A Comparison between Elastic Net Logistic and a Set of Machine Learning Algorithms in Predicting Breast Cancer

Hadeel Imad Naser[1,*], Wakaa Ali Hadba[1]

[1]Teaching, Department of Statistics, College of Administration and Economics, University of Kirkuk, Iraq

Emails: hadeel.imad@uokirkuk.edu.iq; Wakaa2017@uokirkuk.edu.iq

Abstract

Breast cancer is a common type of cancers and the main reason of increased death of women universally. Recently, ML methods have become important in varying fields, such as Logistic Regression, Elastic Net Logistic, Decision Tree, Random Forest, Boosting, Naive Bayes and K Nearest Neighbor. The aim of the current study is to know and predict the type of cancerous tumor whether it is benign or malignant. These above techniques are expected to be helpful. Breast tumor type diagnosis using numerous performance metrics i.e. accuracy, classification error, sensitivity and specificity, both certified and trained models were assessed. The models were developed to determine which model would provide the best performance and comparisons were done. A separate data set from the one used to create the models was utilized to confirm every model. According to the analysis, the findings showed that elastic net logistic model had the highest performance in accurate classification rate (accuracy), classification error and sensitivity. Making it the best classifier for predicting the kind of breast cancer among all other models, privacy and it was also distinguished by reduce the high dimensionality and multicollinearity problems.

Keywords: Logistic Regression; Elastic Net Logistic; Decision Tree; Random Forest; Boosting; Naive Bayes; and K- Nearest Neighbor

1. Introduction

Cancer cells are of two types (benign or malignant). In general, cells are of ability to divide and disseminate different body parts and lead to novel tumors. Which are also considered the second main cause of the rise in deaths among females, which has a greater impact on women than on men. There are several causes of breast cancer, including: It might undoubtedly be heritable (i.e. a family member or close relative is affected), aging (since infection increases over time), postmenopausal obesity or gaining weight, alcohol or smoking, or other factors. Early diagnosis and treatment rises the chance of survival from the disease, which raises the probability of survival [1].

Several studies have investigated breast cancer using varying ML algorithms, which classified the type of tumor and predict it, assisting in the treatment of the disease and reducing its prevalence.

In (2017) Murugan et al. used linear regression methods, decision tree and randomizing forest to classify and expect breast cancer tumors on a set of data attained out of the UCI ML Repository (Wisconsin Breast Cancer), the ratting of success for classification was 84.14% Which was taken out of by linear regression and a prediction rate was 88.14% Which was obtained by random forest [2].

In 2019, DT and KNN algorithms were employed for classifying breast cancer kinds by Rajaguru and Sannasi Chakravarthy. The algorithms were used in the Wisconsin diagnostic breast cancer dataset. They used the principal components analysis to select features and compared the methods using a standard performance matrices. They concluded that the KNN classifier performed better than the DT classifier in classification [3].

DOI: https://doi.org/10.54216/GJMSA.130101

Received: October 12, 2025 Revised: December 04, 2025 Accepted: January 20, 2026

In 2020, Islam et al. compared several ML algorithms (SVM, K-NN,RF, ANN, and LR) with each other using data taken out of the UCI ML database. The performance comparison was centered upon numerous measures, like (sensitivity, accurateness, specificity, negative predictive value, false negative rates, false positive ,F1 score, precision and Matthews correlation coefficient).The comparison results were as follows: the accurateness, precision and F1 score for the ANN were (98.57%, 97.82%, and 0.9890) while (97.14%, 95.65%, and 0.9777) for SVM, so ANN is the best classifier [4].

In the same year, Assegie employed grid search to discover the optimal hyper parameters and proposed the optimized (KNN) model to obtain accuracy in detecting breast cancer. It was used on data obtained out of the (Wisconsin) breast cancer dataset taken out of (Kaggle). After comparing the performance of the (KNN) with tuned hyper parameters and with default hyper parameters, it was found that the accurateness of the suggested enhanced model was (94.35%) more than the default model, which was (90.10%) in diagnosis [5].

Al-harthi et al. (2021) employ the adjusted variances of the features, which in turn act as weights within the L1_ regularization when regularizing the elastic net model to address the inconsistency problem used regularized logistic regression with an adaptive elastic net RLRAEN. They also applied their proposed method to the Wisconsin breast cancer dataset from the UCI repository, and compared it with other penalized ways. The researchers concluded that the method (RLRAEN) is more efficient [6].

In the same year, Assegie et al provided a model for breast cancer prediction by DT and adaptive boosting (Adaboost) on a breast cancer dataset obtained out of (Kaggle) data which consisted of (569 samples, of which 37.25% were benign tumors and 62.74% were malignant tumors). The DT algorithm showed a bias towards benign observations , the researchers found that adaptive boosting was more accurate than DT, with the percentages reaching (92.53% to 88.80%) [7].

In 2022, Nasser and Behadili used DT and KNN algorithms, with feature selection, in DT where the (Gini index) provided accurateness of (87.83%) while with (entropy) the accuracy was (86.77%) in these cases age showed greater effectiveness especially when the age is less than 45.5 while (ki67) appeared the second in effective. KNN was with a rate of 86.24% of accuracy. in the end, the researchers concluded that stage II breast cancer was the most common [8].

In 2023, Bokhare and Jha compared and evaluated ML models, like (KNN, DT, SVM, RF, SVM Kernels, LR , and Naive Bays models) using the accuracy criterion, and concluded that the DT classifier was the most accurate, reaching (97.08%) more than others [9].

In 2024, Ahmed et al compared machine learning algorithms, including (KNN, RF, DT, (CART), SVM and naïve bayes) for diagnosing breast cancer type as well as early prediction of the disease. After analysis, it was concluded that the SVM algorithm outclassed further algorithms for accuracy, data utilization, and precision (Ahmed et al., 2024).

In 2025, Alwazy et al. used statistical learning and machine learning algorithms to diagnose and classify breast, heart, prostate and lung diseases. The algorithms (SVC, RF, DT, XGBoost, ridge and elastic net and lasso) were used on four medical datasets. The researchers compared the algorithms with each other and evaluated the performance based on the measures of accuracy, precision, f1 score, recall and area under the curve. They reached the SVC and RF algorithms that had a maximal accurateness of up to 98% and a score under the curve of 1 in distinguishing between malignant and benign breast cancer samples [11].

The aim of this search is to classify and predict breast cancer tumors by algorithms logistic regression, elastic net, decision tree, random forest, naïve bays and k- nearest neighbor.

This paper divided in 4 section, first included introducing part and reviewing literatures for breast cancer, 2nd section include used methodology, the 3rd presents collecting data and findings and discussion and lastly was the conclusion in the 4th section.

2. Materials and Methods

2.1 Logistic Regression (LR):

Logistic regression (LR) model which also called the logit model, studies the relationship among a definite dependent and many independent variables, which might be continuous, discrete, or qualitative data. If the dependent variable has only two values, then it called binary logistic regression model. And if it is of more than two values, then it is called a multinomial logistic regression model. The logistic regression model is:

$$\log\left(\frac{p}{1-p}\right) = b_0 + b_1x_1 + b_2x_2 + \cdots b_ix_i \qquad (1)$$

DOI: https://doi.org/10.54216/GJMSA.130101
Received: October 12, 2025 Revised: December 04, 2025 Accepted: January 20, 2026

where p points out the event probability, xi are explanatory variables and ($b_0, b_1, .., b_i$,i=1,..,n) are the regression coefficients associated with the reference team [12] ,[13].

Equation (1) can be written in another form:

$$\hat{y}_i = \frac{1}{1+e^{-(b_0+b_1x_1+b_2x_2+\cdots b_ix_i)}} \qquad (2)$$

Where $\hat{y}_i$ indicates to the estimated continuous chance, e (natural logarithm) and $b_0 + bx_1 + b_2x_2 + \cdots bx_i$ stands for the equation of linear regression for independent variables represented in the logit scale [14] .

2.2 Elastic Net Logistic:

A logistic model may contain a large number of variables that may be penalized for neglecting the coefficients of non-significant variables or those close to zero, in this case, the model called regularization model. Regularization's process is important in fitting regression models with a greater amount of variables, as this process can produce models with a greater amount of parameters which in turn makes the estimation process difficult when the amount of features is maximal. The most widely used penalized regression includes [15] :

1. Ridge regression: in cases where the contribution of the variables is small and their coefficients are close to 0, the model will take all variables, as it is beneficial to include all variables in the model It can be expressed by the equation:

$$L_{ridge}(\hat{b}) = \sum_{i=1}^{n}(y_i - x_i'\hat{b})^2 + \lambda\sum_{j=1}^{m}\hat{b}_j^2 = \| y - X\hat{b} \|^2 + \lambda \left|\left|\hat{b}\right|\right|^2 \qquad (3)$$

2. Lasso regression: this model includes the most important variables while the coefficient of the less important variables are equal to 0. It can be expressed by the equation:

$$L_{Lasso}(\hat{b}) = \sum_{i=1}^{n}(y_i - x_i'\hat{b})^2 + \lambda\sum_{j=1}^{m}|\hat{b}_j| \qquad (4)$$

So elastic net regression combines lasso and ridge regression.

From equation (3) and (4) ,can be obtained elastic net equation: [16],[15].

$$L_{Elastic}(\hat{b}) = \left[\frac{1}{2n}\sum_{i=1}^{n}(y_i - x_i'\hat{b})^2 + \lambda\left(\frac{1-\alpha}{2}\sum_{j=1}^{m}\hat{b}_j^2 + \alpha\sum_{j=1}^{m}|\hat{b}_j|\right)\right] \qquad (5)$$

2.3 Decision Tree (DT):

DT is commonly employed algorithms in supervised ML and is applicable to both classification and regression. [17]. DT consists of many nodes including the root node which is the structure of the tree. All nodes are connected to each other by edges, except for the root node. each node has one incoming edge. Some nodes, known as internal nodes or test nodes, have one or more outgoing edges. the instance space of each internal node is divided to dual or further subspaces relative to the discrete functions of the input values [8]. The most advantage for DT algorithm is estimation, prediction and classification [9] .

2.4 Random Forest (RF):

It is considered one of the new techniques in machine learning and its principle is the random selection of the feature in the decision tree training process. RF is used to solve classification and prediction problems as well as reduce dimensions and detect abnormal patterns [18],(Ahmed et al., 2022). The basis of the RF algorithm is bagging, which can be considered a special case of it, where the algorithm consists of a set of m tree classifiers denoted as $H_1(x)$, $H_2(x)$,..., $H_K(x)$. In addition to bagging, there is another source for randomness the forest ensemble. A learning algorithm is haphazardly chosen at every node of the tree, and a subset of m features is evaluated. The clustered classifier can then be given the symbol H(x) which refer to the aggregated classifier , Then we can write the classification margin function which specifies the extent whereby the proportion for the accurate class y may exceed the maximum obtained by class c than y as: [20] .

$$\text{margin}\,(x,y) = p(H(x) = y) - \max_{i=1,i\neq y}^{c} c\ p(H(x) = i) \qquad (6)$$

2.5 Boosting:

It is an alternative method for building models. It works on fitting models to a new set of data by taking samples from the original set. we set the residuals at $r_i=y_i$ and the regression function at $f_0(x)=0$ for all observation. The algorithm then repeats these actions steps for a=1,...,A:

1. A tree model is fitted to the responses r and the features x, then the regression function of this tree is g_a.

DOI: https://doi.org/10.54216/GJMSA.130101
Received: October 12, 2025 Revised: December 04, 2025 Accepted: January 20, 2026

2. put $f_a(x) = f_{a-1}(x) + \lambda g_a(x)$.

3. Then calculate the new residuals from $r_i = r_i - \lambda g_a(x_i)$

Where $f_A(x) = \lambda \sum_{a=1}^{A} g_a(x)$ refers to the final fitted model [15].

2.6 Naïve Bayes:

It is a classifier that classification a statistical model for calculating the chances of a class containing each set of existing attributes, and from this, determines the most ideal class. All attributes are independent of each other, and they all contribute to decision-making [21]. Bayes' theorem is represented in the below equation:

$$P(b_j \mid a_i) = \frac{P(a_i \mid b_j) \times P(b_j)}{P(a_i)} \qquad (7)$$

where $P(a_i)$ and $P(b_j)$ are event probability A and B with no considering each other. $P(b_j \mid a_i)$ is the probability of b_j conditional on a_i and $P(a_i \mid b_j)$ is of a_i conditional on b_j , B is represent the classification variable and A is a tuple of attribute values, (In naïve Bayes classification). When calculating the numerator value for each class and then choosing the class at which the value is the maximum, this rule is called as the maximal a posteriori (MAP) rule. The resulting class, calculated as y for the instance x, is determined using this simple naive Bayes classifier as: [22] .

$$\hat{y} = \arg\max_{b_j} \prod_{i=1}^{n} P(a_i \mid B = b_j) P(B = b_j) \qquad (8)$$

2.7 K Nearest Neighbor (KNN):

It is a non-parametric algorithm which is usually employed to solve the classification problem, the information of points which are adjacent to each other is customs to classify the output labels under this approach [3] .

Algorithm K-NN steps can be written as:

1. input the data set , split it into training and testing sets.

2. select instances out of the test sets, then measure how faraway it is from the training set.

3. Listing the distances in ascending order.

4. The instance's category is the utmost particular one of the first triple training instances (k=3) [23] .

Performance evaluation of models:

To evaluate performance of algorithms, using a confusion matrix which made up of TP, FP, TN, and FN for real data and predicted data, where:

TP: represents the number of cases classified as True which are in fact True.

FN: represents the number of cases classified as False which are in fact True.

TN: represents the number of cases classified as False which are in fact False.

FP: represents the number of cases classified as True which are False [4].

(SE) Sensitivity: represents the probability value that the expected classification will be accurate for the accurate case which is : $SE = \frac{TP}{TP+FN} = \frac{TP}{P}$

(SP) Specificity: represents the value of the probability that the expected classification will be incorrect for the case that is incorrect which is : $SP = \frac{TN}{FP+TN} = \frac{TN}{P'}$

(AC) Accuracy: represents the ratio of correctly classified and calculated as : $AC = \frac{TP+TN}{TP+TN+FP+FN}$.
Another important measure employed to evaluate effectiveness of the classifier is the ROC curve which plots 1- specificity on the x-axis and sensitivity on the y-axis, giving the zone under the curve which ranges through 0 to 1,which is a measure that measures the model's capability of distinguishing among cases [24],[25].

DOI: https://doi.org/10.54216/GJMSA.130101
Received: October 12, 2025 Revised: December 04, 2025 Accepted: January 20, 2026

3. Results and Discussion

3.1 About data:

Cancer is a disease affecting females more than males. In 2015, more than 2.1 million people were diagnosed with the disease, representing 25% of all cancer cases. The disease begins as a tumor in breast cells, growing until they form lumps and begin to disseminate to further organs of the body. The main goal in diagnosing such tumors and distinguishing between malignant and benign ones. A Fine Needle Aspiration (FNA) of a breast lump is used to analyze the picture and assess the features. This is defined meticulously in: [K. P. Bennett and O. L. Mangasarian]: "Robust Linear Programming Discrimination of Two Linearly Inseparable Sets", Optimization Methods and Software 1, 1992, 23-34.This database can be found kaggle https://www.kaggle.com/datasets/uciml/breast-cancer-wisconsin-data

3.2 Attribute Information and results:

Diagnosing (M = malignant, B = benign) is the dependent variable, the predictive variables are (30), the entire attribute values are recorded with four significant digits, and no missing values. The data consisted of a No. of instances: 569 and a No. of attributes: 32 (ID, diagnosis, 30 real-valued input features) Class distribution: 357 benign, 212 malignant, the percentage of B is (0.63) and (0.37) for M. For the purpose of analysis, the data fell into a percentage (0.8) for training and (0.2) for testing, the numeral amount of views for training (456) and testing (113).

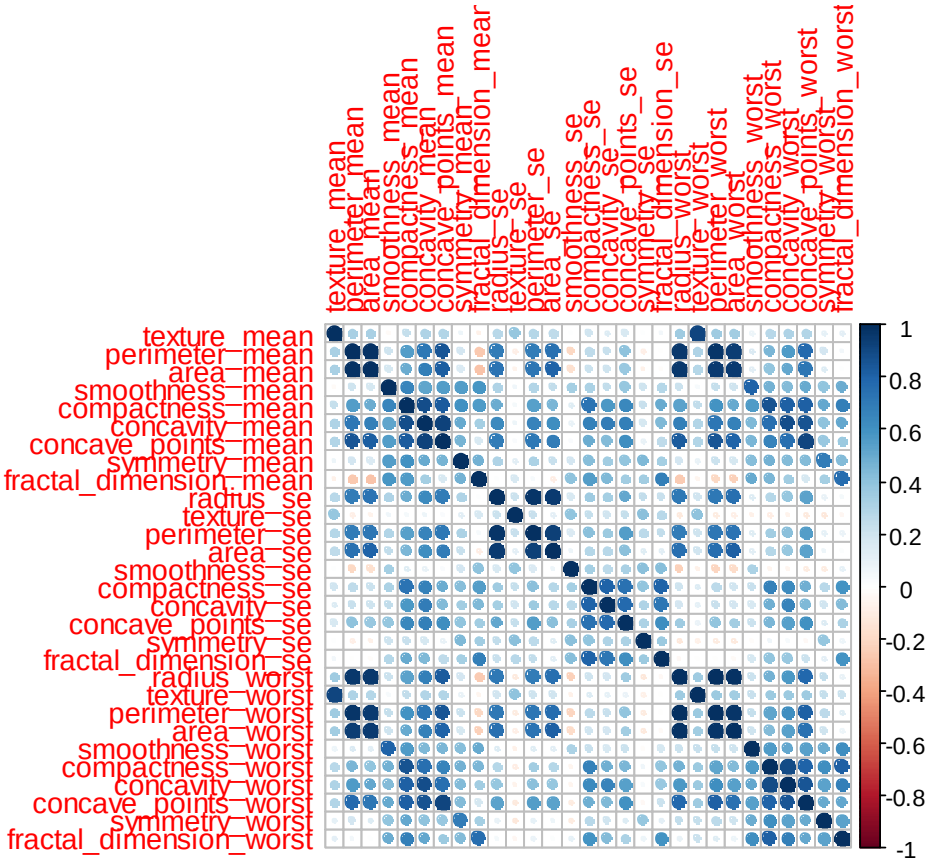

Figure 1. Correlation coefficient between variables and the blue color shows the strength of this coefficient between variables.

DOI: https://doi.org/10.54216/GJMSA.130101
Received: October 12, 2025 Revised: December 04, 2025 Accepted: January 20, 2026

Table 1: Evaluation metrics to compare model performance (algorithmic results for some classification criteria).

	Accuracy	**classification error rate**	**Sensitivity**	**Specificity**
Random Forest	0.9735	0.0265	0.9577	1
Boosting	0.6726	0.3274	0.7465	0.5476
Decision Tree	0.9292	0.0708	0.9014	0.9762
k-Nearest Neighbors	0.9735	0.0265	0.9859	0.9524
Naive Bayes	0.5664	0.4336.	0.9155	0.9155
Logistic Regression	0.9646	0.0354	1	0.9048
Elastic Net Logistic	0.9912	0.0088	1	0.9859

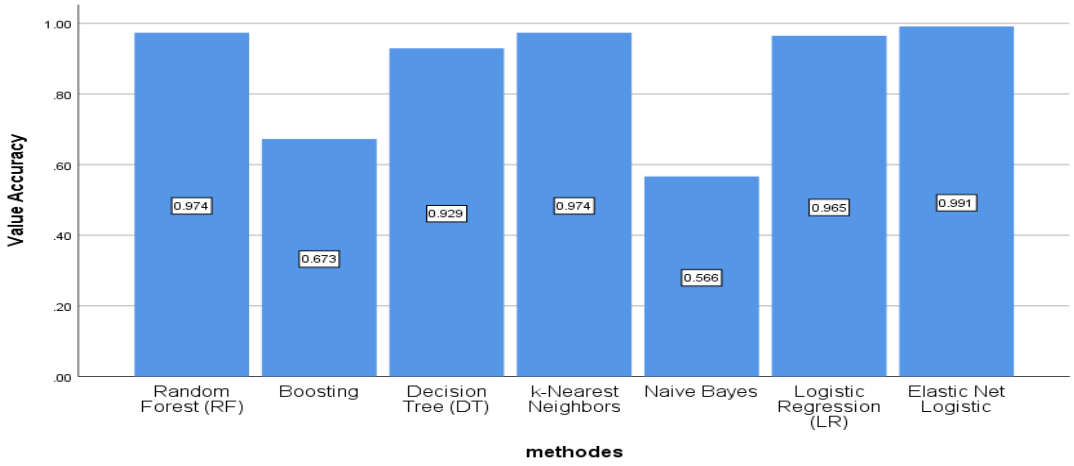

Figure 2. Compare methods using Accuracy criteria

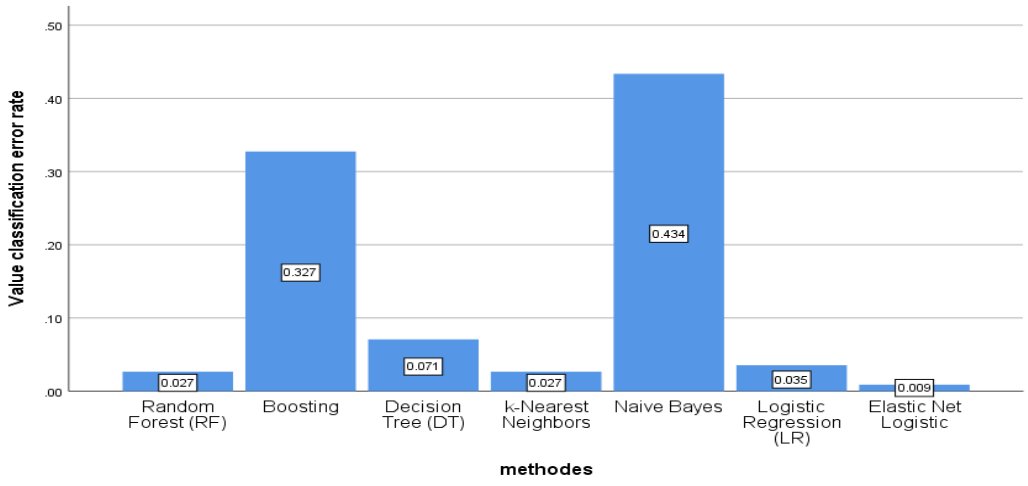

Figure 3. Compare methods using classification error rate criteria

DOI: https://doi.org/10.54216/GJMSA.130101
Received: October 12, 2025 Revised: December 04, 2025 Accepted: January 20, 2026

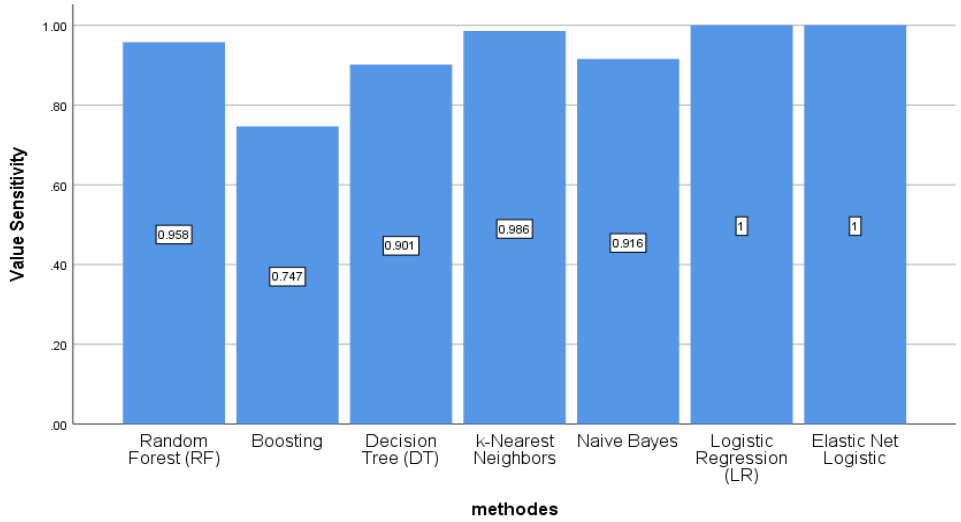

Figure 4. Compare methods using Sensitivity (SE) criteria.

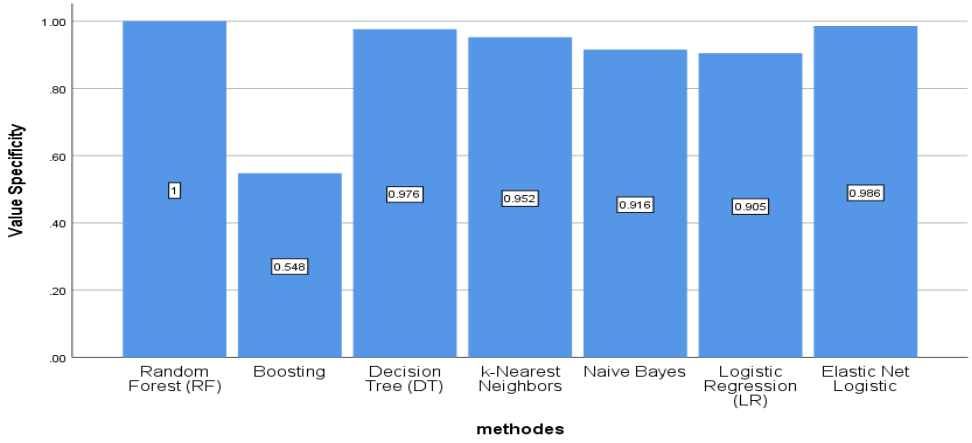

Figure 5. Compare methods using Specificity (SP) criteria.

Logistic Regression, Elastic net Logistic, Decision Tree, Random Forest, Boosting, Naïve Bayes algorithms have broad tasks in classification and prediction. The performance of such algorithms was in comparison by employing a validation dataset in order to verify their performance after adequate training, it was conducted to evaluate their ability to predict future samples. In the current study, the Elastic net logistic model outperformed the rest of the models in breast cancer diagnosis by achieving the highest results, as its accuracy reached (%99.12) with a classification error (0.0088%) sensitivity (1%), Specificity (0.9859).

4. Conclusion

In the present research, the performance of several ML methods namely Logistic Regression, Elastic Net Logistic, Decision Tree, Random Forest, Boosting, Naïve Bayes, and K Nearest Neighbor was compared in diagnosing breast tumors, whether they are benign or malignant, on Wisconsin breast cancer data taken out of UCI ML repository or from Kaggle. The comparison was made on the basis of being sensitive, specific, precise, and classifying of error criteria. The findings uncovered that the algorithm Elastic Net Logistic outperformed the logistic regression algorithm in terms of error classification, sensitivity, specificity, and accuracy on testing data, which makes the model important in diagnosing breast tumors. It also features the ability to reduce variables close to zero, address multicollinearity, and reduce dimensionality issues, as it is a model that combines the Lasso and Ridge methods. Finally, we can recommend using regularization techniques such as Lasso and Ridge and using learning algorithms to compare and classify.

DOI: https://doi.org/10.54216/GJMSA.130101
Received: October 12, 2025 Revised: December 04, 2025 Accepted: January 20, 2026

References

[1] S. Osei-Afriyie, A. K. Addae, S. Oppong, H. Amu, E. Ampofo, and E. Osei, "Breast cancer awareness, risk factors and screening practices among future health professionals in Ghana: A cross-sectional study," *PLoS One*, vol. 16, no. 6, pp. 1–17, 2021, doi: 10.1371/journal.pone.0253373.

[2] S. Murgan, B. Muthu Kumar, and S. Amudha, "Classification and Prediction of Breast Cancer using Linear Regression, Decision Tree and Random Forest," in *2017 International Conference on Current Trends in Computer, Electrical, Electronics and Communication (CTCEEC)*, 2017, pp. 763–766, doi: 10.1109/CTCEEC.2017.8455058.

[3] H. Rajaguru and S. R. Sannasi Chakravarthy, "Analysis of decision tree and k-nearest neighbor algorithm in the classification of breast cancer," *Asian Pacific Journal of Cancer Prevention*, vol. 20, no. 12, pp. 3777–3781, 2019, doi: 10.31557/APJCP.2019.20.12.3777.

[4] 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, no. 5, Sep. 2020, doi: 10.1007/s42979-020-00305-w.

[5] T. A. Assegie, "An optimized K-Nearest neighbor based breast cancer detection," *Journal of Robotics and Control (JRC)*, vol. 2, no. 3, pp. 115–118, May 2020, doi: 10.18196/jrc.2363.

[6] M. Alharthi, M. H. Lee, and Z. Y. Algamal, "Improving the diagnosis of breast cancer using regularized logistic regression with adaptive elastic net," *Universal Journal of Public Health*, vol. 9, no. 5, pp. 317–323, Oct. 2021, doi: 10.13189/ujph.2021.090514.

[7] T. A. Assegie, R. L. Tulasi, and N. K. Kumar, "Breast cancer prediction model with decision tree and adaptive boosting," *IAES International Journal of Artificial Intelligence*, vol. 10, no. 1, pp. 184–190, 2021, doi: 10.11591/ijai.v10.i1.pp184-190.

[8] F. K. Nasser and S. F. Behadili, "Breast Cancer Detection using Decision Tree and K-Nearest Neighbour Classifiers," *Iraqi Journal of Science*, vol. 63, no. 11, pp. 4987–5003, 2022, doi: 10.24996/ijs.2022.63.11.34.

[9] Bokhare and P. Jha, "Machine learning models applied in analyzing breast cancer classification accuracy," *IAES International Journal of Artificial Intelligence*, vol. 12, no. 3, pp. 1370–1377, Sep. 2023, doi: 10.11591/ijai.v12.i3.pp1370-1377.

[10] S. S. Ahmed, Y. Srivastava, and M. G. Khan, "Prediction and Diagnosis of Breast Cancer Using Machine Learning Algorithms," *Asian Journal of Research in Medical and Pharmaceutical Sciences*, vol. 13, no. 3, pp. 54–60, 2024, Art. no. AJRIMPS.117744.

[11] S. H. Alwazy, G. Buyrukoğlu, S. Buyrukoğlu, and M. R. Baker, "Evaluating machine learning and statistical learning techniques for cancer classification and diagnosis," *Iranian Journal of Computer Science*, 2025, doi: 10.1007/s42044-025-00233-z.

[12] L. I. Kework and S. A. Jalal, "Using Multinomial Logistic Regression Model to predict the most Important Factors Affecting in the students' selection for Specialization in Secondary Level for Some Schools Erbil City," *University of Kirkuk Journal of Administrative and Economic Sciences*, vol. 13, no. 2, pp. 1–22, 2023.

[13] S. Sperandei, "Understanding logistic regression analysis," *Biochemia Medica*, vol. 24, no. 1, pp. 12–18, 2014, doi: 10.11613/BM.2014.003.

[14] J. C. Stoltzfus, "Logistic regression: A brief primer," *Academic Emergency Medicine*, vol. 18, no. 10, pp. 1099–1104, Oct. 2011, doi: 10.1111/j.1553-2712.2011.01185.x.

[15] Diana, J. E. Griffin, J. Oberoi, and J. Yao, "Machine-Learning Methods for Insurance Applications – A Survey," 2019.

[16] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 67, no. 2, pp. 301–320, 2005.

[17] D. Shree and J. M. Lincy, "Prediction of chronic kidney disease with supervised machine learning algorithms," *International Journal of Creative Research Thoughts*, vol. 11, no. 3, pp. 2320–2882, 2023.

[18] H. Shi, S. Liu, J. Chen, X. Li, Q. Ma, and B. Yu, "Predicting drug-target interactions using Lasso with random forest based on evolutionary information and chemical structure," *Genomics*, vol. 111, no. 6, pp. 1839–1852, Dec. 2019, doi: 10.1016/j.ygeno.2018.12.007.

[19] H. Ahmed, M. N. A. Al-Hamadani, and I. A. Satam, "Prediction of COVID-19 disease severity using machine learning techniques," *Bulletin of Electrical Engineering and Informatics*, vol. 11, no. 2, pp. 1069–1074, 2022, doi: 10.11591/eei.v11i2.3272.

[20] D. Zhang and J. J.-P. Tsai, *Advances in Machine Learning Applications in Software Engineering*. Idea Group Publishing, 2007.

[21] Parlina et al., "Naive Bayes Algorithm Analysis to Determine the Percentage Level of visitors the Most Dominant Zoo Visit by Age Category," in *Journal of Physics: Conference Series*, vol. 1255, no. 1, Art. no. 012031, Sep. 2019, doi: 10.1088/1742-6596/1255/1/012031.

DOI: https://doi.org/10.54216/GJMSA.130101
Received: October 12, 2025 Revised: December 04, 2025 Accepted: January 20, 2026

[22] D. Berrar, "Bayes' theorem and naive bayes classifier," in *Encyclopedia of Bioinformatics and Computational Biology*, vol. 1, Elsevier, 2018, pp. 403–412, doi: 10.1016/B978-0-12-809633-8.20473-1.

[23] S. F. Khorshid and A. M. Abdulazeez, "Breast Cancer Diagnosis Based on K-Nearest Neighbors: A Review," *Journal of Applied Science and Technology Trends*, vol. 2, no. 01, pp. 1927–1951, 2021, ISSN: 1567-214X.

[24] Korotcov, V. Tkachenko, D. P. Russo, and S. Ekins, "Comparison of Deep Learning with Multiple Machine Learning Methods and Metrics Using Diverse Drug Discovery Data Sets," *Molecular Pharmaceutics*, vol. 14, no. 12, pp. 4462–4475, Dec. 2017, doi: 10.1021/acs.molpharmaceut.7b00578.

[25] F. H. Anad, "Comparison between logistic regression model and Vector machine to classify observations of COVID-19 patients at Al-Hussein General Hospital," *University of Kirkuk Journal of Administrative and Economic Sciences*, vol. 14, no. 3, pp. 261–273, 2024.

DOI: https://doi.org/10.54216/GJMSA.130101
Received: October 12, 2025 Revised: December 04, 2025 Accepted: January 20, 2026