Frontiers in Health InformaticsISSN-Online: 2676-7104 www.healthinformaticsjournal.com
2025; Vol 14: Issue 1 Open Access

Detection of Breast Cancer Using Machine Learning Algorithms: A Study on Logistic Regression, K-Nearest Neighbors, and Decision Trees

Yoo-Shin Park[1], Prof. Tae Hoon Kim[2] and Prof. Byung-Chan Min[3]
[1]Department of Industrial & Management Engineering, Hanbat National University, 34158, Daejeon, Republic of Korea
[2]School of Information and Electronic Engineering and Zhejiang Key Laboratory of Biomedical Intelligent Computing Technology, Zhejiang University of Science and Technology, No. 318, Hangzhou, Zhejiang, China
[3]Hanbat National University, Department of Industrial & Management Engineering
125, Dongseo-daero, Yuseong-gu, 34158, Daejeon, Republic of Korea
[1]winpys@hanmail.net, [2]323020@zust.edu.cn and [3]bcmin@hanbat.ac.kr
Corresponding Author: Prof. Byung-Chan Min

Cite this paper as: Yoo-Shin Park, Tae Hoon Kim and Byung-Chan Min (2025). Detection of Breast Cancer Using Machine Learning Algorithms: A Study on Logistic Regression, K-Nearest Neighbors, and Decision Trees. *Frontiers in Health Informatics,* 14(1), 1-11.

ABSTRACT

Background
Breast cancer is a malignant neoplasm that originates from breast cells. It is a major issue that affects both developed and developing nations. Prompt identification considerably increases the chance of effective therapy. Regular tests and public awareness campaigns are very helpful in the early identification of breast cancer. Machine learning offers potential in breast cancer prediction via the identification of hidden patterns in data. Several ML prediction models have been studied to accurately predict breast cancer.

Methods
In this study, three classifiers LR, KNN and DT were examined to detect malignant and benign samples in the Wisconsin Breast Cancer Dataset (WBCD). Model performance is evaluated using evaluation measures including recall, F1-score, accuracy, precision, and error rate.

Results
Early breast cancer detection is facilitated by the combination of many risk factors. It facilitates the development of necessary care plans. Managing, storing, and gathering a variety of data is essential. Multi-factor intelligent systems improve breast cancer prediction. They work well in the treatment of disease. The Logistic Regression model had the highest accuracy of 99.12% and demonstrated strong predictive capabilities. K-Nearest Neighbors (KNN) had slightly lower accuracy at 96.49%. It still shows promising results in classifying breast cancer instances. Decision Tree, with an accuracy of 92.98%, showed competitive precision, recall, and F1-score metrics for both classes. It indicates strong performance in classifying breast cancer instances based on provided features.

Keywords: Breast Cancer, Malignant, Machine Learning, Accuracy

1. INTRODUCTION

Breast cancer caused 670,000 deaths worldwide in 2022. Other than age and sex, women account for half of all cases without any particular risk factors. Furthermore, out of 185 nations, 157 had women with breast cancer as the most frequent malignancy (WHO, 2024). Breast cancer is a disorder in which aberrant breast cells develop into tumours due to their uncontrolled growth (Philomena, 2020). These tumours have the potential to become fatal if they are not treated. Breast cancer cells first grow in the breast's lobules or milk ducts. Its initial phase (in situ), may be identified early on and poses little danger to life. Cancer cells can invade adjacent breast tissue and result in thickening or lumpy tumours. Metastasis is the term for the process by which invasive tumours move to neighbouring lymph nodes or other organs. Metastasis may be severe and perhaps deadly. Treatment decisions are driven by the patient, the kind of cancer, and the extent of its distribution. It is common to use medicine, radiation treatment, and surgery. Breast cancer risk factors include increasing age, obesity, consuming alcohol, and a family history of the disease. The risk is further increased by variables such as historical radiation exposure, reproductive history (including age at first pregnancy), tobacco use, and postmenopausal hormone treatment. Breast problems in pediatric and adolescent patients may range from benign congenital anomalies to pathological findings. Most breast concerns in this age group are benign. On the other hand, cancer does occur in rare cases. Early detection of these cancerous events is essential (Adekeye et al., 2023; Adewuyi et al., 2023). Machine learning, a subfield of artificial intelligence, is important for the identification of diseases in healthcare (Jiang et al., 2021; Maltare et al., 2023).

The study reported a 0.5% annual increase in female breast cancer incidence in the US over the past four decades, primarily due to localized-stage and hormone receptor-positive disease. However, mortality rates have steadily declined since 1989 (Giaquinto et al., 2022; Al-Dosari & Abdellatif, 2024; Moses et al., 2022; Othman & Abdelwahab, 2021). Ak (2020) studied breast cancer tumor using machine learning techniques like logistic regression, k-nearest neighbors, and support vector machine. Data visualization and analysis were performed using R, Minitab, and Python. The study found that these techniques can significantly impact cancer detection decisions. Allugunti (2022) introduced a computer-aided Diagnosis (CAD) method for breast cancer detection using thermographic images. The study classified patients into cancer, no cancer, and non-cancerous classes using various ML approaches. Pre-processing of mammography pictures was investigated to improve classification success. Boutry et al. (2022) discussed benign tumors, their evolution, ecology, and interactions with hosts.

Researchers investigated the function of viruses, cellular activity, and genetic and epigenetic patterns in benign and malignant tumours. The research examined the relationship between the cell makeup of tumours and the surrounding microenvironment. It highlighted how benign tumours impact evolutionary pathways and individual fitness. Chaurasia and Pal (2020) used the Wisconsin Diagnostic Breast Cancer dataset to evaluate six ML techniques. The dataset characteristics from 32 to 12 were reduced using statistical methods. To reduce misclassification, six integrated models were used. The basic learners use a stacking classifier (Voting Classifier). By dividing the dataset between 80% training and 20% testing, hyperparameters were manually modified. Especially with the smaller data set, all machine learning methods demonstrated remarkable performance, with test accuracies over 90%. The study includes tumor-on-chips (TOCs) in the field of cancer.

Lei and colleagues highlighted their potential for simulating tumour microenvironments and intercellular communication (Lei et al., 2022). The authors emphasised the integration of electrochemical biosensors. These biosensors are highly valued for their selectivity, sensitivity, low cost of synthesis, and ease of use. They suggest future developments towards flexible biosensors for long-term tumor microenvironment monitoring. The article discussed the importance of cognitive computing in healthcare and highlighted its role in clinical decision-making, patient care, and timely, cost-effective treatment (Srivani et al. (2023). Park et al. (2022) explored the potential of flavonoids in breast cancer treatment. They explored genetic alterations linked to breast cancer and the use of radiation in developing radiation-resistant breast cancer. The study by Mahesh et al. (2022) developed an efficient ensemble method for early detection of breast cancer using ML techniques like SVM, KNN, DT Classifier, RF, and LR. They compared these techniques with XGBoost using k-fold cross-validation.

The study used Explainable AI to interpret decision-making in erythemato-squamous diseases. They used Random Forest and XGBoost models on a standard dataset.

The study emphasised the use of quantitative analysis to address ethical concerns with the application of AI in healthcare. It highlighted transparent, causative, and interpretable diagnoses of AI (Rathore et al., 2022).

Ethiopian women from Adama Science and Technology University participated in Tadesse et al.'s research (2022). The cervical cancer knowledge, attitude, and practice (KAP) of these students were rated. This cross-sectional survey included 667 female students who were selected using simple random selection. The findings show that only 2.2% of respondents had undergone a cervical cancer screening. However, 71.7% of respondents felt well about cervical cancer, and 60.6% of respondents had heard about it. An examination for cervical cancer was reportedly hindered by one's lack of knowledge about the disease. Screening programs and public awareness campaigns should be implemented to detect precancerous tumours in women as early as possible.

Comparative research on DCNNs for computer-aided diagnosis (CADx) of breast cancer using mammograms was carried out (Tsochatzidis et al., 2019). The authors trained and evaluated state-of-the-art CNNs using two mammography datasets. These datasets include areas of interest (ROIs) with the representation of benign or malignant cancer cells. Two distinct training events were examined in the study. It starts from scratch and optimises a pre-trained network. The pretrained network optimisation outperformed the freshly trained network. Using the Wisconsin Breast Cancer Dataset (WBCD), Uddin et al, (2023) examined the effectiveness of many ML algorithms in the detection of breast cancer. The effectiveness of many classifiers was assessed using the F1-score, recall, accuracy, precision, and error rate. The vote classifier turned out to be the most effective with the lowest mistake rate. The author also used the Flask micro-framework to develop a website.

Optimisation problems with feature extraction, classification, and segmentation are frequently performed to classify problems using deep convolutional networks (AlZubi, 2023; Cho et al., 2024; Wasik and Pattinson, 2024). These networks are frequently used to identify diseases in fish, animals, plants, and other systems. These methods are also used in industrial processes for defect identification and quality control (Porwal, 2024). These techniques have uses in the environment, biology, and industry. Deep convolutional networks are often used for difficult pattern recognition and diagnostic tasks due to their effectiveness and adaptability for industrial and farming sectors.

The objectives of this work are

- Conduct an experiment using Wisconsin Breast Cancer Dataset (WBCD).
- Identify connections between certain attributes and the dataset's class feature.
- Employ cutting-edge machine learning classifiers on the given dataset.
- Examine and compare the outcomes from various classifiers.
- To select the best model, compare the forecast performance and accuracy of the model.

The proposed work is divided into five parts. Section 1 provides an outline of the introduction and related studies. Section 2 provides materials and methods utilised in this study to evaluate breast cancer classification. Section 3 discusses the results and analysis of the suggested models. Section 5 will provide a conclusion to the entire work. Last section includes limitations and future work of the presented study.

2. MATERIALS AND METHODS

2.1 Data Source

These days, machine learning classifiers are widely used in medical diagnostics. Dr. William H. Wolberg created the Wisconsin Breast Cancer Dataset. that is used in this research. This dataset was collected from the UCI repository. It has 30 features and 569 patient information with no missing values. There are 212 cases of malignant or cancerous breast cancer and 357 cases of benign or non-cancerous breast cancer. An 80:20 ratio is used to randomly divide the dataset into training and testing sets. To prevent replication, the dataset's "id" column is eliminated because it has no bearing on the outcome.

Data Preprocessing (CSV file)

Data preparation is an important step in using a dataset to get the best possible results. The accuracy of the results from ML models may be greatly impacted by the presence of noise, missing values, or imbalanced data in datasets. Thus, before beginning the modelling process, it is essential to clean the dataset by eliminating these undesirable components from the data. Two columns in the dataset that are used to diagnose breast cancer are identified by the symbols M and B, which stand for malignant and benign instances. These verbal labels must be translated into numerical values since machine learning models often work with numerical data. In this case, the encoding of B is 1, which denotes benignity, whereas the encoding of M is 0, which denotes malignancy. Further, normalisation methods are used to make sure comparability and consistency among characteristics. Max-Min normalisation is a widely employed normalisation technique. The data is scaled in the form of 0 and 1. This prevents characteristics with larger numerical values from dominating the model training process. Additionally, data standardisation is carried out to assure that every feature has an equal impact on the distance metric. That is used by certain ML algorithms, such KNN. After standardising, the distribution of the data takes on a mean of 0 and a standard deviation of 1. Through this process, biases resulting from size disparities are minimised and the characteristics become more identical. In short, the data preparation step involves several important steps to build up the dataset for ML modelling. These procedures include cleaning the data and encoding categorical variables using normalisation and standardisation approaches. Carefully carrying out these preprocessing processes can improve the adaptability and performance of ML models.

2.3 Data Visualization

Data visualisation is the process of presenting information visually. So that data characteristics can be understood by individuals. A heatmap was used to examine the relationship between two or more variables in the dataset. Figure 1 displays the correlation matrix for the dataset related to breast cancer.

The dataset used in this research has 30 characteristics. It is high dimensional. An excessive number of features might lead to overfitting. It can complicated to obtain the most effective outcomes. The 30 characteristics are divided into 16 features using Principal Component Analysis (PCA). it boosts the effectiveness of the study.

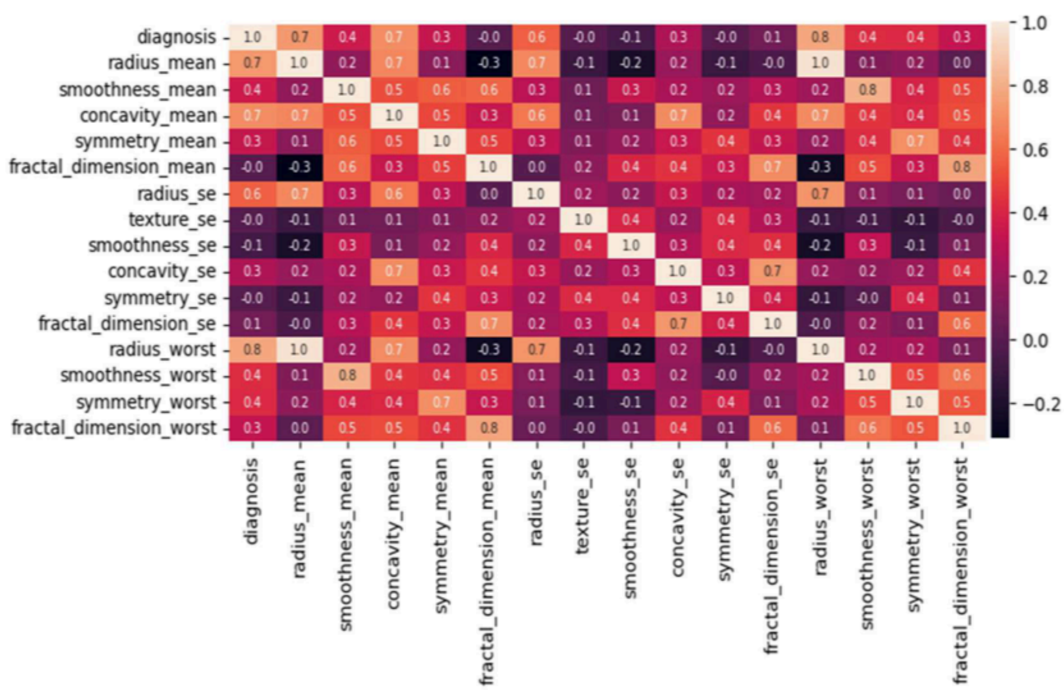

Figure 1: Data visualization

2.4 Computational ML Techniques

Analysis was done using open-source Python programs, Jupiter Notebook in Anaconda Environment, NumPy v1.20, TPOT, and Scikit-Learn 0.18.2 packages. Three learning strategies were used to predict malignant and benign samples. To measure the model's performance in the final examination, a train-test split technique was used. The dataset was divided into 80% for training and 20% for testing.

2.4.1 KNN Classifier

K-Nearest Neighbour is a simple and powerful ML approach. It is commonly employed for both classification and regression problems and well-suited for large training datasets. It creates predictions by using feature similarity as a non-parametric classifier. To predict values and assign labels to a new data point, it compares attributes of the training dataset. The technique employs computations to calculate Euclidean distance. The general formula is

$$\text{Euclidean distance} = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$

The variables x_i and y_i reflect the feature values of the i^{th} dimension for two data points. N represents total number of dimensions (features). The Euclidean distance between the new data point and all other points in the training dataset is determined using the KNN algorithm. The "k" data points with the smallest Euclidean distances are then chosen. Lastly, it uses the majority class of these "k" closest neighbours to determine the label to be applied to the new data point. KNN computation can predict data without assuming any underlying probability distributions. It is an easy-to-understand technique. However, the algorithm's effectiveness may be greatly affected by the value of the parameter k. Therefore, choosing a sensible value for k is essential to getting the best outcomes. In this study, K Fold =n_neighbors: 5 is taken. that represents the value of k is set to 5. When algorithm makes predictions for a new data point, the algorithm will consider the 5 nearest neighbors in the training dataset.

2.4.2 Linear Regression (LR) Method

The logistic regression model offers an appropriate solution for the classification of breast cancer as benign or malignant. It is beneficial when only two outcomes are available for prediction. Because it computes the probability of a given input belonging to a particular class. In logistic regression, the possibility of an input characteristic belonging to a certain class is connected to a logistic function. In this work, sigmoid function is used. the sigmoid function is defined as

$$P(Z) = \frac{1}{1 + e^{-Z}}$$

z stands for the linear combination of the model coefficients and

$$z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \pm \cdots + \beta_n x_n$$

Where β_0, β_1, β_2… β_n represent model coefficients and x_1, x_2, x_3… x_n represent input values. In addition, regularisation strategies to reduce overfitting and enhance generalisation are supported by logistic regression.

2.4.3 Decision Tree Model

In this work, a decision tree model is being built using a DecisionTreeClassifier from the scikit-learn module. The variable 'dt' contains the instantiated DecisionTreeClassifier. This classifier is a kind of tree-based model that divides the feature space into areas and labels. The performance of a decision tree is significantly influenced by its parameters. Some of these settings include the maximum depth of the tree, the minimal number of leaf nodes, and the minimum amount of samples needed to split a node. Using GridSearchCV from the scikit-learn model_selection module, one may find the optimal value for the "max_depth" parameter by grid searching. After analysing a parameter grid in detail, GridSearchCV chooses the hyperparameters that provide the best results. In this instance, a grid search is used since'max_depth' establishes the maximum depth of the decision tree. The value range for max_depth in the parameter grid param is 1 to 19. Decision trees are mostly created using the Iterative Dichotomiser 3 (ID3). ID3 uses measures like information gain and entropy to evaluate which qualities are best for creating trees and splitting data.

The following formula is used to determine the Information Gain (IG).

$$IG(D, A) = H(D) - H(D \mid A)$$

Where, the information gained by dividing dataset D for attribute A is represented as $IG\ (D,\ A)$. (H(D)) is the entropy of dataset D. The conditional entropy of dataset D for attribute A is denoted as H(D|A).

Figure 2 shows a flowchart of procedures used for work.

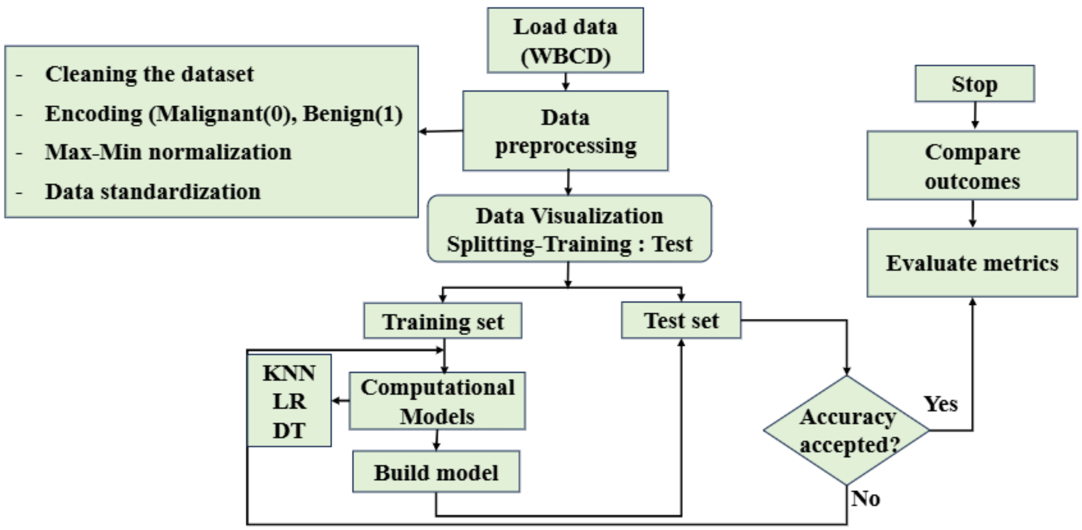

Figure 2: Flowchart of working principle

3. RESULTS AND DISCUSSION

The evaluation parameters demonstrate the predictive capacity of three ML models: LR, KNN and DT for input attributes. The confusion matrix provides a detailed summary of the KNN, LR and DT classifiers' performance to differentiate between instances of malignant and benign breast cancer (Figure 3).

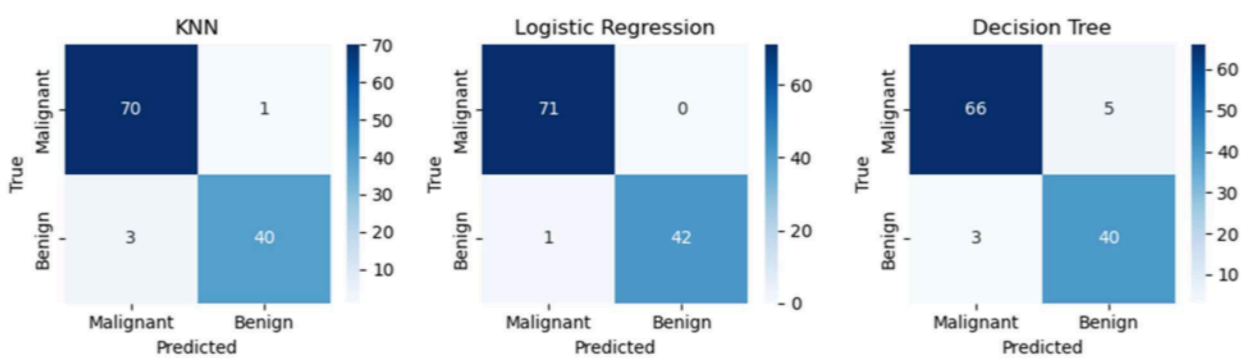

Figure 3: Confusion matrix of different models

The matrix is split into four cells for binary classification of malignant and benign. Each value represents a unique set of actual and predicted classifications. The accurate prediction of malignant cases is shown in the top-left cell, where 70 out of 71 cases of real cancer were accurately detected by the KNN model. However, one instance is incorrectly identified as benign. On the other hand, the bottom-right cell shows the classification of benign cases. The model successfully recognised 40 out of 43 real benign cases. Three benign instances were mistakenly classified as cancer.

The capacity of LR model to accurately detect events of malignant and benign is shown by the confusion matrix. It accurately identified 71 instances as true positives of Malignant events. Remarkably, there were no False Negatives, which indicates that the model did not mistakenly classify any Malignant events as benign. The model correctly classified 42 cases as Benign (True Negatives) in the Benign class and demonstrates model's capacity to distinguish non-malignant cases. However, the model did identify one benign event as cancer in a single False Positive.

Within the malignant class, the DT model correctly identified 66 events as malignant (true positives). However, it misclassified five instances of cancer as benign (false negatives). The tool accurately identified 40 cases as Benign (True Negatives) for the Benign class. However, it did identify three benign cases as malignant (False Positives). Finally, confusion matrix is a useful tool to evaluate the performance of the KNN, LR and DT classifiers. It points out both advantages and disadvantages of breast cancer diagnosis.

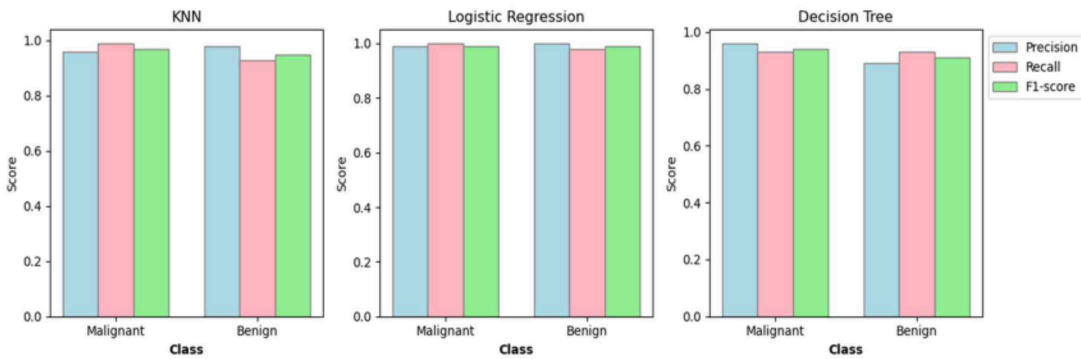

Figure 4: Classification report.

Figure 4 presents the classification metrics of KNN, LR and DT models. KNN model demonstrates a strong precision of 96% for malignant cases and indicates a remarkable level of precision in separating true malignant cases from all expected positive events. The model can identify the majority of real malignant cases, as shown by its 99% recall rate.

The total performance of the computational method is represented by an F1-score of 97%, which is obtained by adding accuracy and recall. The KNN model has an overall accuracy rate of 96%. It demonstrates good prediction ability of breast cancer. The LR approach detects actual malignant events with 99% accuracy. This model shows good performance of classification.

Moreover, it correctly diagnoses the majority of true benign events, with a 98% recall rate for benign incidences. LR achieves an F1-score of 99% for benign cases, which is similar to KNN's overall performance. LR produces correct predictions, with a total accuracy record of 99%. The model accurately detects most malignant events with a high degree of precision and recall (96% and 93%, respectively), and it captures a significant percentage of real malignant instances. The DT classifier has the potential to assist medical professionals in cancer detection, as seen by its balanced F1-score and overall accuracy of 93%. However, its slightly lower precision of 89% suggests an increased risk of false positives for benign patients. In conclusion, DT still works rather well overall and helps in the detection of cancer even if its overall accuracy has been significantly lowered. However, when it comes to accuracy and general performance, KNN lags much below LR.

4. COMPARISON OF PERFORMANCE OF DIFFERENT MODELS WITH EXISTING LITERATURE

The ML models used to predict cancer are shown in Table 1. Ak (2020) presents a precise logistic regression model with an accuracy rating of 98.10%. Chaurasia & Pal (2020) presented a range of models with accuracy levels of more than 90%, including CART, SVM, NB, and others. While Mahesh et al. (2022) published XGBoost with an incredible accuracy of 97.81%, Gupta & Garg (2020) gave many models with accuracy ranging from 95% to 98%. Rathore et al. (2022) presented RF model with 98.21% accuracy. Uddin et al. (2023) presented many models with accuracy ranging from 91.04% to 98.77%. These tests show the effectiveness of ML in diagnosing cancer. In the Presented study, the computational LR model has an accuracy of 99% for the classification of malignant and benign breast cancer tasks. DT model shows 93% overall accuracy. Furthermore, the performance of a 96% accurate K-Nearest Neighbours (KNN) model is good.

Table 1: Accuracy of predicted models with existing literature

Model	Accuracy	Author
LR	98.10%	Ak (2020)
CART, SVM, NB, KNN, LR, MLP, AB, GBC, RF, ET, Bagging, XGB, VC	>90%	Chaurasia & Pal (2020)
NN, LR, DT, RF, SVM, ANN,	95.80%, 95.80%, 95.80%, 97.20%, 97.20%, 98.24%	Gupta & Garg (2020)
XGBoost	97.81%	Mahesh et al. (2022)
Random Forest	98.21%	Rathore et al., 2022
SVM, RF, KNN, DT, NB, LR, AB, GB, MLP, NCC, VC	98.07%,94.20%, 96.84%, 94.20%, 91.04%, 98.42%, 96.31%, 95.78%, 97.54%, 93.15% 98.77%,	Uddin et al. (2023)
K-Nearest Neighbors (KNN)	96%	This work
Logistic Regression (LR)	99%	This work
Decision Tree (DT)	93%	This work

5. CONCLUSION

In the medical industry, ML is widely utilized for the classification and prediction of diseases among humans, animals and legumes. Especially when it comes to the analysis of big datasets with several attributes. Breast cancer is one of the most common kinds of cancer worldwide and often results in early death. On the other hand, early diagnosis is essential for lowering death rates and saving lives. By predicting breast cancer based on input data, this program helps users save the time and costs associated with standard diagnostic techniques. The study offered a methodology that uses three machine learning algorithms to predict breast cancer. The performance of the LR is better than the other methods with 99% accuracy. This work aims to enhance the breast cancer classification model while preserving its interpretability, transparency, accuracy, and fairness via the use of explainable machine learning techniques.

6. LIMITATIONS AND FUTURE WORK

It is essential to acknowledge fundamental limitations and highlight areas of presented work where more work and development are needed. The performance of models is dependent on the characteristics of the training and evaluation datasets. So, more attributes must be added to the dataset. The variations in the distributions of the dataset, imbalances across classes, and complexity may have a significant effect on the performance. More research should concentrate on evaluating the adaptability of the models over a wide range of datasets to determine their stability and suitability in practical contexts. Furthermore, the claimed accuracies often do not account for feature selection, cross-validation methods, and hyperparameter tuning. All of which can significantly affect the whole model construction process. In conclusion, machine learning may advance by promoting the development of more dependable, intelligible, and deployable models that effectively solve issues in the real world by addressing the constraints that have been found and suggesting new research avenues. This collaborative endeavour will facilitate machine learning solutions' wider use and positive societal impact and also enhancing the validity and reliability of machine learning research.

ACKNOWLEDGEMENT

Funding Details

This research received no external funding.

Authors' contributions

All authors contributed toward data analysis, drafting and revising the paper and agreed to be responsible for all the aspects of this work.

Declaration of Conflicts of Interests

Authors declare that they have no conflict of interest.

Availability of data and materials

The datasets used in the current study available from the corresponding author on reasonable request.

Use of Artificial Intelligence

Not applicable

Declarations

Authors declare that all works are original and this manuscript has not been published in any other journal

REFERENCES

Adekeye, A., Lung, K. C., Brill, K. L., & Department of Surgery, Thomas Jefferson University, Philadelphia, Pennsylvania. (2023). Pediatric and adolescent breast Conditions: a review. In *Journal of Pediatric and Adolescent Gynecology* (Vol. 36, Issue 1, pp. 5–13). https://doi.org/10.1016/j.jpag.2022.11.001

Adewuyi, R., Aweda, J., Ogunwoye, F., Omoniyi, P., & Jen, T.-C. (2023). Prediction of Impact Strength of TIG Welded Cr-Mo Steel Using Artificial Neural Networks. Metallurgical and Materials Engineering, 30(1), 61–69. https://doi.org/10.56801/MME1031

Ak, M. F. (2020). A comparative analysis of breast cancer detection and diagnosis using data visualization and machine learning applications. *Healthcare*, *8*(2), 111. https://doi.org/10.3390/healthcare8020111

Al-Dosari, M. N. A. & Abdellatif, M. S. (2024). The Environmental Awareness Level Among Saudi Women and Its Relationship to Sustainable Thinking. Acta Innovations, 52, 28–42. https://doi.org/10.62441/ActaInnovations.52.4

Allugunti, V. R. (2022). Breast cancer detection based on thermographic images using machine learning and deep learning algorithms. *International Journal of Engineering in Computer Science*, *4*(1), 49–56. https://doi.org/10.33545/26633582.2022.v4.i1a.68

AlZubi, A.A. (2023). Artificial Intelligence and its Application in the Prediction and Diagnosis of Animal Diseases: A Review. *Indian Journal of Animal Research*. 57(10): 1265-1271. https://doi.org/10.18805/IJAR.BF-1684

Boutry, J., Tissot, S., Ujvari, B., Capp, J., Giraudeau, M., Nedelcu, A. M., & Thomas, F. (2022). The evolution and ecology of benign tumors. *Biochimica Et Biophysica Acta. Reviews on Cancer/Biochimica Et Biophysica Acta, CR. Reviews on Cancer*, *1877*(1), 188643. https://doi.org/10.1016/j.bbcan.2021.188643

Chaurasia, V., & Pal, S. (2020). Applications of machine learning techniques to predict diagnostic breast cancer. *SN Computer Science/SN Computer Science*, *1*(5). https://doi.org/10.1007/s42979-020-00296-8

Cho, O.H., Na, I.S. and Koh, J.G. (2024). Exploring Advanced Machine Learning Techniques for Swift Legume Disease Detection. *Legume Research*. https://doi.org/10.18805/LRF-789

Giaquinto, A. N., Sung, H., Miller, K. D., Kramer, J. L., Newman, L. A., Minihan, A., Jemal, A., & Siegel, R. L. (2022). Breast Cancer Statistics, 2022. *Ca*, *72*(6), 524–541. https://doi.org/10.3322/caac.21754

Gupta, P., & Garg, S. (2020). Breast Cancer Prediction using varying Parameters of Machine Learning Models. *Procedia Computer Science*, *171*, 593–601. https://doi.org/10.1016/j.procs.2020.04.064

Jiang, N., Zhao, Z., & Xu, P. (2021). Predictive Analysis and Evaluation Model of Chronic Liver Disease Based on BP Neural Network with Improved Ant Colony Algorithm. Journal of Healthcare Engineering, 2021, 1–7. https://doi.org/10.1155/2021/3927551

Lei, L., Ma, B., Xu, C., & Liu, H. (2022). Emerging tumor-on-chips with electrochemical biosensors. *TrAC. Trends in Analytical Chemistry*, *153*, 116640. https://doi.org/10.1016/j.trac.2022.116640

Maltare, N. N., Sharma, D. & Patel, S. (2023). An Exploration and Prediction of Rainfall and Groundwater Level for the District of Banaskantha, Gujrat, India. *International Journal of Environmental Sciences*, 9(1), 1-17. https://www.theaspd.com/resources/v9-1-1-Nilesh%20N.%20Maltare.pdf

Mahesh, T.R., Kaladevi, A. C., Balajee, J. M., Vivek, V., Prabu, M., & Muthukumaran, V. (2022, December 31). *An efficient ensemble method using K-Fold cross validation for the early detection of benign and malignant breast cancer*.

Moses, M. B., Nithya, S. E. & Parameswari, M. (2022). Internet of Things and Geographical Information System based Monitoring and Mapping of Real Time Water Quality System. *International Journal of Environmental Sciences*, 8(1), 27-36.

Othman, K. M. M. & Abdelwahab, H. (2021). Prediction of the Optimum Asphalt Content Using Artificial Neural Networks. *Metallurgical and Materials Engineering,* 27(2):227-242. https://doi.org/10.30544/579.

Park, M. Y., Kim, Y., Ha, S. E., Kim, H. H., Bhosale, P. B., Abusaliya, A., Jeong, S. H., & Kim, G. S. (2022). Function and application of flavonoids in the breast cancer. *International Journal of Molecular Sciences*, *23*(14), 7732. https://doi.org/10.3390/ijms23147732

Philomena, V. S. (2020). A Statistical Approach towards Breast and Vaginal Cancer in Women. Bio-Science Research Bulletin, 36(1), 22-32.

Porwal, S., Majid, M., Desai, S. C. Vaishnav, J. & Alam, S. (2024). Recent advances, Challenges in Applying Artificial Intelligence and Deep Learning in the Manufacturing Industry. *Pacific Business Review* (International), 16(7), 143-152.

Rathore, A. S., Arjaria, S. K., Gupta, M., Chaubey, G., Mishra, A. K., & Rajpoot, V. (2022). Erythemato-Squamous Diseases Prediction and Interpretation using Explainable AI. *IETE Journal of Research*, 1–20. https://doi.org/10.1080/03772063.2022.2114953

Srivani, M., Murugappan A., Mala T., (2023). Cognitive computing technological trends and future research directions in healthcare — A systematic literature review. *Artificial Intelligence in Medicine*, *138*, 102513. https://doi.org/10.1016/j.artmed.2023.102513

Tadesse, A., Segni, M. T., & Demissie, H. F. (2022). Knowledge, Attitude, and Practice (KAP) toward Cervical Cancer Screening among Adama Science and Technology University Female Students, Ethiopia. *International Journal of Breast Cancer*, *2022*, 1–9. https://doi.org/10.1155/2022/2490327

Tsochatzidis, L. T., Costaridou, L., & Pratikakis, I. (2019). Deep Learning for Breast Cancer Diagnosis from Mammograms—A Comparative Study. *Journal of Imaging*, *5*(3), 37. https://doi.org/10.3390/jimaging5030037

Uddin, K. M. M., Biswas, N., Rikta, S. T., & Dey, S. K. (2023). Machine learning-based diagnosis of breast cancer utilizing feature optimization technique. *Computer Methods and Programs in Biomedicine*

Update, *3*, 100098. https://doi.org/10.1016/j.cmpbup.2023.100098

Wasik, S. and Pattinson, R. (2024). Artificial Intelligence Applications in Fish Classification and Taxonomy: Advancing Our Understanding of Aquatic Biodiversity. *FishTaxa*, 31: 11-21.

World Health Organization: WHO & World Health Organization: WHO. (2024, March 13). Breast cancer. https://www.who.int/news-room/fact-sheets/detail/breast-cancer [Accessed Date: 18/05/2024]