Bulletin of Electrical Engineering and Informatics
Vol. 13, No. 6, December 2024, pp. 4306∼4315
ISSN: 2302-9285, DOI: 10.11591/eei.v13i6.7866

An interpretable machine learning-based breast cancer classification using XGBoost, SHAP, and LIME

Monoronjon Dutta, Khondokar Md. Mehedi Hasan, Alifa Akter, Md. Hasibur Rahman, Md. Assaduzzaman
Department of Computer Science Engineering, Daffodil International University, Dhaka, Bangladesh

Article Info

Article history:

Received Nov 13, 2023
Revised Apr 6, 2024
Accepted May 17, 2024

Keywords:

Breast cancer detection
Chi-square
Feature selection
SHAP and LIME
Machine learning techniques

ABSTRACT

Globally, breast cancer is among the most prevalent and deadly tumors that affect women. Early and accurate identification of breast cancer is essential for effective treatment planning and improving patient outcomes. This research focuses on improving breast cancer classification accuracy through machine learning (ML) methodologies, emphasizing interpretability. The study utilized the chi-square method to enhance model testing performance by pinpointing the most significant features for further analysis. The study also improved data quality by identifying and removing outliers, thus minimizing the influence of data irregularities on the performance of the models. For classification, the study evaluated six different ML algorithms—namely extreme gradient boosting (XGBoost), decision tree (DT), AdaBoost (AB), support vector machine (SVM), gradient boosting (GB), and K-nearest neighbors (KNN)—each applied to distinguish between the two variants of breast cancer. Among these, the XGBoost classifier emerged as the most accurate, achieving an impressive 99.30% accuracy rate. Moreover, the research incorporated shapley additive explanations (SHAP) and local interpretable model-agnostic explanations (LIME) methods to boost the interpretability of the proposed model, offering crucial insights into the model's decision-making process. Applying these interpretability techniques provided significant insights into the predictive factors influencing healthcare outcomes, ensuring the classification approach's transparency and reliability.

Corresponding Author:

Md. Assaduzzaman
Department of Computer Science and Engineering, Daffodil International University
Dhaka, Bangladesh
Email: assaduzzaman.cse@diu.edu.bd

1. INTRODUCTION

Breast cancer affects women globally and is a common and potentially deadly illness. It must be detected early and accurately to improve patient outcomes and survival rates. Surpassing lung cancer as the most prevalent cancer globally, it saw around 2.3 million new cases diagnosed in both men and women combined [1]. Among females, breast cancer constituted a quarter of all cancer cases in 2020, making it the most commonly diagnosed cancer among women [2]. Its incidence has been steadily increasing, especially in transitioning countries, posing a significant burden on public health. Sadly, an estimated 685,000 women died of breast cancer in 2020, representing 16% of all female cancer deaths or one in every six cancer-related deaths in women. Recognizing the gravity of the situation, the World Health Organization has recently launched the Global Breast Cancer Initiative to address it [3]. The conventional diagnosis procedure, which mainly depends on histological analysis, is both time-consuming and subject to interpretation variability. Machine learning (ML) has become a transformative tool, allowing improved, precise, and reliable cancer classification. Among ML algorithms, extreme gradient boosting (XGBoost) is a standout ML algorithm known for its efficacy in managing complex datasets, making it well-suited for medical diagnostic purposes.

However, the effectiveness of ML in the medical field is frequently hampered by its ”black box” character, which makes decision-making ambiguous and difficult for medical professionals to trust. Interpretability techniques such as shapley additive explanations (SHAP) and interpretable model-agnostic explanations (LIME) increase transparency and promote informed treatment decisions. These strategies clarify the predictions of ML models by emphasizing the impact of specific features on outcomes.

Several recent methodologies for breast cancer prediction have been developed, demonstrating great technological advances. Recent research in this field by Michael *et al.* [4] introduced a CAD framework for breast ultrasound image classification, achieving remarkable results using LightGBM with a 99.80% F1 score, 100.0% precision, 99.86% accuracy, and 99.60% recall. Amethiya *et al.* [5] explored ML algorithms and biosensors for early breast cancer classification, with the fuzzy ELM-RBF ML models achieving an impressive accuracy of 98.05%. Elsadig *et al.* [6] evaluated classification models for breast cancer prediction, identifying support vector machine (SVM) as the top performer with an accuracy of 97.7%. Egwom *et al.* [7] utilized the LDA-SVM model for breast cancer diagnosis, achieving remarkable accuracy and precision by effectively handling missing values. Tarawneh *et al.* [8] applied a decision tree (DT)-based data mining approach for early breast cancer detection, with the DT classifier demonstrating high accuracy. Ebrahim *et al.* [9] compared classical and deep learning methods, with DT and ensemble techniques achieving an accuracy of 98.7%. Kadhim and Kamil [10] predicted breast cancer using the GB model with an impressive 96.77% F1 score, underscoring the importance of ML in breast cancer prediction and diagnosis. Using the white blood cell (WBC) dataset, Birchha and Nigam [11] assessed the averaged-perceptron machine-learning classifier. They found that it achieved a remarkable accuracy of 0.984 in breast cancer detection, with no false-negative predictions, outperforming other models in similar studies. Massari *et al.* [12] developed an ontological model using the DT algorithm for breast cancer detection, achieving a remarkable prediction accuracy of 97.10%. The fast learning network (FLN) model gained outstanding accuracy rates of 98.37% and 96.88% on the WBCD and WDBC datasets, indicating its potential for broader medical applications in [13]. Research by Omotehinwa *et al.* [14], incorporating LightGBM, SMOTE, and tree-structured parzen estimator (TPE), achieved a notable 99.12% accuracy on the WDBC dataset, highlighting the importance of data augmentation and hyperparameter optimization in ML diagnostics. Liza *et al.* [15] exploration of early disease detection revealed that Random Forest (RF) and AdaBoost (AB) models outperformed others with an impressive 99.20% accuracy and a near-99% receiver operating characteristic (ROC) score. Gad *et al.* [16] emphasized the significance of feature selection and, using a pigeon-inspired optimizer-based approach with the RF model on the WBC dataset, achieved an accuracy of 97.2% along with uniform F-score, recall, and precision rates of 97.3%.

The main contribution of the research is to enhance breast cancer classification through ML techniques while identifying the shortest performance time to deliver optimal classification results. The study employed the chi-square feature selection technique to identify the most significant features and outlier detection methods to improve the data quality, contributing to the proposed model's high performance. Additionally, integrated SHAP and LIME approaches enhance the interpretability of the proposed model, providing vital insight into the decision-making process and ensuring transparency and trustworthiness in clinical applications.

This research paper is formatted as follows: section 2 presents a method. Section 3 shows the experimental environment's setup and result analysis. Section 4 is a comparative analysis of the previous study. Section 5 about discusses and conclusions.

2. METHOD

The primary goal of this study is to determine the most precise classification method for breast cancer diagnosis. The full structure is shown in Figure 1. The work commences with data collection. Following the dataset collection, various data preprocessing techniques are applied, including data clearing, outlier removal and Chi-square. Labeled data is obtained to evaluate the algorithms. This dataset is divided into two sets, with 75% dedicated to training and test data, or 25% of the data utilized for evaluating the model's effectiveness. After splitting the dataset, multiple ML models are trained, and the best models are compared. Then, the best models are subjected to the implementation of SHAP and LIME to determine the best-predicting features. The ultimate goal of the research is to determine the optimal algorithm for classifying breast cancer.

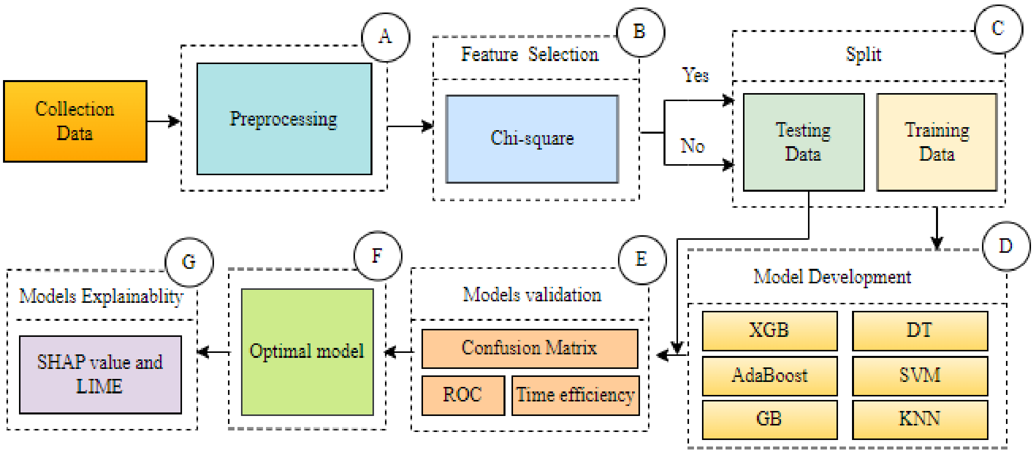


Figure 1. Proposed method of the breast cancer classification

2.1. Describe the dataset

The dataset central to this investigation was sourced from Kaggle, a prominent online repository renowned for its vast collection of datasets. The compilation of data, encompassing 569 items with 33 features, was meticulously assembled by Dr. William H. Wolberg, Wisconsin, USA. The dataset's features encompass a comprehensive array of measurements, including the radius (mean of distances from the center to the perimeter), texture (standard deviation of grey-scale values), perimeter, area, smoothness (local variations in radius lengths), compactness (calculated as the square of the perimeter divided by the area, less one), concavity (reflecting the severity of the concave portions of the contour), concave points (denoting the number of concave portions of the contour), symmetry, and fractal dimension, described as the ”coastline approximation” minus one. Each entry within this dataset has been meticulously categorized as either ”malignant” (M) or ”benign” (B), terms that medically signify the nature of tumor cells [17]. Of all the data records considered, 357 (62.7%) are identified with benign tumor cells, and the remaining 212 (37.3%) indicate malignant tumor cells.

2.2. Preprocessing

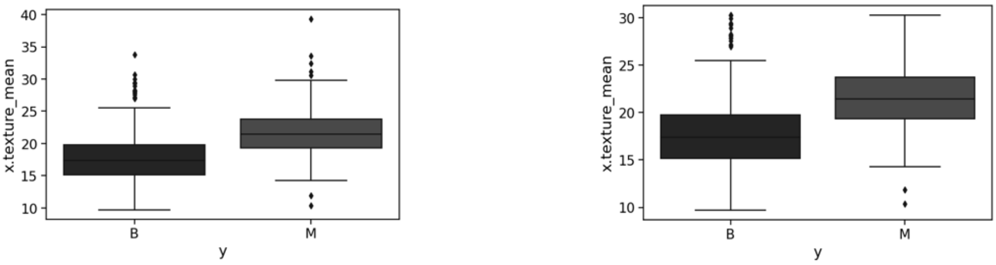
This study applies preprocessing techniques, such as outlier remove and the chi-square test, to improve the model performance.

2.2.1. Outlier detection and removal

The dataset contains various features related to breast cancer characteristics, each with specific outlier records identified through upper limits and quartile ranges. For instance, in the feature x.radius_mean, there are 14 outlier records with values exceeding the upper limit of 21.9. Similarly, in x.texture_mean, seven outlier records exceed the upper limit of 30.2. This pattern repeats across other features, such as x.area_mean, x.radius_se, and x.perimeter_worst, with varying numbers of outlier records detected in each feature (see Figures 2(a) and (b)). When these outliers are eliminated, the dataset's statistical analyses and prediction models can become more accurate and reliable [18], [19].

2.3. Feature selection

The study employs the chi-square statistical method for feature selection to analyze categorical data. This technique underscores the significance of the relationship between two categorical variables by determining whether the observed frequency of a category significantly deviates from what would be expected under the assumption of independence [20]. Utilizing this method, the researcher identifies the top 20 features deemed most relevant for analysis: $/x_{\text{radius\_mean}}$, $x_{\text{texture\_mean}}$, $x_{\text{perimeter\_mean}}$, $x_{\text{area\_mean}}$, $x_{\text{compactness\_mean}}$, $x_{\text{concavity\_mean}}$, $x_{\text{concave\_pts\_mean}}$, $x_{\text{radius\_se}}$, $x_{\text{perimeter\_se}}$, $x_{\text{area\_se}}$, $x_{\text{compactness\_se}}$, $x_{\text{concavity\_se}}$, $x_{\text{radius\_worst}}$, $x_{\text{texture\_worst}}$, $x_{\text{perimeter\_worst}}$, $x_{\text{area\_worst}}$, $x_{\text{smoothness\_worst}}$, $x_{\text{compactness\_worst}}$, $x_{\text{concavity\_worst}}$, $x_{\text{concave\_pts\_worst}}$, and $x_{\text{symmetry\_worst}}$. These selected features, indicative of the variables' impact on the dataset, are critical for further analysis and modeling in research.



(a) (b)

Figure 2. Outlier detection and removal in this proposed dataset; (a) before and (b) after

2.4. Machine learning algorithms

The study analyzes various machine-learning algorithms' approaches to breast cancer diagnosis. These models include XGB, DT, AB, SVM, gradient boosting (GB), and KNN.

2.4.1. Extreme gradient boosting

The XGBoost library, encompassing GB algorithms designed for speed and performance, is adept at solving modern data science challenges and fits well into the latest toolsets. It operates using a scalable ensemble of DT and applies GB to minimize loss functions iteratively, thereby improving the model's predictive accuracy. A distinctive feature of XGBoost is its use of DT as base learners, which significantly contributes to its high accuracy levels in (1) and (2) [21], [22]:

$$L_{\text{xgb}} = \sum_{i=1}^{N} L(y_i, F(x_i)) + \sum_{m=1}^{M} \Omega(h_m) \tag{1}$$

$$\Omega(h) = \gamma T + \frac{1}{2}\lambda \|w\|^2 \tag{2}$$

where T denotes the number of tree leaves and w represents the leaf weights. XGBoost also incorporates a shrinkage factor to control the learning rate, preventing overfitting by reducing the step size of each additive model expansion.

2.4.2. Decision tree

Employing a graph resembling a tree for decision-making and data classification, the DT model is supervised learning. Starting with a single node, it divides into multiple possible outcomes: an attribute test is indicated by each internal node, each branch represents a test result, and a class label is shown by each leaf node in (3) and (4) [23], [24]:

$$Entropy = -\sum_{i=1}^{n} p_i \times \log(p_i) \tag{3}$$

$$Gini\ index = 1 - \sum_{i=1}^{n} p_i^2 \tag{4}$$

The entropy measures the disorder or uncertainty, and the Gini index quantifies the distribution of classes within a subset of the dataset; both are used to choose the most informative features at each decision node.

2.4.3. AdaBoost

AB, also known as adaptive boosting, is an ensemble technique that builds a powerful classifier by combining weak classifiers. The method dynamically modifies the weights of erroneously classified examples, directing succeeding classifiers' attention toward more complex cases in (5) [25]:

$$H_k(p) = \pm \left(\sum_{k=1}^{k} a_k h_k(p) \right) \tag{5}$$

where H_k is the combined hypothesis, a_k are the weights for each hypothesis, and $h_k(p)$ are the individual weak hypotheses.

2.4.4. Support vector machine

SVM is a sophisticated classification approach that identifies the best hyperplane to maximize the margin between classes. SVM seeks to identify the optimum border that divides the data into classes (6) [26]:

$$L(\omega) = \sum_{i=1} \max(0, 1 - y_i(\omega^T x_i + b)) + \lambda ||\omega||^2 \tag{6}$$

where the loss function is defined by the sum of the distances of misclassified points from the decision boundary, and the regularization term penalizes the complexity of the model.

2.4.5. Gradient boosting

GB is a sequential learning method that corrects the mistakes of prior models and adds them to form a final strong model. It is an effective method for regression and classification problems in (7) [27]:

$$F_m(X) = F_{m-1}(X) + \eta \times f_m(X) \tag{7}$$

where F_m is the improved model, f_m is the weak learner, and η is the learning rate.

2.4.6. K-Nearest Neighbors

The non-parametric KNN technique is used for regression and classification. The input of a KNN is the k closest training instances in the feature space; the output of the KNN is contingent upon whether it is employed for classification or regression in (8) and (9) [28]:

$$\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \tag{8}$$

$$f_{ij} = 1 - \frac{6 \sum_{i=1}^{n} (\text{rank}(p_i) - \text{rank}(q_j))^2}{n(n^2 - 1)} \tag{9}$$

where ’P’ represents the training tuples, while ’Q’ stands for the testing tuples, with ’n’ indicating the overall count of observations. Typically, the value of ’Fij’ ranges between -1 and 1.

3. RESULT ANALYSIS

The study employed a Core i5-10210U CPU with a clock speed of 1.60 GHz. An NVIDIA GeForce MX110 was utilized for graphics processing, complemented by 8 GB of RAM. The model construction was facilitated using Python alongside libraries such as Numpy, Pandas, and Matplotlib. These components were integrated within the Jupyter Notebook platform, forming the computational foundation for the research.

3.1. Accuracy

Accuracy is the ratio of successfully classified instances to total instances evaluated (10) [29]. This statistic assesses the model’s ability to predict outcomes throughout the full dataset accurately [30].

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{10}$$

3.2. Precision

Precision is the percentage of successfully predicted positive samples among all cases projected as positive by the model in (11) [31], [32]:

$$Precision = \frac{TP}{TP + FP} \tag{11}$$

3.3. Recall

Recall estimates the proportion of correctly predicted positive instances from all actual positive samples in the dataset in (12) [33]:

$$Recall = \frac{TP}{TP + FN} \tag{12}$$

3.4. F1-Measure

F1-Measure is the harmonic mean of precision and recall in (13) [34]. It supplies a balanced benchmark between precision and recall, primarily when an imbalance exists between the categories.

$$F1\text{-}Measure = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{13}$$

3.5. Performance analysis

The XGB model stands out for its outstanding performance, achieving the highest accuracy, precision, recall, and F1 score among the assessed models in (10) to (13). In contrast, models such as the DT, AB, SVM, and KNN exhibit lower success metrics, with KNN notably registering the most insufficient accuracy. While not reaching the pinnacle of accuracy demonstrated by the XGB model, the GB model still showcases commendable performance metrics (see Tables 1 and 2). This study underscores the varied capabilities of different ML algorithms, highlighting the XGB model's exceptional ability to predict outcomes and computational efficiency accurately in this comparison.

Table 1. Models performance analysis

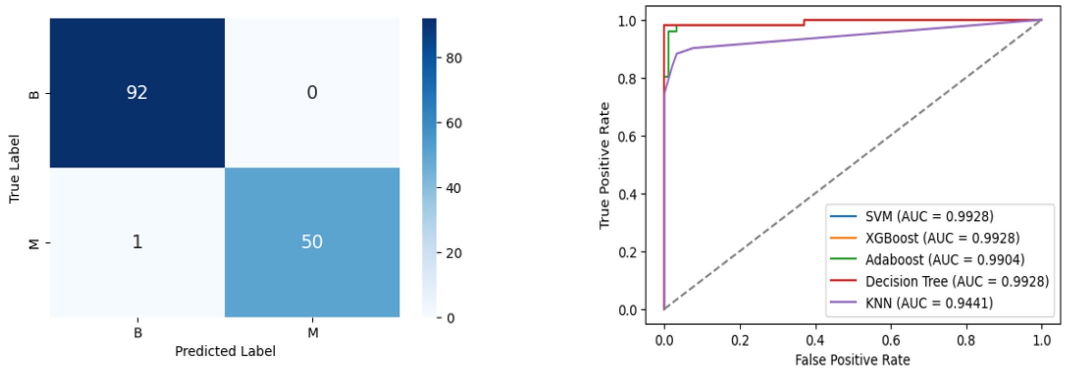
Models	Accuracy	Precision	Recall	F1-measure	Findings
XGB	0.9930	0.9946	0.9902	0.9923	Highest accuracy
DT	0.9510	0.9425	0.9532	0.9473	Lowest accuracy
AB	0.9790	0.9792	0.9750	0.9770	Lowest accuracy
SVM	0.9650	0.9639	0.9597	0.9617	Lowest accuracy
GB	0.9860	0.9848	0.9848	0.9848	Slightly reduce from highest accuracy
KNN	0.9301	0.9370	0.9107	0.9216	Lowest accuracy

Table 2. Models computational efficiency

Models	Without feature selection			With feature selection		
	Train(s)	Test(s)	Accuracy(%)	Training(s)	Test(s)	Accuracy(%)
XGB	0.0883	0.0138	99.30	0.0623	0.0050	99.30
DT	0.0129	0.0034	95.10	0.0077	0.0024	93.01
AB	0.3746	0.0199	97.90	0.1518	0.0100	94.41
SVM	0.8452	0.0035	96.50	0.4853	0.0026	97.20
GB	0.4497	0.0035	98.60	0.3255	0.0032	98.60
KNN	0.0040	0.0811	93.01	0.0035	0.0123	93.01

The confusion matrix, esteemed for mapping the counts of predicted versus actual values, emerges as a pivotal tool in assessing the effectiveness of classification models, encompassing both binary, and multiclass strategies. Employing an N × N matrix, this tool delineates the precision of model predictions about actual outcomes [35]. A closer inspection of Figure 3(a) reveals that XGBoost secures a commendable classification prowess, distinctly outperforming KNN, demonstrating diminished accuracy in classification tasks.

The indicated ROC curve visually compares classifier performance, where SVM, XGBoost, and DT models stand out with near-identical, exemplary AUC scores of 0.9928, reflecting their high effectiveness in distinguishing between classes (see Figure 3(b)). Except for KNN, these models exhibit high accuracy in classification tasks, as deduced from their proximity to the top left corner of the ROC space [36].

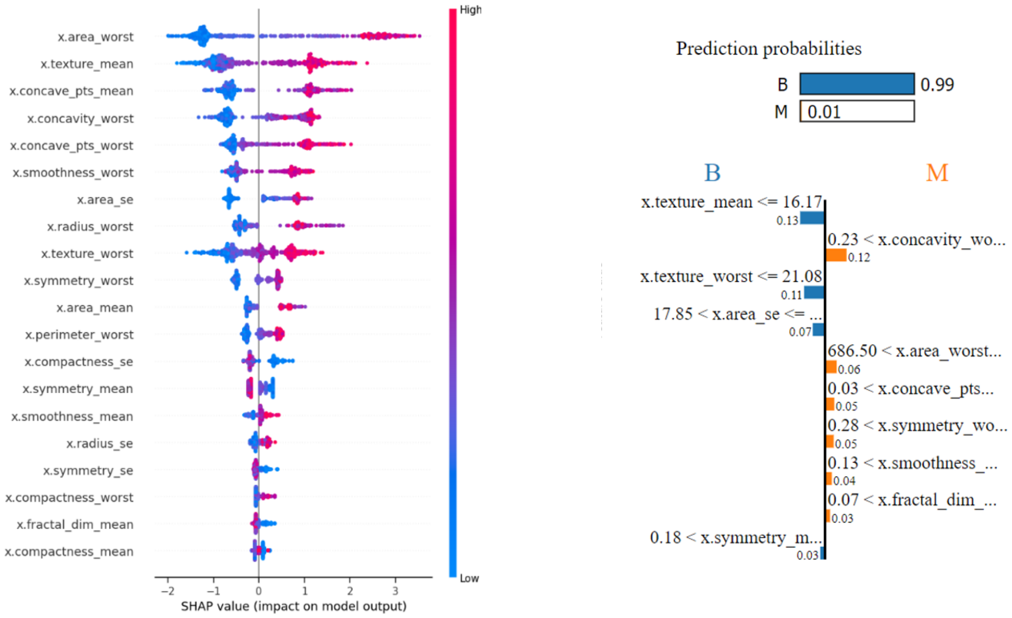


(a) (b)

Figure 3. Proposed model: (a) confusion matrix and (b) ROC curve

3.6. Discussions using LIME and SHAPE

The study aimed to demystify the ’black box’ character of the XGBoost algorithm by adding SHAP parameters and explaining its mechanism in breast cancer predicted outcomes (created SHAP python package) [37], [38]. Furthermore, they used SHAP dependence analysis to show how a single variable affected the output of the XGBoost model, as shown in Figure 4(a). A more granular examination revealed the top 20 critical clinical features influencing the XGBoost model’s output, with ”X.area_worst” registering the highest value. Conversely, ”x.texture_mean” and ”x.compactness_mean” were identified as the second and last lowest values, respectively. Subsequently, the study employed SHAP force analysis and the LIME algorithm for a nuanced explanation of individual predictions concerning death from breast cancer, categorizing cases into Benign and Malignant, as shown in Figure 4(b). Using SHAP force analysis and the LIME method, Figure 4(b) illustrates a particular deceased example. The XGBoost model projected the probability of breast cancer as benign at 0.99% and malignant at 0.01%. Notably, the ”x.text _Worst” value was 21.08 for Benign, while the highest value for Malignant was 686.50 [39].



(a) (b)

Figure 4. The SHAP summary plot; (a) shows the top 20 clinical features that contribute to the XGBoost model and explaining individual prediction results and (b) using the LIME algorithm

4. COMPARATIVE ANALYSIS

In the comparative analysis conducted by various authors in 2023, Birchha and Nigam [11] utilized the Avg-perceptron method, yielding an accuracy of 0.98%, the lowest among the methods assessed. Similarly, Massari *et al.* [12] employed ontology, achieving an accuracy of 0.977%, ranking the most down. Albadr *et al.* [13] applied FLN, with a reported accuracy of 0.9883%, still falling within the lower range of accuracies. Omotehinwa *et al.* [14] utilized TBL, attaining an accuracy of 0.9912%, again recording one of the weakest accuracies. The researchers in [15], [16] employed RF, achieving accuracies of 0.9790% and 0.973%, respectively, among the lowest in the analysis, and no one applied to shape and Lime. In contrast, the proposed model, XGB, presented by an unspecified author, demonstrated the highest accuracy among the methods assessed, achieving an accuracy of 0.9930% (see Table 3). This indicates a notable improvement in accuracy compared to the other methods evaluated in the study.

Table 3. Comparative analysis

Authors name	Publish year	Proposed method	Accuracy(%)	XAI used	Remarks
Birchha and Nigam [11]	2023	Avg-perceptron	0.98	No	Lowest accuracy
Massari *et al.* [12]	2023	Ontology	0.977	No	Lowest accuracy
Albadr *et al.* [13]	2023	FLN	0.9883	No	Lowest accuracy
Omotehinwa *et al.* [14]	2023	TBL	0.9912	No	Lowest accuracy
Liza *et al.* [15]	2023	RF	0.9790	No	Lowest accuracy
Gad *et al.* [16]	2023	RF	0.973	No	Lowest accuracy
Proposed model		XGB	0.9930	Yes	Highest accuracy

5. CONCLUSION

ML models are increasingly utilized to model and predict spatial phenomena. A significant challenge these models face, limiting their broader adoption in spatial data modeling, is their interpretability, especially when the goal is to uncover underlying processes and features. The paper presented contributes to breast cancer classification, distinguishing between benign and malignant cases. It addresses this gap by demonstrating which features from the dataset are more important for deriving spatial insights from ML models, utilizing SHAP and LIME with the best-performing XGBoost model. Applying the values of SHAP and the LIME method in ML can help physicians make clinical decisions about cancer classification. The future work plan includes training the model with additional data and improving the model's testing time efficiency.

REFERENCES

[1] H. Sung *et al.*, "Global Cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: a Cancer Journal for Clinicians*, vol. 71, no. 3, pp. 209–249, Feb. 2021, doi: 10.3322/caac.21660.

[2] E. Heer, A. Harper, N. Escandor, H. Sung, V. McCormack, and M. M. Fidler-Benaoudia, "Global burden and trends in premenopausal and postmenopausal breast cancer: a population-based study," *The Lancet Global Health*, vol. 8, no. 8, pp. e1027–e1037, Aug. 2020, doi: 10.1016/s2214-109x(20)30215-1.

[3] B. O. Anderson *et al.*, "The Global Breast Cancer Initiative: a strategic collaboration to strengthen health care for non-communicable diseases," *The Lancet Oncology*, vol. 22, no. 5, pp. 578–581, May 2021, doi: 10.1016/S1470-2045(21)00071-1.

[4] E. Michael, H. Ma, H. Li, and S. Qi, "An optimized framework for breast cancer classification using machine learning," *BioMed Research International*, vol. 2022, pp. 1–18, Feb. 2022, doi: 10.1155/2022/8482022.

[5] Y. Amethiya, P. Pipariya, S. Patel, and M. Shah, "Comparative analysis of breast cancer detection using machine learning and biosensors," *Intelligent Medicine*, vol. 2, no. 2, Oct. 2021, doi: 10.1016/j.imed.2021.08.004.

[6] M. A. Elsadig, A. Altigani, and H. T. Elshoush, "Breast cancer detection using machine learning approaches: a comparative study," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 13, no. 1, pp. 736-745, Feb. 2023, doi: 10.11591/ijece.v13i1.pp736-745.

[7] O. J. Egwom, M. Hassan, J. J. Tanimu, M. Hamada, and O. M. Ogar, "An LDA–SVM machine learning model for breast cancer classification," *BioMedInformatics*, vol. 2, no. 3, pp. 345–358, Sep. 2022, doi: 10.3390/biomedinformatics2030022.

[8] O. Tarawneh, M. Otair, M. Husni, Hayfa. Y. Abuaddous, M. Tarawneh, and M. A. Almomani, "Breast cancer classification using decision tree algorithms," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 4, 2022, doi:10.14569/ijacsa.2022.0130478.

[9] M. Ebrahim, A. A. H. Sedky, and S. Mesbah, "Accuracy assessment of machine learning algorithms used to predict breast cancer," *Data*, vol. 8, no. 2, p. 35, Feb. 2023, doi: 10.3390/data8020035.

[10] R. R. Kadhim and M. Y. Kamil, "Comparison of machine learning models for breast cancer diagnosis," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 12, no. 1, pp. 415-421, Mar. 2023, doi: 10.11591/ijai.v12.i1.pp415-421.

[11] V. Birchha and B. Nigam, "Performance analysis of averaged perceptron machine learning classifier for breast cancer detection," *Procedia Computer Science*, vol. 218, pp. 2181–2190, 2023, doi: 10.1016/j.procs.2023.01.194.

[12] H. El Massari, N. Gherabi, S. Mhammedi, Z. Sabouri, H. Ghandi, and F. Qanouni, "Effectiveness of applying machine learning techniques and ontologies in breast cancer detection," *Procedia Computer Science*, vol. 218, pp. 2392–2400, 2023, doi: 10.1016/j.procs.2023.01.214.

[13] M. A. A. Albadr, M. Ayob, S. Tiun, F. T. AL-Dhief, A. Arram, and S. Khalaf, "Breast cancer diagnosis using the fast learning network algorithm," *Frontiers in Oncology*, vol. 13, Apr. 2023, doi: 10.3389/fonc.2023.1150840.

[14] T. O. Omotehinwa, D. O. Oyewola, and E. G. Dada, "A light gradient-boosting machine algorithm with tree-structured parzen estimator for breast cancer diagnosis," *Healthcare Analytics*, pp. 100218–100218, Jun. 2023, doi: 10.1016/j.health.2023.100218.

[15] F. T. Liza, M. C. Das, P. P. Pandit, A. Farjana, A. M. Islam and F. Tabassum, "Machine Learning-based relative performance analysis for breast cancer prediction," *2023 IEEE World AI IoT Congress (AIIoT)*, Seattle, WA, USA, 2023, pp. 0007-0012, doi: 10.1109/AIIoT58121.2023.10174469.

[16] I. Gad, M. Elmezain, M. M. Alwateer, M. Almaliki, G. Elmarhomy and E. Atlam, "Breast cancer diagnosis using a machine learning model and swarm intelligence approach," Jan. 2023, *2023 1st International Conference on Advanced Innovations in Smart Cities (ICAISC)*, Jeddah, Saudi Arabia, 2023, pp. 1-5, doi: 10.1109/ICAISC56366.2023.10085393.

[17] "Breast Cancer Wisconsin (Diagnostic) Data Set," www.kaggle.com. Avaiable online: https://www.kaggle.com/datasets/uciml/breast-cancer-wisconsin-data (accessed: June 2023).

[18] O. Koren, M. Koren, and O. Peretz, "A procedure for anomaly detection and analysis," *Engineering Applications of Artificial Intelligence*, vol. 117, part A, p. 105503, Jan. 2023, doi: 10.1016/j.engappai.2022.105503.

[19] S. Zhang, R. Yao, C. Du, E. Essah, and B. Li, "Analysis of outlier detection rules based on the ASHRAE global thermal comfort database," *Building and Environment*, vol. 234, pp. 110155–110155, Apr. 2023, doi: 10.1016/j.buildenv.2023.110155.

[20] I. S. Thaseen and C. A. Kumar, "Intrusion detection model using fusion of chi-square feature selection and multi class SVM," *Journal of King Saud University - Computer and Information Sciences*, vol. 29, no. 4, pp. 462–472, Oct. 2017, doi: 10.1016/j.jksuci.2015.12.004.

[21] T. E. Mathew, "Breast cancer classification using an extreme gradient boosting model with F-score feature selection technique," *Journal of Advances in Information Technology*, vol. 14, no. 2, pp. 363–372, Jan. 2023, doi: 10.12720/jait.14.2.363-372.

[22] T. Suresh, T. A. Assegie, S. Ganesan, R. L. Tulasi, R. Mothukuri, and A. O. Salau, "Explainable extreme boosting model for breast cancer diagnosis," *International Journal of Power Electronics and Drive Systems (IJPEDS)*, vol. 13, no. 5, pp. 5764–5764, Oct. 2023, doi: 10.11591/ijece.v13i5.pp5764-5769.

[23] B. Charbuty and A. Abdulazeez, "Classification based on decision tree algorithm for machine learning," *Journal of Applied Science and Technology Trends*, vol. 2, no. 01, pp. 20–28, Mar. 2021, doi: 10.38094/jastt20165.

[24] A. K. Singh, "Breast cancer classification using ML on WDBC," *Lecture Notes in Electrical Engineering*, pp. 609–619, Jan. 2023, doi: 10.1007/978-981-99-0189-0_48.

[25] R. Wang, "AdaBoost for feature selection, classification and its relation with SVM, a review," *Physics Procedia*, vol. 25, pp. 800–807, 2012, doi: 10.1016/j.phpro.2012.03.160.

[26] S. Suthaharan, "Support vector machine," *Machine Learning Models and Algorithms for Big Data Classification*, vol. 36, pp. 207–235, 2016, doi: 10.1007/978-1-4899-7641-3_9.

[27] C. Bentéjac, A. Csörgő, and G. Martínez-Muñoz, "A comparative analysis of gradient boosting algorithms," *Artificial Intelligence Review*, vol. 54, Aug. 2020, doi: 10.1007/s10462-020-09896-5.

[28] S. Sun and R. Huang, "An adaptive k-nearest neighbor algorithm," *2010 Seventh International Conference on Fuzzy Systems and Knowledge Discovery*, Yantai, China, 2010, pp. 91-94, doi: 10.1109/FSKD.2010.5569740.

[29] M. Kumar, S. Singhal, S. Shekhar, B. Sharma, and G. Srivastava, "Optimized stacking ensemble learning model for breast cancer detection and classification using machine learning," *Sustainability*, vol. 14, no. 21, p. 13998, Oct. 2022, doi: 10.3390/su142113998.

[30] A. K. Das, S. Kr. Biswas, A. Mandal, A. Bhattacharya, and S. Sanyal, "Machine learning based intelligent system for breast cancer prediction (MLISBCP)," *Expert Systems with Applications*, vol. 242, p. 122673, May 2024, doi: 10.1016/j.eswa.2023.122673.

[31] N. Al-Azzam and I. Shatnawi, "Comparing supervised and semi-supervised Machine learning models on diagnosing breast cancer," *Annals of Medicine and Surgery*, vol. 62, pp. 53–64, Feb. 2021, doi: 10.1016/j.amsu.2020.12.043.

[32] S. M. Malakouti, M. B. Menhaj, and A. A. Suratgar, "ML: early breast cancer diagnosis," *Current Problems in Cancer: Case Reports*, pp. 100278–100278, Jan. 2024, doi: 10.1016/j.cpccr.2024.100278.

[33] W. N. L. W. H. Ibeni, M. Z. M. Salikon, A. Mustapha, S. A. Daud, and M. N. M. Salleh, "Comparative analysis on bayesian classification for breast cancer problem," *Bulletin of Electrical Engineering and Informatics (BEEI)*, vol. 8, no. 4, pp. 1303–1311, Dec. 2019, doi: 10.11591/eei.v8i4.1628.

[34] R. H. Hridoy, A. D. Arni, S. K. Ghosh, N. R. Chakraborty, and I. Mahmud, "Performance enhancement of machine learning algorithm for breast cancer diagnosis using hyperparameter optimization," *International Journal of Power Electronics and Drive Systems (IJPEDS)*, vol. 14, no. 2, pp. 2181–2181, Apr. 2024, doi: 10.11591/ijece.v14i2.pp2181-2190.

[35] 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*, vol. 20, no. 6, p. 5103, Mar. 2023, doi: 10.3390/ijerph20065103.

[36] S. Wang, D. Li, N. Petrick, B. Sahiner, M. G. Linguraru, and R. M. Summers, "Optimizing area under the ROC curve using semi-supervised learning," *Pattern Recognition*, vol. 48, no. 1, pp. 276–287, Jan. 2015, doi: 10.1016/j.patcog.2014.07.025.

[37] Z. Li, "Extracting spatial effects from machine learning model using local interpretation method: An example of SHAP and XGBoost," *Computers, Environment and Urban Systems*, vol. 96, p. 101845, Sep. 2022, doi: 10.1016/j.compenvurbsys.2022.101845.

[38] C. Hu *et al.*, "Application of interpretable machine learning for early prediction of prognosis in acute kidney injury," *Computational and Structural Biotechnology Journal*, vol. 20, pp. 2861–2870, 2022, doi: 10.1016/j.csbj.2022.06.003.

[39] M. Laatifi *et al.*, "Explanatory predictive model for COVID-19 severity risk employing machine learning, shapley addition, and LIME," *Scientific Reports*, vol. 13, no. 1, Apr. 2023, doi: 10.1038/s41598-023-31542-7.

BIOGRAPHIES OF AUTHORS

Monoronjon Dutta is an accomplished computer science and engineering student at Daffodil International University in Dhaka, Bangladesh. He has extensive knowledge and expertise in research, focusing on the complex and complicated fields of machine learning, deep learning, and computer vision. Currently, he is working on many research projects and publishes research papers. He can be contacted at email: monoronjon15-3660@diu.edu.bd and monoronjon.cse@gmail.com.

Khondokar Md. Mehedi Hasan is a dedicated student at Daffodil International University in Dhaka, Bangladesh, where he is enrolled in the Department of Computer Science and Engineering. His career and academic interests focus on software development, with a particular focus on deep learning and machine learning. He actively does research to contribute to the developing field of artificial intelligence since he is motivated by the advancement of technology. He can be contacted at email: mehedi360272@gmail.com and khondokar15-13804@diu.edu.bd.

Alifa Akter is a student at Daffodil International University in Dhaka, Bangladesh, pursuing her studies in computer science and engineering. Her interest mainly lies in specializing in machine learning for her future research. He can be contacted at email: alifa15-3639@diu.edu.bd and alifa.akter255@gmail.com.

Md. Hasibur Rahman is an accomplished computer science and engineering student at Daffodil International University in Dhaka, Bangladesh. He is a web developer with research skills. He has extensive knowledge and expertise in web development, research machine learning, and deep learning. He also has a keen interest in cyber security. He can be contacted at email: hasibur15-3162@diu.edu.bd and hasibur.personal.bd@gmail.com.

Md. Assaduzzaman is a Senior Lecturer in the Department of Computer Science and Engineering, at Daffodil International University, Dhaka, Bangladesh. He has an MSc. in computer science and engineering and a bachelor's degree in computer science and engineering. He has extensive expertise in both teaching and research. His current research interests are machine learning and deep learning. He has published many research papers in several international peer-reviewed journals and conferences. He can be contacted at email: assaduzzaman.cse@diu.edu.bd.