**Enhancing Breast Cancer Diagnosis with XGBoost and SHAP: Toward Transparent and Trustworthy Clinical Decision Support**

Zarin Hadika1[0009-0002-3204-7244], Abdur Rahman Saad2[0009-0003-2078-0702], Mirza Maria Moon3[0009-0002-9201-0478]

1,2,3 Sonargaon University, Dhaka, Bangladesh

lncs@springer.com

**Abstract.** Early detection and accurate classification of breast cancer remain essential challenges in global healthcare, particularly in resource-constrained settings like South Asia. This study presents an interpretable machine learning–based approach for diagnosing breast cancer using the Breast Cancer Wisconsin (Diagnostic) dataset. A scalable and efficient classification system was developed centered on XGBoost owing to its aptness for structured clinical data and support for interpretability. Addressing the paramount need for transparency in clinical decision-making, SHAP (SHapley Additive exPlanations) was integrated to calculate and visualize feature importance, both locally and globally. The findings reveal that perimeter\_worst, area\_worst, and radius\_worst are the features that significantly influence malignancy predictions. The model not only outperformed several established classifiers, including SVM, Random Forest, and AdaBoost, but also offered interpretable outputs to support clinical trust and decision-making. The SHAP-enhanced visualization tools, including summary and force plots, provide clinicians with insights into model rationale, helping bridge the gap between black-box AI systems and practical clinical integration. This study demonstrates the promise of interpretable machine learning to improve diagnostic accuracy, transparency, and workflow efficiency in breast cancer screening and makes the case for broader implementation in real-world clinical practice.

**Keywords:** Breast Cancer, XGBoost, SHAP, Clinical Decision Support, Interpretable Machine Learning, Healthcare Analytics, Predictive Modeling, Explainable AI, Diagnostic Classification, Tabular Data Analysis

1. **Introduction**

Breast cancer is among the most common cancers across the globe. It is one of the most common women's cancers worldwide and a major international health issue. It happens when breast cells grow uncontrollably and abnormally. The early symptoms in a patient often serve as warning signs of the development of cancerous cells in the body. According to the GLOBOCAN 2020 report, cancer is still a major cause of global health problems, with 19.3 million new cases and 10 million deaths reported in 2020[13]. Among these, female breast cancer accounted for 2.3 million new cases, highlighting a concerning rise in incidence and mortality [14]. Statistically, approximately 1 in 8 women gets breast cancer in their lifetime. Breast cancer ranks as the second most common cancer in women in South Asian countries like Bangladesh, and typically happens at a late stage, reducing the chances of effective treatment [15].  
  
Traditional treatment strategies include surgery, chemotherapy, and radiation therapy, each playing a paramount role in disease control. Human frailties in the form of errors, late diagnosis, and inaccessibility of medical services—most forcefully in resource-scarce settings—are, however, still presenting challenges to early and accurate diagnosis. Early diagnosis and accurate staging of breast cancer, therefore, remain central to the improvement of outcomes and reduction of health system burden.  
  
Over the last decade, machine learning (ML) has been one of the promising technologies in medical diagnosis that can recognize faint patterns in clinical and diagnostic data that are not visible to conventional methods.ML models have been effective in tumor detection at early stages, risk assessment for patients, and diagnostic decision support. Most of the current ML-based breast cancer prediction models are black boxes, lacking interpretability and clinical trustworthiness. In addition to that, the majority of the existing studies prioritize the enhancement of classification performance, regardless of the need for transparency and compatibility with actual clinical workflows. This leaves a wide gap between technical achievement in ML and real-world usability in daily clinical practice, especially in low-resource settings.

The purpose of this study is to develop a machine learning–based breast cancer prediction and clinical decision support system. The research objectives are as follows:

1. To identify the most influential features contributing to breast cancer diagnosis using interpretability tools like SHAP.
2. To create and evaluate a valid predictive model that will aid clinicians in diagnosing breast cancer.
3. To develop a clinically applicable decision support system that incorporates machine learning predictions into the diagnostic workflow in a clinical environment.

The following questions direct this investigation:

1. Which are the most significant features to predict malignant vs. benign tumors?
2. How can SHAP or other interpretability methods render model decisions interpretable and understandable to clinicians?
3. What are the strengths and limitations of data-driven predictive models in the clinical setting?

To address these objectives, we utilize the Breast Cancer Wisconsin dataset and apply the XGBoost algorithm for binary classification. For interpretability, we use SHAP (SHapley Additive exPlanations) to highlight the most influential features impacting the model's predictions. We visualize the model explanations using SHAP’s summary\_plot to present overall feature importance, and force\_plot to provide local explanations for individual predictions. These tools help bridge the gap between model accuracy and clinical transparency, making the system more trustworthy and usable in real-world diagnostic settings.

1. **Literature Review**

Shen et al.(2019) developed a CNN model to detect Breast Cancer from biopsies and microscopic images. The efficacy of Shen et al.'s approach was rigorously evaluated on two widely recognized public datasets: the Digital Database for Screening Mammography (DDSM) / CBIS-DDSM and INbreast. For the INbreast dataset, the best single model achieved an impressive per-image AUC of 0.95, with four-model averaging further improving it to 0.98 (sensitivity: 86.7%, specificity: 96.1%), and for the DDSM dataset, the best single model achieved a per-image Area Under the Curve (AUC) score of 0.88, which improved to 0.91 with three-model averaging[1].

Siham et al.(2020) focused on how to preprocess data to deal with imbalanced data that have missing values using resampling techniques to enhance the classification accuracy of detecting breast cancer. The three classifiers (NB, SMO, J48) were tested over original data. The accuracy was respectively 71.67%, 69.58%, 75.52%. After applying a discretization filter and removing the records with missing values, results improved. After that, a resample filter was applied for 7 times, then the accuracy was 98.20%(J48), 76.61%(NB), 95.32%(SMO)[2].

The study presented to detect Breast cancer that machine learning technique is good enough on linear data. But, when the data form is imaged the machine learning technique fails. For the classification of the breast cancer images data Kumar (2021) claimed that a deep learning based technique CNN gives better results as compared to machine learning techniques[3].

Kumar al(2022) trained machine learning models with some python libraries such as numpy, pandas, matplotlib. They trained K-nearest neighbor(KNN), Support vector machine(SVM), Decision tree classifier with that library. They used the Breast malignant growth dataset which was recovered utilizing the UCI archive. The accuracy of those models were 95%(KNN), 96%(SVM) and 93%(DTC)[4].

The study introduced the internal functionality of machine learning algorithms. Abien (2018) used the Wisconsin Diagnostic Dataset to train models. GRU-SVM, Linear Regression, MLP, Nearest Neighbor Softmax Regression, SVM algorithms used to train data. The activation function was used for GRU-SVM as like "Sigmoid" or "ReLU". For Nearest Neighbor used Norm function L1 and L2 both. MLP used ReLU as an activation function. Then, trained the models with 128 batch size and 3000 epochs (Nearest Neighbor's is just 1 epoch). The accuracy of those models was above 90%[5].

Breast al (2020) proposed a new method that Deep Neural Network with Support Value(DNNS) which introduced better quality images and fixed other performance parameters. They discussed the internal process of DNNS. Then, they analyzed the performance according to some other methodology like Naive Bayes, SVM, RCNN classifier, Bidirectional Recurrent Neural Networks.The accuracy of the DNNS method was 97.21 and it was the best result of all of those methods[6].

Liu et al. (2024) developed a clinical decision support tool using SHAP values to predict breast cancer recurrence. The study showed a high predictive accuracy of 0.97 with Extra Trees and 0.96 with Random Forest but also pointed out key limitations, especially concerning data quality and the limited size and diversity of the dataset[7].

Gurcan (2025) introduced an advanced framework for breast cancer diagnosis using deep learning combined with stacking ensemble techniques. The framework incorporates models such as LightGBM, CatBoost, and a CNN-based meta-predictor. The proposed model delivers high accuracy, enhanced F1 scores, and faster training times to support efficient healthcare decision-making. However, the use of complex deep learning and ensemble models increases computational demands, potentially limiting real-time use in resource-constrained settings[8].

Ayepeku (2024) conducted a comprehensive analysis of breast cancer prediction, comparing various machine learning models such as Logistic Regression, Random Forest, Support Vector Classifier, and ensemble methods like Gradient Boosting and AdaBoost. The study uses comprehensive metrics and visualization to assess models but faces typical issues with data quality and availability, while lacking discussion on real-time deployment, which limits practical clinical integration despite strong analytical insights[9].

Patil et al. (2023) investigated early breast cancer prediction by comparing multiple machine learning and deep learning techniques, including SVM, KNN, Naïve Bayes, Logistic Regression, Random Forest, Decision Tree, XGB Classifier, and Artificial Neural Networks (ANN). The study highlights feature selection’s role in boosting accuracy, with SVM reaching 98.24%, but also notes ongoing challenges with data quality and availability. Moreover, the paper highlights model performance comparisons but overlooks real-time deployment and clinical integration, limiting practical use for early breast cancer detection [10].

Öznacar & Ergene (2024) examined the potential of machine learning techniques, including AdaBoost, SVM, Random Forest, and Logistic Regression, for early detection and malignancy prediction in breast cancer. The study highlights that the AdaBoost model showed the highest performance reaching 93.60% AUC and 95.65% precision. However, the study emphasizes the limitations of traditional breast cancer diagnostics, advocating for AI-based improvements, yet it overlooks challenges like data bias and the need for broad clinical validation [11].

Rb et al. (2024) proposed a novel approach for early breast cancer prediction using an ensemble of machine-learning algorithms, including KNN, Naive Bayes, SVM, and Decision Tree Classifier, enhancing predictive capabilities and improving early detection and patient outcomes. The authors demonstrated that using ensemble machine learning algorithms enhances early breast cancer detection accuracy, offering practical benefits for regions with limited medical resources. Additionally, the study emphasizes that Logistic Regression offers efficient and interpretable results, while SVM excels in accuracy with high-dimensional data [12].

**3 Methodology**

**3.1** **Data Preprocessing**

The Breast Cancer Wisconsin (Diagnostic) Data Set is used for this research. Since the dataset is highly systematic, it does not contain any missing values. The target feature "Diagnosis" was encoded using the LabelEncoder() class, where malignant cases were encoded as 1 (positive) and benign cases as 0 (negative). Min-Max scaling is then used to normalize and standardize features to the same range for every continuous variable in [0, 1]. Scaling your data enables your algorithms, such as distance-based and decision-tree-based algorithms like XGBoost, to learn faster and make better predictions.

**3.2** **Model Selection**

XGBoost (eXtreme Gradient Boosting) is selected as the classification algorithm since it works well with structured and tabular data. XGBoost is very accurate, possesses regularization methods to stop overfitting, and can handle linear and non-linear patterns very effectively. It also allows feature importance analysis, which would make it appropriate for interpretability research. Besides, the model is very efficient and scalable and can handle small- and large-scale datasets.

**3.3** **Model Training and Validation**

The data is divided into training and testing sets, where 60% is utilized for training and 40% is kept for testing. The division is done to test the model's generalization capability on new data.

**3.4** **Evaluation Metrics**

The performance of the machine learning model on the Wisconsin Breast Cancer dataset was evaluated based on several existing classification metrics. Indeed, these metrics are very important in clinical diagnostics, where either FPs or FNs may result in severe medical consequences.

* **Accuracy:** The accuracy is the correct prediction ratio over the total number of predictions. Despite being employed extensively, this measure is not enough when dealing with imbalanced class scenarios, e.g., breast cancer datasets, where the proportion of benign subjects may be much higher than that of malignant ones.
* **Precision:** Precision is another key metric to consider, especially for categorizing problems, and one needs to grasp how it works in conjunction with other metrics such as accuracy and recall. Accuracy measures out of all cases, the model has predicted as malignant, how many were actually malignant.
* **Recall:** Recall (also referred to as sensitivity or true positive rate) is the proportion of how many true malignant cases that are identified correctly. It is the proportion out of all the cases of malignant tumors, identified correctly by the model.

* **F1-score:** F1 score is an essential measure in machine learning, particularly for classification issues. Since it gives a sole balanced measure that takes into account both precision and recall.
* **ROC-AUC:** Discrimination between malignant and benign classes at any classification threshold provided is measured by ROC-AUC for a model. Higher AUC indicates higher discriminative power and is useful to measure ranking power for binary classifiers in the clinical domain.
* **Confusion Matrix:** Confusion matrix gives an overview of model classification predictions in the form of true positives, true negatives, false positives, and false negatives. It distinctly tells us the kind of error that has been committed, which is very useful during diagnosis in medicine. In predicting breast cancer, false negatives (false malignant cases) should be minimized, whereas false positives should be reduced to prevent unnecessary procedures.

**4 Experimental Setup**

**4.1 Development Environment**

The implementation was done in Python, with the major libraries being Scikit-learn for machine learning models and evaluation metrics, Pandas for data manipulation. All experiments were run on a typical personal computer with Windows 10, an Intel Core i5 processor, and 8 GB of RAM.

**4.2 Reproducibility protocol**

For reproducibility, the random\_state parameter was always set to 32 in all the concerned functions, such as data splitting and model training. No other control of randomness or cross-validation was performed. The code has a simple and clear workflow with popular open-source libraries.

**5 Results and Discussion**

**5.1 Model Performance**

The XGBoost classifier was evaluated using the following standard metrics. This model had a strong predictive power on the Wisconsin Breast Cancer dataset.

The confusion matrix (Fig. 1) demonstrates that there were 133 TP benign cases and 85 TP malignant cases detected by the model, with only 4 FP and 6 FN. It further indicates that good reliability, especially in minimizing the proportion of inaccurate instances, is a property highly valued in medical diagnostics.

A chart of a variety of colored squares

AI-generated content may be incorrect.

**Fig.1.** Confusion Matrix

The ROC curve (Fig. 2) demonstrates the classifier's capacity for separating malignant from benign cases with an Area Under the Curve (AUC) score of 0.99—a good sign of outstanding discriminative performance.

A graph of a function

AI-generated content may be incorrect.

**Fig.2.** ROC-AUC Curve

Performance of the suggested model was trained and tested by using popular classifiers such as Support Vector Machine (SVM), Decision Tree (DT), Random Forest (RF), Gradient Boosting Classifier (GBC), K-Nearest Neighbors (KNN), Logistic Regression (LR), and AdaBoost. The performance measure for the proposed model from these comparative evaluation contexts is shown in Table 1.

**Table 1.** Performance Comparison of Different Classification Models

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model** | **Accuracy** | **Precision** | **F1-Score** | **Recall** |
| SVM | 90.35% | 93.67% | 87.06% | 87.06% |
| Decision Tree | 90.79% | 89.77% | 88.27% | 86.81% |
| Random Forest | 94.74% | 93.41% | 93.41% | 93.40% |
| Gradient Boost | 95.61% | 96.55% | 94.38% | 92.31% |
| K-Nearest Neighbors | 90.35% | 89.65% | 87.64% | 85.71% |
| Logistic Regression | 90.78% | 91.67% | 88% | 84.61% |
| AdaBoost | 95.17% | 94.44% | 93.92% | 93.41% |
| **XGBoost** | **95.61%** | **95.51%** | **93.41%** | **94.44%** |

The assessment indicated that XGBoost exhibited better balanced performance by posting the best outcomes for accuracy and precision, along with recall among all models tested. The 94.44% F1-score of XGBoost illustrates its capability to control both false positives and false negatives, thereby offering reliable clinical decision support in the diagnosis of breast cancer.

**5.2 Feature Importance & Interpretability**

In order to explain the XGBoost model's predictions and the most contributing features, SHAP (Shapley Additive Explanations) was used. SHAP provides both global and local interpretability by assigning a value to each feature for its contribution to the output of the model.

**Global Interpretability**

Global feature importance was assessed using a SHAP summary plot. This visualization (see Fig. 3) displays the distribution of SHAP values for each feature across all samples. The SHAP value, representing a feature's magnitude and direction of contribution to the prediction, is the x-axis, and the color gradient represents the feature value, ranging from low (blue) to high (red).A screen shot of a graph

AI-generated content may be incorrect.

**Fig.3** SHAP Summary Plot showing the global feature importance based on SHAP values

According to the plot, perimeter\_worst, texture\_mean, and area\_worst have the highest influence on model predictions. These features, associated with tumor size and texture, consistently show strong SHAP values, suggesting a substantial contribution to malignancy classification. Additional impactful features include concave points\_mean, radius\_worst, and area\_se, all of which display wide spreads of SHAP values, indicating frequent involvement in shaping model output. In contrast, features such as fractal\_dimension\_se, texture\_se, and perimeter\_mean exhibit lower SHAP magnitudes, implying less influence overall.

**Local Interpretability**

Local interpretability was further examined using a SHAP force plot, which illustrates how individual feature values contribute to a single prediction. The force plot (see Fig. 4) shows features that push the prediction toward either the malignant or benign class.



**Fig.4.** SHAP Force Plot explaining individual prediction

Here, features such as compactness\_se = 0.01236, radius\_worst = 17.38, area\_worst = 932.7, and perimeter\_worst = 113.7 exert positive SHAP values (in red), thereby increasing the likelihood of a malignant classification. Conversely, features like texture\_mean = 12.91, area\_se = 20, and concave points\_mean = 0.03157 contribute negatively (in blue), reducing the malignancy probability.The total SHAP value for this instance is –2.73, pushing the prediction towards the benign class despite several strong malignancy features.

**5.3 Implications for Clinical Practice**

These findings show the potential usefulness of machine learning models, XGBoost in this instance with interpretable SHAP outputs, to support clinical decision-making in the diagnosis of breast cancer. Through the identification of the most predictive features, e.g., perimeter\_worst, area\_worst, and radius\_worst—the model underscores biologically and clinically meaningful tumor features related to malignancy.

The incorporation of these types of models into clinical workflows can have various functions:

* **Early Detection:**

The patient model also serves as a decision-support system by recognizing cases at heightened risk using objective morphological and morphometric evaluation from imaging modalities. It offers marked benefits to oncologists and radiologists through early identification of potentially suspicious lesions, which can help improve patient outcomes and shorten diagnostic procedure duration.

* **Risk Stratification:**

They would then be stratified into several risk groups using model-estimated probabilities, and then would facilitate clinicians to strike a balance between additional diagnostic workup (e.g., biopsy or surveillance imaging) among patients at greater predicted risk.

* **Transparency & Trust:**

SHAP values also offer transparency regarding how individual features contribute to making specific predictions. Such interpretability leads to increased trust and acceptability by clinicians, especially in high-stakes settings where model explainability is very important.

* **Resource Optimization:**

By excluding low-risk cases with high confidence, the model can prevent redundant diagnostic tests, thereby optimizing clinical resources and reducing patient anxiety.

**5.4 Limitations**

There are a number of limitations to the research that can restrict the generalizability of its results. It was trained on a pre-defined dataset with pre-extracted features, limiting the potential for it to capture real-world clinical diversity. It is unable to learn visual diagnostic patterns since it is not provided with raw image data, for example, mammograms and histopathology slides. It only gets static, single-modality input, denying it temporal and multi-source clinical information that would maximize predictive performance along with clinical utility.

**6 Conclusion and Future Work**

**6.1 Summary of Contributions**

This study offers an XGBoost-based, data-driven approach to the early diagnosis of breast cancer, augmented by SHAP-driven interpretability to aid model transparency. On the Breast Cancer Wisconsin dataset, the model demonstrated outstanding predictive capability with 95.61% accuracy and 0.99 AUC. Comparison with several baseline classifiers confirmed the superiority of XGBoost in controlling false positives and false negatives, which is paramount in medical diagnosis. SHAP value utility not only uncovered insightful predictive features—i.e., perimeter\_worst and area\_worst—but also initiated global and local explanations that are complementary to clinical reasoning. Collectively, these results set the stage for eventual model deployment within clinical decision support systems (CDSS) to assist in driving improved diagnostic accuracy and workflow efficiency.

**6.2 Research Impact**

The incorporation of explainable machine learning in clinical diagnosis has tremendous potential to enhance the delivery of healthcare. This study closes the gap between clinical usability and predictive performance by focusing on transparency and interpretability. The ability to explain predictions for individual patients enhances clinician trust and allows informed decision-making. Additionally, the approach offers the possibility of resource optimization via patient prioritization according to high risk, with potential implications of early intervention and improved patient outcomes, especially in low-resource settings like Bangladesh and other South Asian nations where delayed diagnosis is still a significant issue.

**6.3 Future Directions**

Looking forward, a number of directions can be pursued to build upon the present work. Clinical validation against real patient data is required to determine the generalizability and usefulness of the model across varied healthcare environments. Deployment in electronic health record (EHR) systems would facilitate automated risk estimation and decision support at the point of care in real time. The model's performance could also be enhanced by integrating multi-modal data, including imaging, genomic data, and longitudinal records, that may be capable of capturing more complicated patterns of malignancy. Further collaboration with clinicians on the design of interpretability interfaces could also facilitate real-world use and trust. Collaboration with hospitals and clinical institutions will also be required in order to pilot test the model in real-world clinical settings and adapt it for wider deployment.

**References**

1. Shen, L., Margolies, L. R., Rothstein, J. H., Fluder, E., McBride, R., & Sieh, W. (2019). Deep learning to improve breast cancer detection on screening mammography. *Scientific reports*, *9*(1), 12495.
2. Mohammed, S. A., Darrab, S., Noaman, S. A., & Saake, G. (2020). Analysis of breast cancer detection using different machine learning techniques. In *Data Mining and Big Data: 5th International Conference, DMBD 2020, Belgrade, Serbia, July 14–20, 2020, Proceedings 5* (pp. 108-117). Springer Singapore.
3. Priyanka, K. S. (2021). A review paper on breast cancer detection using deep learning. In *IOP conference series: materials science and engineering* (Vol. 1022, No. 1, p. 012071). IOP Publishing.
4. Sharma, S., Aggarwal, A., & Choudhury, T. (2018, December). Breast cancer detection using machine learning algorithms. In *2018 International conference on computational techniques, electronics and mechanical systems (CTEMS)* (pp. 114-118). IEEE.
5. Agarap, A. F. M. (2018, February). On breast cancer detection: an application of machine learning algorithms on the wisconsin diagnostic dataset. In *Proceedings of the 2nd international conference on machine learning and soft computing* (pp. 5-9).
6. Vaka, A. R., Soni, B., & Reddy, S. (2020). Breast cancer detection by leveraging Machine Learning. *Ict Express*, *6*(4), 320-324.
7. Liu, Y., Fu, Y., Peng, Y., & Ming, J. (2024). Clinical decision support tool for breast cancer recurrence prediction using SHAP value in cooperative game theory. *Heliyon*. <https://doi.org/10.1016/j.heliyon.2024.e24876>
8. Gurcan, F. (2025). *Enhancing Breast Cancer Prediction Through Stacking Ensemble and Deep Learning Integration*. <https://doi.org/10.7717/peerj-cs.2461>
9. Ayepeku, O. F. (2024). Analysis and Visualization of Breast Cancer Prediction through Machine Learning Models. *Jurnal Sistem Informasi*, *13*(3), 1178. <https://doi.org/10.32520/stmsi.v13i3.4100>
10. Patil, S. B., Patil, R. V., & Mahalle, P. N. (2023). Early Breast Cancer Prediction using Machine Learning and Deep Learning Techniques. *International Journal on Recent and Innovation Trends in Computing and Communication*. <https://doi.org/10.17762/ijritcc.v11i10s.7603>
11. Öznacar, T., & Ergene, N. (2024). A Machine Learning Approach to Early Detection and Malignancy Prediction in Breast Cancer. *International Journal of Computational and Experimental Science and Engineering*. <https://doi.org/10.22399/ijcesen.516>
12. Rb, S., V, S., U, M., & Ts, Y. (2024). *Early Breast Cancer Prediction Using Machine-Learning Algorithms*. 1–7. <https://doi.org/10.1109/adics58448.2024.10533612>
13. Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians, 71*(3), 209–249. <https://doi.org/10.3322/caac.21660>
14. Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., & Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians, 68*(6), 394–424. <https://doi.org/10.3322/caac.21492>
15. International Agency for Research on Cancer. (2022). *Global Cancer Observatory*. <https://gco.iarc.fr>