didn't apply any feature optimization or selection techniques, keeping the approach straightforward. [23].

Aamir et al. highlighted the development of computer-aided diagnosis systems to assist radiologists and reduce misdiagnoses. However, due to the serious consequences of delayed or incorrect diagnoses, there is an urgent need to improve these systems. Machine learning has emerged as a promising technique for the early and accurate detection of breast cancer, addressing this critical need. [24]. Hernández-Julio et al. proposed that DSS are valuable tools for helping decision-makers. These systems rely on a knowledge database and rules to function effectively. Their research focused on validating clinical decision support systems by applying Mamdani-type fuzzy set theory. They tested their results using the Wisconsin breast cancer dataset and compared their findings with existing studies in the literature. [25]. Dubey et al. tried to find out whether the k-means clustering algorithm can be utilized for the early detection of breast cancer, which is a key element for the real and successful treatment of this disease. The investigation tested the power of a few different computational methods, like centroid initialization, distance metrics, split methods, epoch thresholds, attributes, and iterations, on the BCW. The Breast Cancer Wisconsin (BCW) diagnostic dataset has been used in the study. [26]. Nemade and Fegade explored various machine-learning classification techniques on a breast cancer dataset. They used methods like Naïve Bayes, Logistic Regression, Support Vector Machine, K-Nearest Neighbor, and Decision Tree, along with ensemble techniques like Random Forest, AdaBoost, and XGBoost. By evaluating these techniques with various performance measures, they discovered that the Decision Tree and XGBoost classifiers achieved the highest accuracy at 97%. Notably, the XGBoost classifier also recorded the highest AUC of 0.999 [27].

Chaurasia and Pal's study compares six machine learning algorithms to assess how the WDBC dataset impacts their prediction accuracy. The primary goal is to demonstrate the performance of

the classifier libraries. Innovations are possible with advancements in new hardware graphical interfaces, potentially leading to better outcomes. The study explores different configurations of these models, treating them as specifically tailored to the reader. The study found that all models performed exceptionally well, with an average accuracy of over 90%, even with a limited feature set [28]. Mushtaq et al. introduced a method that blends the Naive Bayes algorithm with a Gaussian distribution and Chi-squared-based attribute selection. To manage the data's dimensionality, they used an extended Kernel Principal Component Analysis (K-PCA) and tested five kernels. After thorough pre-processing, they evaluated their system's performance using accuracy, precision, sensitivity, and specificity metrics. Impressively, their method, which utilized six selected features and the sigmoid K-PCA, achieved an accuracy of 99.28% [29]. Pascal [30] extensively reviewed deep learning-based approaches for classifying breast ultrasound (US) images, comparing the performance of CNN architectures like AlexNet, VGG, ResNet, GoogleNet, and EfficientNet. Additionally, transformer models were examined, which have recently gained popularity for their comparable performance to CNNs in medical imaging. The study demonstrated that transformer and CNN models effectively classify breast US images using the publicly available BUSI dataset. Isik and Pascal developed a method to classify breast cancer in ultrasound images using few-shot learning. They tested two meta-learning approaches, ProtoNet and MAML, on the challenging BUSI dataset. Using a ResNet50 backbone with ProtoNet, they achieved an accuracy of 0.882–0.889, improving the baseline by over 6%. ProtoNet consistently outperformed MAML in their tests [16].

2. Methodology

2.1 Dataset and Preprocessing

2.1.1 Introduction to the Wisconsin Breast Cancer Dataset

The dataset used in the study consists of breast cancer data obtained from the University of Wisconsin Hospital. The dataset consists of 569 samples and 31 attributes. Each sample represents a patient's tumor characteristics; the classification objective is whether the tumor is benign (0) or malignant (1). There are 357 benign and 212 malignant patients features [31]. Singh and colleagues presented three feature selection strategies that use metaheuristic algorithms: the Gravitational Search Optimization Algorithm (GSA), Emperor Penguin Optimization (EPO), and a hybrid of GSA and EPO (hGSAEPO). Although previous research has utilized basic algorithms for feature selection in various machine learning tasks, their use in breast cancer (BC) classification has not been fully explored. This study aims to fill that gap by combining these methods. They chose to focus on breast cancer because it is the second leading cause of death among women. Early detection is crucial, as it can lead to effective treatment and help avoid unnecessary medical procedures [32].

2.1.2 characteristics and classes of the dataset

The dataset used in this study comprises various characteristics of tumor cells that are essential for breast cancer detection. Each feature provides crucial information regarding the physical attributes of the tumor cells. The characteristics and their corresponding descriptions are depicted in Table 1.

Table 1

Dataset features and descriptions

Features	Description
x.radius_mean	Mean radius of the tumor cells

x.radius_mean	Mean radius of the tumor cells
x.texture_mean	Mean texture of the tumor cells
x.perimeter_mean	Mean perimeter of the tumor cells
x.area_mean	Mean area of the tumor cells
x.smoothness_mean	Mean smoothness of the tumor cells
x.compactness_mean	Mean compactness of the tumor cells
x.concavity_mean	Mean concavity of the tumor cells
x.concave_points_mean	Mean number of concave portions of the contour of the tumor cells
x.symmetry_mean	Mean symmetry of the tumor cells
x.fractal_dimension_mean	Mean "coastline approximation" of the tumor cells
x.radius_se	Standard error of the radius of the tumor cells
x.texture_se	Standard error of the texture of the tumor cells
x.perimeter_se	Standard error of the perimeter of the tumor cells
x.area_se	Standard error of the area of the tumor cells
x.smoothness_se	Standard error of the smoothness of the tumor cells
x.compactness_se	Standard error of the compactness of the tumor cells
x.concavity_se	Standard error of the concavity of the tumor cells
x.concave_points_se	Standard error of the number of concave portions of the contour of the tumor cells
x.symmetry_se	Standard error of the symmetry of the tumor cells
x.fractal_dimension_se	Standard error of the "coastline approximation" of the tumor cells
x.radius_worst	Worst (largest) radius of the tumor cells
x.texture_worst	Worst (most severe) texture of the tumor cells
x.perimeter_worst	Worst (largest) perimeter of the tumor cells
x.area_worst	Worst (largest) area of the tumor cells
x.smoothness_worst	Worst (most severe) smoothness of the tumor cells
x.compactness_worst	Worst (most severe) compactness of the tumor cells
x.concavity_worst	Worst (most severe) concavity of the tumor cells
x.concave_points_worst	Worst (most severe) number of concave portions of the contour of the tumor cells
x.symmetry_worst	Worst (most severe) symmetry of the tumor cells
x.fractal_dimension_worst	Worst (most severe) "coastline approximation" of the tumor cells
y	target

2.1.3 Data preprocessing steps

Several pre-processing steps were performed on the dataset to ensure the quality and effectiveness of the training process. First, the ID column in the dataset was removed. The ID column, which uniquely identifies each sample, does not contain any useful information for the predictive modeling task. Including the ID column in the training process could lead to overfitting, as the model might learn to recognize the IDs rather than the actual patterns in the data. Removing this column helps to improve the quality of the training data and ensures that the model focuses on the relevant features.

Next, missing values in the dataset were addressed. Any rows with missing or null values were either removed or imputed using statistical methods, such as the mean or median of the respective feature. Additionally, the dataset was normalized to bring all features onto a similar scale, which is essential when working with machine learning algorithms that rely on distance calculations. Categorical variables, if present, were encoded into numerical values using techniques such as one-hot encoding or label encoding. Finally, logarithmic transformations of the target attributes to numerical values were not applied in this study, as the target variable was already in a suitable format for the classification task.

2.2 Algorithms Used

2.2.1 Logistic Regression (LR)

LR is an approach that is widely known and used in the recognition of supervised learning, usually, it is used to solve problems that involve binary classification [33]. The LR model relies on the sigmoid function to estimate the likelihood of the target variable, which results in an output value in the range of 0 to 1, where 1 is the probability of the target variable and 0 is the opposite. But its major goal is to detect the connection between the independent variables and the dependent variable. The model parameters are usually found by the minimization of the Mean Squared Error (MSE) loss function using the Stochastic Gradient Descent (SGD) optimization method. Moreover, the features such as the interpretability and explainability of the LR model's outcomes provide the algorithm by the readability and accountability. To sum up, LR is a learning-based model that can be used for small and medium-sized businesses to solve classification problems as depicted in Eq. 1.

$$\mathrm{P(Y=1|X)} = \frac{1}{1+\mathrm{e}^{-(\beta_0+\beta_1 \mathrm{X}_1+\beta_2 \mathrm{X}_2+\cdots+\beta_\mathrm{n} \mathrm{X}_\mathrm{n})}} \quad (1)$$

2.2.2 Gaussian Naive Bayes (GNB)

The GNB classifier is the model best suited for supervised machine learning, evolving from the Bayes' Theorem of an object in a certain class [34]. However, GNB and the LR are similarly in that each are probabilistic in nature, which originates from the "naive" belief that a subjectivity's logical dependence on the process of another object is not affected by that object's being in another part of the class. The idea of notion and technical foundation of the Naive Bayes theorem is almost the same as of the Bayes' theorem, a famous mathematical formula for calculating conditional probabilities. This theorem relates the link between P(A|B), the chance of the happening of event A if the event A is observed, and the a priori probabilities P(A) and P(B) of events A and B. More explicitly, P(B|A) is the probability of event B under the condition of the occurrence of event A. The symbolization of Bayes' Theorem is depicted Eq. 2.

$$P(C_k|x) = \frac{P(C_k)\prod_{i=1}^{n} P(x_i|C_k)}{P(x)}$$

$$P(C_k|x) \propto P(C_k)\prod_{i=1}^{n} P(x_i|C_k) \quad (2)$$

This function brings forward an understandable way of calculating the a posteriori probability of the Naïve Bayes classification method, which is the main component. The algorithm assumes such things as features used in the classification process as the Gaussian or normal distribution which makes them perfect for iterative processing In the end, The GNB classifier is an extensively used machine learning technique of the probabilistic type, which performs efficiently and with a high degree of classification accuracy on Bayes's theorem's most elemental principles.

2.2.3 Bernoulli Naive Bayes (BNB)

Used in Bayesian probability, BNB is a famous algorithm that is successfully implemented for learning in different problems such as breast cancer detection [34]. A new model of the Naive Bayes algorithm is discussed, which is typically expressed as the classification having characters that are independent of each other. One of the Bernoulli versions is Good Enough for binary classifications, as it assumes that they are binary features. With this algorithm, which models the distribution of

these binary features, the BNB algorithm is looking into the relations between the input variables and the target class. This quality makes it a nice choice for breast cancer diagnosis.

2.2.4 Multinomial Naive Bayes (MNB)

The MNB is another variant of the Naive Bayes classifier that is more ideal for text classification tasks [34]. As the Bernoulli version, it may not be as widely used for breast cancer detection. However, the MNB algorithm too may be beneficial. Especially in the case of high-dimensional data, such as those that are representable by frequency vectors like gene expression profiles, it can also be suitable for high-dimensional data especially for gene expression profile systems. Through quantifying the distributions of these feature frequencies, the MNB algorithm can be brought into the cycle of new diagnostic solutions and it can enlightened create the difference between classes perhaps of breast cancer, so that it becomes a foreseeable scenario for more accurate and efficient diagnostic processes.

2.2.5 Support Vector Machines (SVM)

SVM are another common supervised machine learning method research that is utilized for classification tasks [35]. This strategy is grounded on the idea of a hyperplane that splits the set into different subgroups as well as for the highest accuracy, the optimal margin is provided. Here, regardless of the margin being either high or low each means the distance between the two closest data points that are of different classes and are so close to the hyperplane. These advanced models can learn and work out different actions like classifying, predicting, and detecting anomalies. The general idea of the method-SVM is to come up with a hyperplane that sea parameters data used for the various classes most effectively that is also distant from each class points - close or far away distances.

The SVM classifier's scope of implementation takes in the hyperplane in such a way that the largest minimum distance between the desired classes and the hypoplant class is achieved. The model is then said to suffer from the generalization error of the model all depending on this configuration. The SVM algorithm attempts to identify the most effective hyperplane among many through this optimization process. The AI system has quite an original sentence but eventually, it becomes easier and takes less effort to develop it more. So, in conclusion Support Vector Machine is a powerful supervised learning approach which models the concepts of hyperplane and margin to provide efficient classification as well as prediction and anomaly detection on complicated datasets.

2.2.6 Decision Trees (DT)

A Decision tree (DT) models belong to widely-used supervised machine learning algorithms and are commonly used for decision-making [36]. These models are very similar and include such features as decision points, strategy (event) points having a structure of the tree that allows to study various alternatives at specific circumstances. The prime focus of DT algorithms is to enhance the future performance or reduce predicted costs by imaging what would be effective and in comparison, sort it out as depicted in Eq. 3. Another famous method of classification task used in the supervised machine learning community is Decision Tree classifier. The tree is a classifier that attempts to partition the dataset into different classes creating a tree like structure with internal nodes being features of dataset used in decision making, branches representing the decision rules and leafs represent outcome (class labels). Basically, the DT classifier asks a series of questions and on the basis

of the answers (yes/no) splits the tree into subtrees, thus enabling efficient decision making. Among the evaluation of DT models, with a high relation to entropy representing it as is that in reality which means how much information tool necessary for portrayal about a given sample. Entropy assumes a value of zero when the sample is entirely homogeneous -sensitivity to all elements are equal-, and it reaches one in an intervened halfway split metric. In addition, Gini index is a common statistical criterion for measuring the inequality of a sample commonly used in DT algorithms. A Gini coefficient of 0 represents a perfectly identical sample with each element equal to all others, whereas an index of one depicts the highest level30of inequality among the individuals within that particular subgroup.

$$Entropy(D) = -\sum_{i=1}^{C} p_i \log_2(p_i) \quad (3)$$

2.2.7 Random Forest (RF)

RF is a machine learning model developed by an ensemble in which a large number of models are joined together, each operating differently but the combinations of outputs win at the end of each prediction [37]. The basis on which RF is built is the capability to make good predictions from a set of many rather of just one decision tree, each of which has been derived having used a random subset of the data according to a principle. This idea differs from using only one decision tree, which can sometimes produce a model that is shorter than expected or more specific than anticipated. Rewarding the RF model with the opportunity to process fuzzy data and balance the less significant cases (minimizing false negatives) is a non-trivial requirement in the detection of cancer. Listening to the fragmentation of a benign or malignant tumor where the malignant category might have come off with a portion of the input data as depicted in Eq. 4. The RF algorithm develops with the help of its recursion, that involves repeatedly taking data samples with replacement and discarding the unpartitioned data, thus making it possible to handle class imbalance problems and give dependable forecasts.

$$\hat{y} = \text{mode}\{h_1(\boldsymbol{x}), h_2(\boldsymbol{x}), \dots, h_m(\boldsymbol{x})\} \quad (4)$$

The RF algorithm, at first, is realized by the generation of several decision trees, which is used to train each on a randomly chosen portion of the input data. These individually developed DT are intertwined in a majority decision process determining the final classification or regression outcome. The model's predictive accuracy is intensified and overfitting is avoided, since the decision trees' collective results tend to be more robust and generalized than a single tree. The RF model additionally brings randomness within the process by randomly selecting features for node splitting, which makes the DT more diverse and raises the overall model efficiency.

2.2.8 K-Nearest Neighbors (KNN)

KNN algorithm that can be seen as a non-parametric supervised learning-based classifier that uses the pertinent information of proximity to carry out the tasks of classification and prediction [38]. The algorithm takes note of the attribute vectors and the labels needed during the training phase which can then be used to retrain it as and when needed as depicted in Eq. 5. In the case of the unlabeled vector, the KNN algorithm determines the 'K' the user wants and then labels it accordingly. If there are any training attributes that are a bit extensive, one can be very excited about them, and they are included in addition to other attributes that are relevant in the relevant vector. Exclusively for continuous variable types, the most used distance metric known as Euclidean distance is going to

be based on the formula that has been given. This distance metric is designed for real-valued vectors only. In the case of discrete variables, the Zhao-Bhatta metric is commonly the one used.

Exploiting extraordinary results in the various stages, KNN proved to be the proceedings classification and predictive one of the leading models. Also, in data processing, it has shown flabbergasting finding on the resolution of inexact value in prognosis, automatic recommendations, and forecasting outcomes in the areas of finance, credit data, and healthcare. Specifically, the KNN algorithm was able to make correct predictions of, for instance, cardiovascular diseases, breast cancer, and prostate cancer. Not like other machine learning algorithms, the KNN method is nonparametric, that is, it does not make any assumptions on the underlying data distribution. This makes possible the processing of nonlinear data effectively. Moreover, the KNN algorithm is impractical because it certainly needs numerous data storage spaces, and it is more worryingly being dependent on storing the entire training dataset for making predictions on new data points.

$$\hat{y} = \text{mode}\{y_i : \boldsymbol{x_i} \in \mathcal{N}_k(\boldsymbol{x})\} \tag{5}$$

2.2.9 Gradient Boosting (GB)

Gradient Boosting is an astute machine learning mechanism that binds several less competent models together to give a stable and precise forecasting system [39]. The basic concept behind this technique is to add new models sequentially to the ensemble, each new member is targeted at eliminating the errors resulting from the earlier models as depicted in Eq. 6. Gradient Boosting is different from other methods that boost in that it does not require the user to define a specific base estimator. It uses an already fixed base estimator which can be used for classification and regression tasks at the same time. As for regression problems, the mean squared error (MSE) is typically utilized as the cost function, while for classification, the log loss function is often considered the best.

The formula for math underpinning Gradient Boosting depicts how the final team 'F' is formed by repeatedly including weak learners' 'f' (with a given learning rate 'η' and the input data 'X'). However, this process does not necessarily ensure that the model will be output-ordered. It could still be a problem when this candidate comes across. No worries, though; this iterative process guarantees at least one of those situations to never appear again. Gradient Boosting is quite frequently used in different areas, ranging from data mining and machine learning to artificial intelligence. This robust technique has proved to be successful in solving numerous problems over time, with the help of the input from both researchers and practitioners, even in the case of breast cancer prognosis, the results of which were quite amazing. Gradient Boosting is an incredible technique that you can use to change and adapt in the context of the new conditions present nowadays. As this method unifies a number of weak models to create a very reliable predictive system, the results are such a high accuracy and dependability, that it has a big following among engineers and scientists.

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x) \tag{6}$$

2.2.10 Extreme Gradient Boosting (XGBoost)

XGBoost or eXtreme Gradient Boosting is an empowered machine learning algorithm that is widely favored in recent years because of its lightning speed and accuracy [40]. This mechanism is built on the principle of gradient boosting, which is the combination of several weak models (such as decision trees) in order to obtain a good data model that is predictive. XGBoost, specifically, is known for its efficiency, scalability, and the variety of data types it can handle, including complex, high-

dimensional disease data in breast cancer research. Through the ability of the algorithm to capture fine relationships and complex structures in data, researchers and practitioners may build breast cancer detection and diagnosis models that are highly accurate and robust.

2.2.11 AdaBoost

AdaBoost, or Adaptive Boosting, is a robust machine learning caliber algorithm which operates by building a stabilization model through the uniting of several simple or feeble classifiers [41]. The primary principle of this methodology is to dynamically weight the input data, picking the most revealing bits of information to train each weak classifier successively. In boosting at each step, weak classifiers are trained and their weights of importance are adjusted based on data samples until the ensemble model is created. As a result, single errors are given more weight and the classifier learns a little bit from each of them, leading eventually to a final classifier that is good at classifying everything.

Based on the image, the AdaBoost method builds many weak classifiers each of which is assigned its respective weight on the individual's performance. The last result is then found by adding all weak classifiers and using a voting scheme with their contribution One of the very important benefits of AdaBoost is that it is a very simple algorithm that is needed a few input parameters and a basic understanding of the weak learners that are being combined. Furthermore, through its flexible nature, the algorithm can be easily integrated with other methods that can be used to identify and deal with weak clarifications, which means that it is a tool that can be used in a variety of ways besides machine learning. The mathematical form of AdaBoost is presented in the equation and it inklings about how the algorithm adds the weighed outputs of the weak individual classifiers to come up with the final decision as depicted in Eq. 7. The fundamental basis of AdaBoost's accomplishment, which is the iterative process, is the point in fact that each new classifier aims at correcting the errors of the previous ones.

$$H(x) = \text{sign}(\sum_{t=1}^{T} \alpha_t h_t(x)) \quad (7)$$

2.2.12 LightGBM (LGBM)

LightGBM beats Expectation which is another algorithm based on boosting that has become very popular and is known as the preferred selection for several machine learning tasks, one of them which is breast cancer detection. This algorithm was made in a way that it is memory-friendly while focusing on minimizing the use of the memory and speeding-up the computational speed. With the help of histogram-based feature splitting and gradient-based one-side sampling, LightGBM can easily tackle large, high-dimensional datasets, which makes it an amazing option for breast cancer research. The option to automatically pick the most essential features and the fact that it performed stably have been the major factors boosting the project's expansion in the field of breast cancer detection and diagnosis.

2.2.13 Linear Discriminant Analysis (LDA)

Linear Discriminant Analysis (LDA) is a great analysis method for breast cancer detection and diagnosis. LDA is mostly used to identify the most relevant features that are able to distinguish between the classes [42]. LDA is a method that tries to minimize the confusion in a classification problem by maximizing the distance between two different classes and, therefore, helps to

determine the maximum amount of distinction as formulated in Eq. 8. This technique is the reason LDA frequently outperforms PCA in classification problems by the most discriminative features. The algorithm derives the eigenvalues of the covariance matrix and produces the best transformation for separation. During the study, the team separated LDA training into different datasets and successfully validated the algorithm powerfully so as to prevent the model from just fitting the training data and to do some reliable testing. By leveraging LDA, researchers can enhance breast cancer detection and diagnosis accuracy and reliability, improving patient outcomes and advancing healthcare.

$$W = {S_W}^{-1} S_B \tag{8}$$

2.2.4 Artificial Neural Networks (ANN)

ANNs are a model that acts as if it were a living being and can understand a network of biological cells and process complex relationships through the provision of many layers [43]. Cancer screening accuracy and detection are the most defining characteristics of the decision support system. An ANN consists of an input layer, one or more hidden layers, and an output layer as in Figure 1. Every layer operates on the input data processed from the previous layer and transmits it to the next layer as formulated in Eq. 9. The model is trained via weights and the activation functions [44]. The most common ANNs are structured as feedforward networks and use the backpropagation algorithm for learning. The weights are changed by backpropagation each time here is the error to be minimized. The model can be mathematically expressed using the formula.

$$a_j = \phi\left(\sum_{i=1}^{n} w_{ji} x_i + b_j\right) \tag{9}$$

Where W is the weight matrix, b is the bias term, and f is the activation function. An ANN is a type of computational model that aims to replicate how the human brain operates. It has three main parts: the input layer, hidden layers, and the output layer. The input layer is where the model receives data, with each neuron corresponding to a specific input. Hidden layers, which can be one or more, process the inputs from the previous layer and generate outputs using activation functions. These layers are linked by weights and bias terms that are fine-tuned during training. The output layer produces the model's final predictions or classifications, with each neuron representing an output category or value. This architecture enables ANNs to understand complex data patterns and make accurate predictions. Figure 1 illustrates a typical ANN structure.

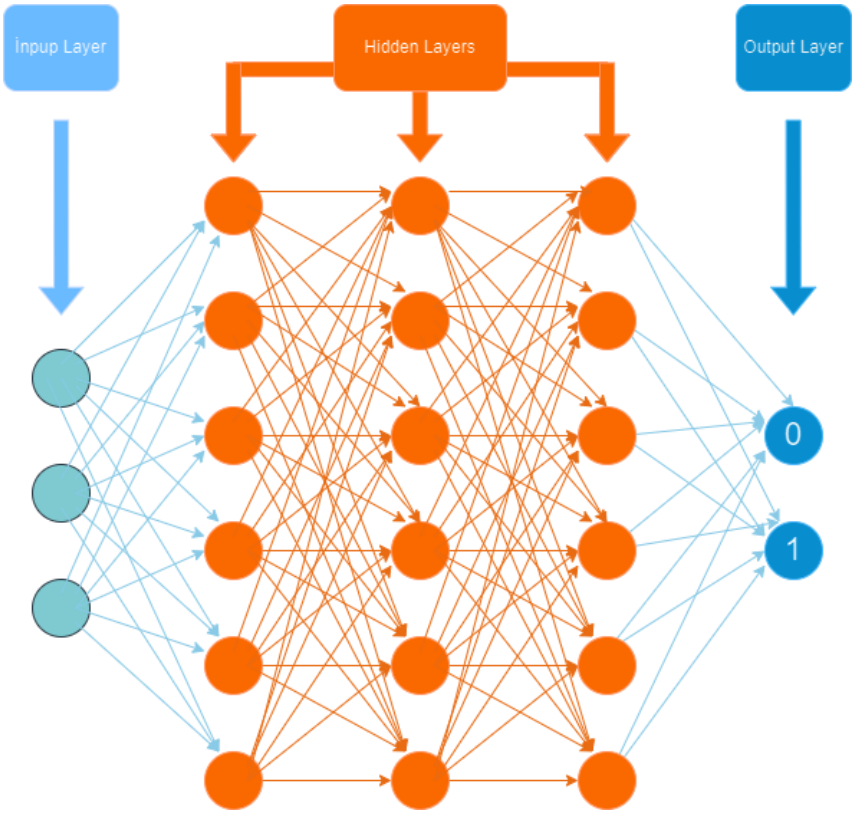

Fig. 1. A typical ANN structure

3. Results

In this study, we employed several key metrics: Accuracy, Precision, Recall, and F1-Score, to assess the performance of our machine learning algorithms and a custom-developed artificial neural network (ANN) model for breast cancer detection. Accuracy provides an overall indication of how well the model performs by measuring the proportion of correctly classified cases. Precision focuses on the model's ability to correctly identify positive cases without misclassifying negative ones, highlighting its effectiveness in minimizing false positives. Recall, or Sensitivity, evaluates how well the model identifies all actual positive cases, ensuring that true cases are not overlooked. The F1-Score balances Precision and Recall, offering a comprehensive view of the model's performance by accounting for both false positives and false negatives.

Using these metrics, we evaluated the performance of various machine learning algorithms and our ANN model, as detailed in Table 2. The dataset was carefully preprocessed to create optimal conditions for training and testing. Our analysis provides valuable insights into the effectiveness of each model in accurately detecting breast cancer, emphasizing their respective strengths and potential limitations. By examining these performance metrics, we aim to identify the most suitable models for clinical applications in breast cancer detection.

Table 2
Performance metrics of machine learning algorithms and ANN model for breast cancer detection

Model	Accuracy %	Precision%	Recall%	F1%
LR	94	95	93	94
GNB	96	97	95	96

BNB	59	30	50	59
MNB	88	91	86	88
SVM	91	91	91	91
DT	92	93	92	92
RF	96	97	95	96
KNN	91	93	89	91
GB	96	97	95	96
XGBoost	97	97	96	97
AdaBoost	97	97	96	97
LGBM	97	98	97	97
LDA	97	97	96	97
ANN	96	96	96	96

Based on the performance metrics from Table 2, the machine learning models and the ANN model for breast cancer detection show varied results. The top-performing models are XGBoost, AdaBoost, LGBM, and LDA, each achieving 97% accuracy. LGBM stands out with the highest precision at 98%. These models demonstrate balanced performance across accuracy, precision, recall, and F1 score, indicating their robustness and reliability in breast cancer detection. RF, GB, and GNB also perform well, each with 96% accuracy, making them strong candidates for clinical applications. On the other hand, BNB has the lowest performance, with only 59% accuracy and 30% precision, indicating it is less suitable for this task. Models like LR, SVM, DT, and KNN show moderate performance, with accuracy ranging from 88% to 94%. The ANN model achieves a high accuracy of 96%, but without detailed precision, recall, and F1 score metrics, its full effectiveness cannot be assessed. Overall, the study highlights the significant potential of advanced machine learning algorithms in improving the accuracy and efficiency of breast cancer diagnosis, with XGBoost, AdaBoost, LGBM, and LDA leading the way.

3.1 Performance of the ANN model with different hyperparameters

The article contains, a report on our project of breast cancer detection that is held by the technology of the ANN that is made in the PyTorch framwork. We experienced many different combinations of the hyperparameters and the results are given in the table below as illustrated in Table 3.

Table 3
Impact of Different Training Parameters and
Activation Functions on Neural Network Performance

Epoch	Test %	Layer	Neuron Number	Activation Function	Learning Rate	Accuracy %
100	20	3	128	Relu	0.01	90.35%
500	20	3	128	Relu	0.01	96.49%
1000	20	3	128	Relu	0.01	96.49%
1000	10	3	128	Relu	0.01	98.25%
1000	30	3	128	Relu	0.01	96.49%
1000	20	3	128	Relu	0.01	96.49%
1000	20	3	128	Relu	0.01	96.49%
1000	20	5	128	Relu	0.01	96.49%
1000	20	8	128	Relu	0.01	96.49%
1000	20	5	256	Relu	0.01	94.74%
1000	20	5	64	Relu	0.01	97.37%
1000	20	5	128	Relu	0.01	96.49%
1000	20	5	128	Relu	0.01	96.49%

1000	20	5	128	Leaky Relu	0.01	92.11%
1000	20	5	128	Sigmoid	0.01	96.49%
1000	20	5	128	Tanh	0.01	62.28%
1000	20	5	128	Leaky Relu	0.1	94.74%
1000	20	5	128	Leaky Relu	0.01	92.11%
1000	20	5	128	Leaky Relu	0.001	96.49%
1000	10	5	128	Relu	0.01	96.49%

Table 3 provides a detailed comparison of the performance of neural network models with various training parameters and activation functions. Among the models, the one with 100 epochs, 20% test data, 3 layers, 128 neurons per layer, and a learning rate of 0.01 using the ReLU activation function achieved the lowest accuracy of 90.35%. On the other hand, the highest accuracy of 98.25% was observed in the model with 1000 epochs, 10% test data, 3 layers, 128 neurons per layer, and the same learning rate and activation function. Several models achieved an accuracy of 96.49%, indicating consistency in performance across different configurations. For example, models with 500 and 1000 epochs, 20% test data, 3 layers, 128 neurons, and a learning rate of 0.01 using the ReLU activation function consistently hit this mark. Increasing the number of layers to 5 or even 8 did not significantly impact the accuracy, as long as the other parameters remained the same. Similarly, variations in the number of neurons per layer (such as 64 or 256) and using the Leaky ReLU or Sigmoid activation functions with the same learning rate still resulted in a comparable accuracy of around 94.74% to 97.37%. Interestingly, using the Tanh activation function with the same configuration led to a significant drop in accuracy to 62.28%, highlighting the importance of choosing the right activation function. Additionally, lowering the learning rate to 0.001 or increasing it to 0.1 while using the Leaky ReLU function resulted in varied performances, with accuracies of 96.49% and 94.74% respectively. This underscores the sensitivity of neural networks to learning rates and activation functions, suggesting that while certain configurations can yield high accuracies, others can lead to substantial performance drops. As a result, the model with the lowest performance achieved 90.35% accuracy, using 100 epochs, 20% test data, 3 layers, 128 neurons per layer, and a learning rate of 0.01 with the ReLU activation function. Conversely, the highest performance was achieved with 98.25% accuracy, using 1000 epochs, 10% test data, 3 layers, 128 neurons per layer, and the same learning rate and activation function.

4. Discussion

The experiments we conducted provide deep insights into the performance of neural networks, especially how they change based on different training parameters and activation functions. Our research showed that some configurations consistently produced high accuracy, demonstrating just how important it is to carefully tune these parameters. For example, when we trained our network for 1000 epochs using a 10% test data split, 128 neurons across three layers, and the ReLU activation function, we achieved an impressive accuracy of 98.25%. This configuration really stood out, showcasing the strength of ReLU in neural network training, as it performed better than other activation functions in similar settings.

Even when we adjusted the test data split to 20%, the model still maintained a strong accuracy of 96.49%. This suggests that the combination of extensive training and the ReLU function is particularly robust, regardless of the test data split. On the other hand, when we used the Tanh activation function, the accuracy dropped significantly to 62.28%. This indicates that Tanh may not be as effective for the specific dataset and conditions we tested. Similarly, while the Leaky ReLU function

sometimes performed better than Tanh, it generally didn't reach the high accuracy levels of standard ReLU, highlighting that the success of alternative activation functions can vary greatly depending on the context.

These findings bring to light the crucial role of hyperparameter optimization in neural network training. The choices you make about activation functions, the number of training epochs, and the structure of the network, including how many layers and neurons used, are all pivotal in determining how well your model performs. Our study emphasizes that even minor tweaks in these parameters can lead to significant differences in accuracy and overall performance.

In summary, while it's clear that you can achieve high levels of accuracy with the right configurations, our research underscores the need for a comprehensive and thoughtful approach to selecting these parameters. This isn't just about finding a quick solution; it's about understanding the underlying mechanisms that drive performance. Future research should continue to explore why certain activation functions work better than others and look into additional parameters that could further improve the training outcomes of neural networks. By diving deeper into these aspects, we can enhance the accuracy and efficiency of neural networks, ultimately leading to better applications in fields like medical diagnostics, where precision is critical. This kind of optimization not only pushes the boundaries of what neural networks can do but also has real-world implications, potentially leading to better patient outcomes and advancements in critical areas like oncology.

5. Conclusion

Breast cancer remains one of the leading causes of death among women and has long been a global health threat. Early detection is crucial as it significantly increases the chances of effective treatment and improves patient survival rates. In this study, we assessed 14 different machine-learning algorithms along with a custom-developed ANN model. Our findings revealed that the neural network's performance varied considerably based on different parameters and activation functions. Notably, the fine-tuned ANN model achieved the highest accuracy of 98.25% when using 128 neurons and the ReLU activation function, outperforming the other machine learning algorithms tested. These results underscore the importance of parameter optimization in developing high-performance neural network models for breast cancer detection. Furthermore, this study highlights the critical role of hyperparameter tuning in training neural networks for medical applications. Future research should explore the mechanisms behind the performance differences across various activation functions and investigate additional parameters to enhance neural network outcomes further. By applying these insights, the accuracy and efficiency of breast cancer diagnosis can be improved, leading to advancements in oncology and better patient care.

References

[1] L. R. Tao, Y. Ye, and H. Zhao, "Early breast cancer risk detection: a novel framework leveraging polygenic risk scores and machine learning," *J Med Genet*, vol. 60, no. 10, pp. 960–964, Oct. 2023, doi: 10.1136/JMG-2022-108582.

[2] D. Brinkley and J. L. Haybittle, "THE CURABILITY OF BREAST CANCER," *The Lancet*, vol. 306, no. 7925, pp. 95–97, Jul. 1975, doi: 10.1016/S0140-6736(75)90003-3.

[3] A. G. Waks and E. P. Winer, “Breast Cancer Treatment: A Review,” *JAMA*, vol. 321, no. 3, pp. 288–300, Jan. 2019, doi: 10.1001/JAMA.2018.19323.

[4] T. J. Key, P. K. Verkasalo, and E. Banks, “Epidemiology of breast cancer,” *Lancet Oncology*, vol. 2, no. 3, pp. 133–140, Mar. 2001, doi: 10.1016/S1470-2045(00)00254-0.

[5] J. Ferlay *et al.*, “Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012,” *Int J Cancer*, vol. 136, no. 5, pp. E359–E386, Mar. 2015, doi: 10.1002/ijc.29210.

[6] A. K. Dubey, U. Gupta, and S. Jain, “Breast cancer statistics and prediction methodology: A systematic review and analysis,” *Asian Pacific Journal of Cancer Prevention*, vol. 16, no. 10, pp. 4237–4245, 2015, doi: 10.7314/APJCP.2015.16.10.4237.

[7] A. K. Dubey, U. Gupta, and S. Jain, “A Survey on Breast Cancer Scenario and Prediction Strategy,” *Advances in Intelligent Systems and Computing*, vol. 327, pp. 367–375, 2015, doi: 10.1007/978-3-319-11933-5_40.

[8] T. M. Hanis, M. A. Islam, and K. I. Musa, “Diagnostic Accuracy of Machine Learning Models on Mammography in Breast Cancer Classification: A Meta-Analysis,” *Diagnostics*, vol. 12, no. 7, Jul. 2022, doi: 10.3390/diagnostics12071643.

[9] A. Turcu-Stiolica *et al.*, “Diagnostic Accuracy of Machine-Learning Models on Predicting Chemo-Brain in Breast Cancer Survivors Previously Treated with Chemotherapy: A Meta-Analysis,” *Int J Environ Res Public Health*, vol. 19, no. 24, p. 16832, Dec. 2022, doi: 10.3390/IJERPH192416832/S1.

[10] M. A. Naji, S. El Filali, K. Aarika, E. H. Benlahmar, R. A. Abdelouhahid, and O. Debauche, “Machine Learning Algorithms For Breast Cancer Prediction And Diagnosis,” *Procedia Comput Sci*, vol. 191, pp. 487–492, Jan. 2021, doi: 10.1016/J.PROCS.2021.07.062.

[11] D. Coşkun *et al.*, “A comparative study of YOLO models and a transformer-based YOLOv5 model for mass detection in mammograms,” *Turkish Journal of Electrical Engineering and Computer Sciences*, vol. 31, no. 7, pp. 1294–1313, Nov. 2023, doi: 10.55730/1300-0632.4048.

[12] I. Pacal, “MaxCerVixT: A novel lightweight vision transformer-based Approach for precise cervical cancer detection,” *Knowl Based Syst*, vol. 289, Apr. 2024, doi: 10.1016/j.knosys.2024.111482.

[13] M. Burukanli and N. Yumuşak, “COVID-19 virus mutation prediction with LSTM and attention mechanisms,” *Comput J*, Jul. 2024, doi: 10.1093/COMJNL/BXAE058.

[14] E. Aslan, “LSTM-ESA HİBRİT MODELİ İLE MR GÖRÜNTÜLERİNDEN BEYİN TÜMÖRÜNÜN SINIFLANDIRILMASI,” *Adıyaman Üniversitesi Mühendislik Bilimleri Dergisi*, vol. 11, no. 22, pp. 63–81, Apr. 2024, doi: 10.54365/adyumbd.1391157.

[15] E. Aslan and Y. Özüpak, “Classification of Blood Cells with Convolutional Neural Network Model,” *Bitlis Eren Üniversitesi Fen Bilimleri Dergisi*, vol. 13, no. 1, pp. 314–326, Mar. 2024, doi: 10.17798/bitlisfen.1401294.

[16] G. Işık and İ. Paçal, "Few-shot classification of ultrasound breast cancer images using meta-learning algorithms," *Neural Comput Appl*, Jul. 2024, doi: 10.1007/s00521-024-09767-y.

[17] I. Kunduracioglu and I. Pacal, "Advancements in deep learning for accurate classification of grape leaves and diagnosis of grape diseases," *Journal of Plant Diseases and Protection*, Mar. 2024, doi: 10.1007/s41348-024-00896-z.

[18] I. Pacal, "A novel Swin transformer approach utilizing residual multi-layer perceptron for diagnosing brain tumors in MRI images," *International Journal of Machine Learning and Cybernetics*, Mar. 2024, doi: 10.1007/s13042-024-02110-w.

[19] I. Pacal, M. Alaftekin, and F. D. Zengul, "Enhancing Skin Cancer Diagnosis Using Swin Transformer with Hybrid Shifted Window-Based Multi-head Self-attention and SwiGLU-Based MLP," *Journal of Imaging Informatics in Medicine*, Jun. 2024, doi: 10.1007/s10278-024-01140-8.

[20] O. Attallah, F. Anwar, N. M. Ghanem, and M. A. Ismail, "Histo-CADx: duo cascaded fusion stages for breast cancer diagnosis from histopathological images," *PeerJ Comput Sci*, vol. 7, pp. 1–38, Apr. 2021, doi: 10.7717/peerj-cs.493.

[21] D. A. Ragab, O. Attallah, M. Sharkas, J. Ren, and S. Marshall, "A framework for breast cancer classification using Multi-DCNNs," *Comput Biol Med*, vol. 131, Apr. 2021, doi: 10.1016/j.compbiomed.2021.104245.

[22] Z. Mushtaq, A. Yaqub, S. Sani, and A. Khalid, "Effective K-nearest neighbor classifications for Wisconsin breast cancer data sets," *Journal of the Chinese Institute of Engineers, Transactions of the Chinese Institute of Engineers,Series A*, vol. 43, no. 1, pp. 80–92, Jan. 2020, doi: 10.1080/02533839.2019.1676658.

[23] M. H. Alshayeji, H. Ellethy, S. Abed, and R. Gupta, "Computer-aided detection of breast cancer on the Wisconsin dataset: An artificial neural networks approach," *Biomed Signal Process Control*, vol. 71, Jan. 2022, doi: 10.1016/j.bspc.2021.103141.

[24] S. Aamir *et al.*, "Predicting Breast Cancer Leveraging Supervised Machine Learning Techniques," *Comput Math Methods Med*, vol. 2022, 2022, doi: 10.1155/2022/5869529.

[25] Y. F. Hernández-Julio, L. A. Díaz-Pertuz, M. J. Prieto-Guevara, M. A. Barrios-Barrios, and W. Nieto-Bernal, "Intelligent Fuzzy System to Predict the Wisconsin Breast Cancer Dataset," *International Journal of Environmental Research and Public Health 2023, Vol. 20, Page 5103*, vol. 20, no. 6, p. 5103, Mar. 2023, doi: 10.3390/IJERPH20065103.

[26] A. K. Dubey, U. Gupta, and S. Jain, "Analysis of k-means clustering approach on the breast cancer Wisconsin dataset," *Int J Comput Assist Radiol Surg*, vol. 11, no. 11, pp. 2033–2047, Nov. 2016, doi: 10.1007/s11548-016-1437-9.

[27] V. Nemade and V. Fegade, "Machine Learning Techniques for Breast Cancer Prediction," *Procedia Comput Sci*, vol. 218, pp. 1314–1320, Jan. 2023, doi: 10.1016/J.PROCS.2023.01.110.

[28] V. Chaurasia and S. Pal, "Applications of Machine Learning Techniques to Predict Diagnostic Breast Cancer," *SN Comput Sci*, vol. 1, no. 5, Sep. 2020, doi: 10.1007/s42979-020-00296-8.

[29] Z. Mushtaq, M. F. Qureshi, M. J. Abbass, and S. M. Q. Al-Fakih, "Effective kernel-principal component analysis based approach for wisconsin breast cancer diagnosis," *Electron Lett*, vol. 59, no. 2, p. e212706, Jan. 2023, doi: 10.1049/ELL2.12706.

[30] İ. PACAL, "Deep Learning Approaches for Classification of Breast Cancer in Ultrasound (US) Images," *Iğdır Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, vol. 12, no. 4, pp. 1917–1927, Dec. 2022, doi: 10.21597/jist.1183679.

[31] "Breast Cancer Wisconsin Diagnostic Dataset." Accessed: Jul. 17, 2024. [Online]. Available: https://www.kaggle.com/datasets/utkarshx27/breast-cancer-wisconsin-diagnostic-dataset

[32] L. K. Singh, M. Khanna, and R. Singh, "An enhanced soft-computing based strategy for efficient feature selection for timely breast cancer prediction: Wisconsin Diagnostic Breast Cancer dataset case," *Multimed Tools Appl*, pp. 1–66, Feb. 2024, doi: 10.1007/S11042-024-18473-9/TABLES/26.

[33] M. P. LaValley, "Logistic regression," *Circulation*, vol. 117, no. 18, pp. 2395–2399, May 2008, doi: 10.1161/CIRCULATIONAHA.106.682658/ASSET/472996AA-9EBA-4322-A948-192E511AC922/ASSETS/GRAPHIC/19FF1.JPEG.

[34] E. M. K. Reddy, A. Gurrala, V. B. Hasitha, and K. V. R. Kumar, "Introduction to Naive Bayes and a Review on Its Subtypes with Applications," *Bayesian Reasoning and Gaussian Processes for Machine Learning Applications*, pp. 1–14, Apr. 2022, doi: 10.1201/9781003164265-1.

[35] W. S. Noble, "What is a support vector machine?," 2006. [Online]. Available: http://www.nature.com/naturebiotechnology

[36] C. Kingsford and S. L. Salzberg, "What are decision trees?," *Nat Biotechnol*, vol. 26, no. 9, p. 1011, Sep. 2008, doi: 10.1038/NBT0908-1011.

[37] M. Belgiu and L. Drăgu, "Random forest in remote sensing: A review of applications and future directions," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 114, pp. 24–31, Apr. 2016, doi: 10.1016/J.ISPRSJPRS.2016.01.011.

[38] P. Cunningham and S. J. Delany, "K-Nearest Neighbour Classifiers-A Tutorial," Jul. 01, 2021, *Association for Computing Machinery*. doi: 10.1145/3459665.

[39] A. Natekin and A. Knoll, "Gradient boosting machines, a tutorial," *Front Neurorobot*, vol. 7, no. DEC, p. 63623, Dec. 2013, doi: 10.3389/FNBOT.2013.00021/BIBTEX.

[40] S. S. Azmi and S. Baliga, "An Overview of Boosting Decision Tree Algorithms utilizing AdaBoost and XGBoost Boosting strategies," *International Research Journal of Engineering and Technology*, 2020, [Online]. Available: www.irjet.net

[41] X. Li, L. Wang, and E. Sung, "AdaBoost with SVM-based component classifiers," *Eng Appl Artif Intell*, vol. 21, no. 5, pp. 785–795, Aug. 2008, doi: 10.1016/J.ENGAPPAI.2007.07.001.

[42] C. J. Huberty, “Discriminant Analysis,” *Rev Educ Res*, vol. 45, no. 4, pp. 543–598, Dec. 1975, doi: 10.3102/00346543045004543/ASSET/00346543045004543.FP.PNG_V03.

[43] M. Gevrey, I. Dimopoulos, and S. Lek, “Review and comparison of methods to study the contribution of variables in artificial neural network models,” *Ecol Modell*, vol. 160, no. 3, pp. 249–264, Feb. 2003, doi: 10.1016/S0304-3800(02)00257-0.

[44] A. Krogh, “What are artificial neural networks?,” 2008. [Online]. Available: http://www.r-project.org/