Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Comparative Analysis of Explainable AI Methods Across Machine Learning Classifiers for Breast Cancer Detection

Rajdeep Mondal

rajdeepm180@gmail.com

Chandigarh University

Rohan Kumar Mondal

Chandigarh University

Ansh Jamwal

Chandigarh University

Research Article

Keywords: Explainable AI, SHAP, LIME, Breast Cancer Detection, Logistic Regression, Random Forest, Interpretability

Posted Date: March 25th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9210604/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

Comparative Analysis of Explainable AI Methods Across Machine Learning Classifiers for Breast Cancer Detection

Rajdeep Mondal,Rohan Kumar Mondal,Ansh Jamwal
Department of Computer Science and Engineering
Chandigarh University
Mohali,India
rajdeepm180@gmail.com

***Abstract*—Breast cancer has been one of the leading causes of death among women in the world and the importance of early and precise detection of the disease has been of utmost importance in the world. Machine learning models have been proven to be highly predictive when applied in medical diagnosis; their inability to offer interpretability can however act as a barrier when it comes to their use in clinical practice. Explainable Artificial Intelligence (XAI) systems, including SHAP (Shapley Additive Explanations) and LIME (Local Interpretable Model-agnostic Explanations) have become one of the most effective to combat the issue. This work provides a comparative study of SHAP and LIME on two popular classifiers- Logistic Regression and random forest on the Wisconsin Breast Cancer data set. Along with measuring the performance of a model, the paper presents a quantitative measure of agreement to evaluate the reliability of SHAP and LIME in the presence of important features. The experimental findings demonstrate that the two approaches reach a great degree of consensus (approximately 77 percent), which would lead to consistent and high-quality explanations, and slight deviations can be observed in more complicated models. These results indicate the complementary relationship between SHAP and LIME and their possibility in improving medical AI system transparency.**

Keywords—Explainable AI, SHAP, LIME, Breast Cancer Detection, Logistic Regression, Random Forest, Interpretability

I. INTRODUCTION

Breast cancer is one of the most common forms of cancer in women around the world and the earlier it is detected the more chances one stands to survive. Medical diagnosis has extensively been approached with machine learning techniques because of the capacity to find complex patterns of data [1]. Logistic Regression and Random Forest are examples of models that have proven to perform highly in predicting breast cancer classification tasks [2]. Although most machine learning models are effective, they act like black boxes, and clinicians find it hard to comprehend wether they are rationale in making predictions. This interpretability is a serious obstacle to their implementation in healthcare facilities, which require transparency and trust [3]. Explainable Artificial Intelligence (XAI) has become one of the answers to this issue. The most popular XAI are SHAP, which offers both global and local features importance, as a game theory [4], and LIME, which interprets single predictions through game theory [5].

II. RELATED WORK

Of recent years, the field of explainable artificial intelligence (XAI) as it applies to healthcare has been thoroughly studied. SHAP is a coherent model prediction interpretation model based on Shapley value that was proposed by Lundberg and Lee [4] as a model offering both global and local interpretability. Ribeiro et al. [5] suggested LIME that explains the individual predictions using local surrogate models and has been most commonly used in model-agnostic explanations.

XAI has gained much attention in medical diagnostics. Tjoa and Guan [8] reviewed explainable AI in the healthcare sector, and came to the conclusion that explainable AI is essential in developing the trust of the medical decision-making machine. Similarly, Holzinger et al. [9] claimed that certain fields such as medicine which are high stakes need interpretable models.

Wisconsin Breast Cancer dataset [6] has gained a lot of use in testing machine learning models in medical diagnosis tasks. Random Forest is an ensemble learning algorithm which is a decision tree based classification problem method and has been widely researched on [11]. Also, ROC analysis is generally applied to analyze the performance of classification models [10].

Other comparisons of XAI methods have also been made recently. Molnar [12] has given an elaborate discussion of interpretable machine learning methods, such as SHAP and LIME. Carvalho et al. [13] reflected upon the interpretation methods and mentioned the drawbacks and advantages of their use in practice. Also, reliability and robustness of LIME explanations have been studied as in the case of Slack et al. [14].

Nevertheless, despite these developments, the majority of current studies are interested in utilizing XAI methods individually, and not comparatively. It still lacks quantitative assessment systems of agreement among the approaches to explanation. In this paper, the gap is sealed by presenting an agreement-based comparison between SHAP and LIME between various classifiers.

III. METHODOLOGY

A. Dataset

This research will utilize Wisconsin Breast Cancer dataset [6] that has 569 samples and 30 numerical characteristics of tumors. The data is popular in benchmarking the classification models in medical studies.

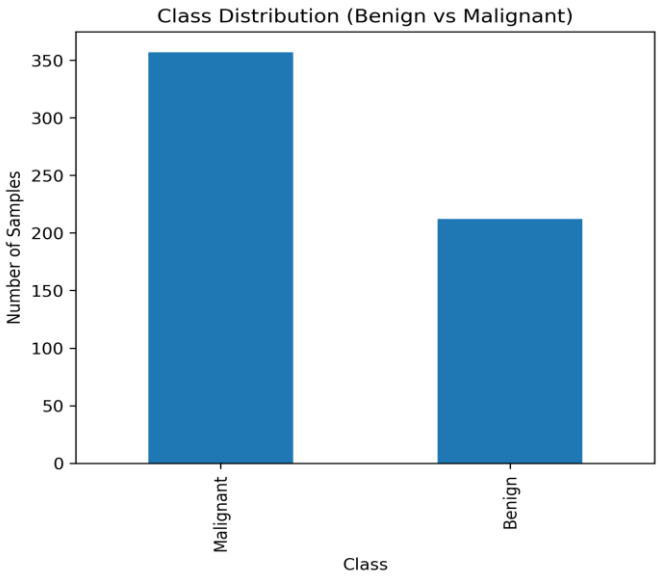


Fig. 1. Class Distribution of Breast Cancer Dataset (Benign vs Malignant)

B. Data Preprocessing

The dataset was split into training and testing sets using an 80–20 ratio with stratification to preserve class distribution. Feature scaling was performed using StandardScaler to normalize input features. The models were implemented using the scikit-learn library [7].The model used Accuracy, precision, recall, F1-score and ROC-AUC to determine performance

C. Machine Learning Models

Two classifiers were used in this. First Logistic Regression which is a linear model known for its interpretability and second Random Forest which is an ensemble learning method based on decision trees [11] D. Explainable AI Methods

D. Explainable AI methods

- SHAP: SHAP explains model predictions by assigning contribution values to each feature based on Shapley values from cooperative game theory [4].
- LIME: LIME generates local explanations by approximating the model with an interpretable surrogate model around a specific prediction [5].

E. Aggrement Metric(Proposed)

To quantify the similarity between SHAP and LIME methods top 10 features were extracted from each method and the overlap between features was computed

Agreement score:

Agreement = |SHAP ∩ LIME| / 10

This process was repeated across all test samples, and the average agreement was calculated.

IV. RESULTS

A. Model Performance

Table I. Model Performance Comparison

Model	Accuracy	Precision	Recall	F1-score	ROC-AUC
Logistic Regression	0.982	0.986	0.986	0.986	0.995
Random Forest	0.956	0.958	0.972	0.965	0.993

B. Confusion Matrix

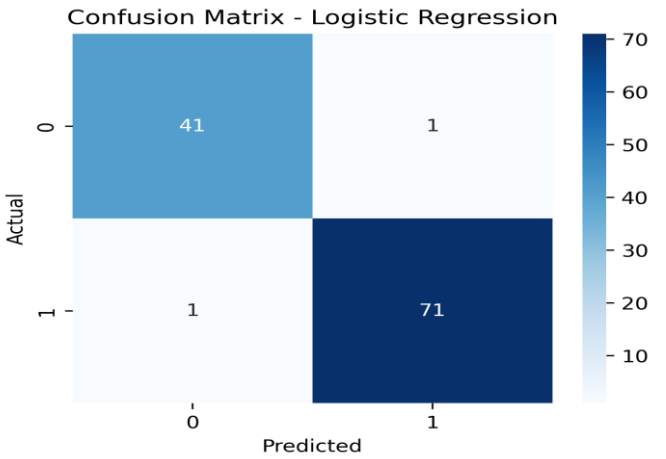


Fig. 2. Confusion Matrix for Logistic Regression

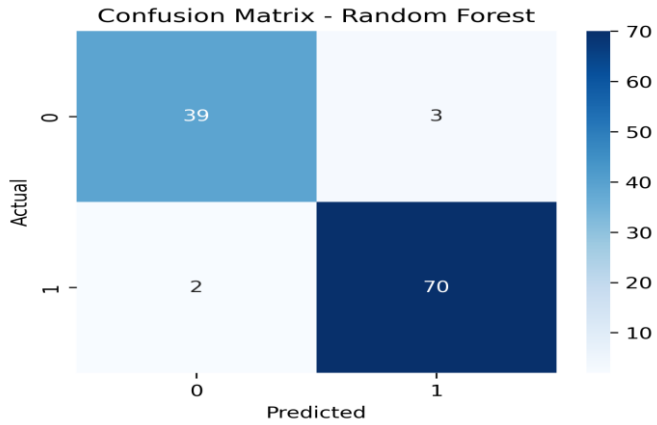


Fig. 3. Confusion Matrix for Random Forest models.

C. ROC Curve

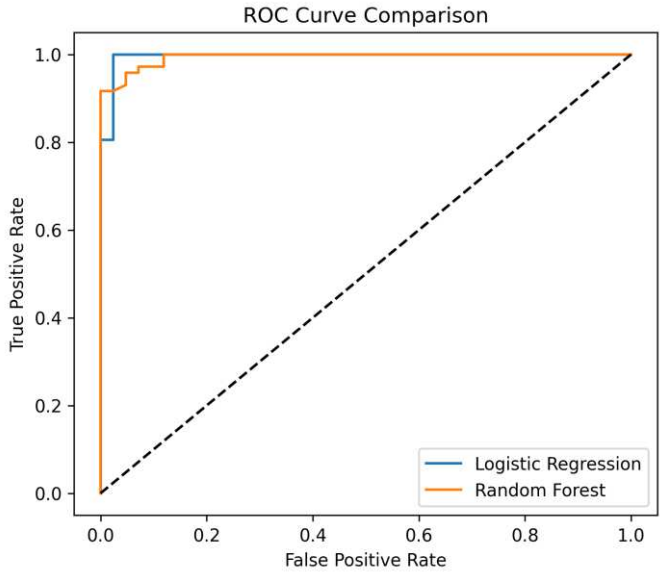


Fig. 4. ROC Curve comparison between Logistic Regression and Random Forest.

ROC curves are widely used to evaluate classification performance [10].

D. Shap Analysis

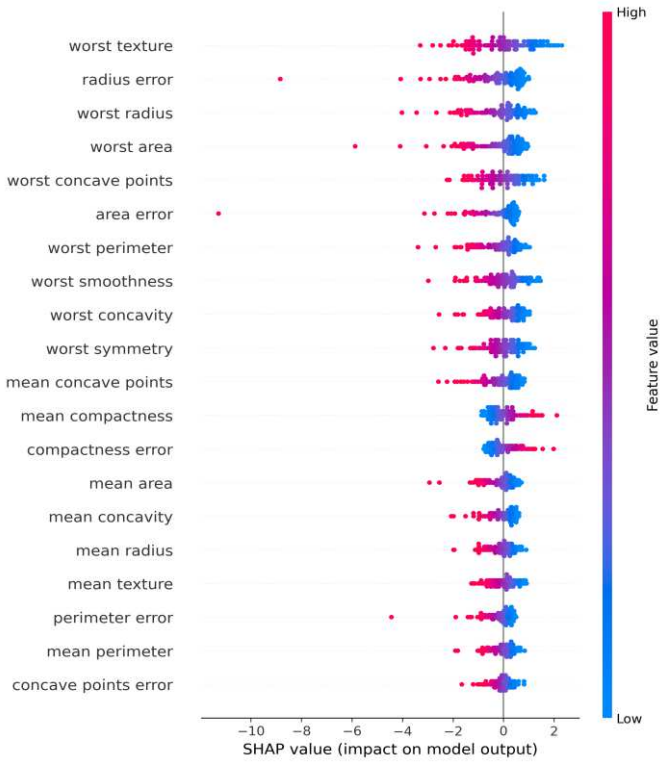


Fig. 5. SHAP beeswarm plot for Logistic Regression model.

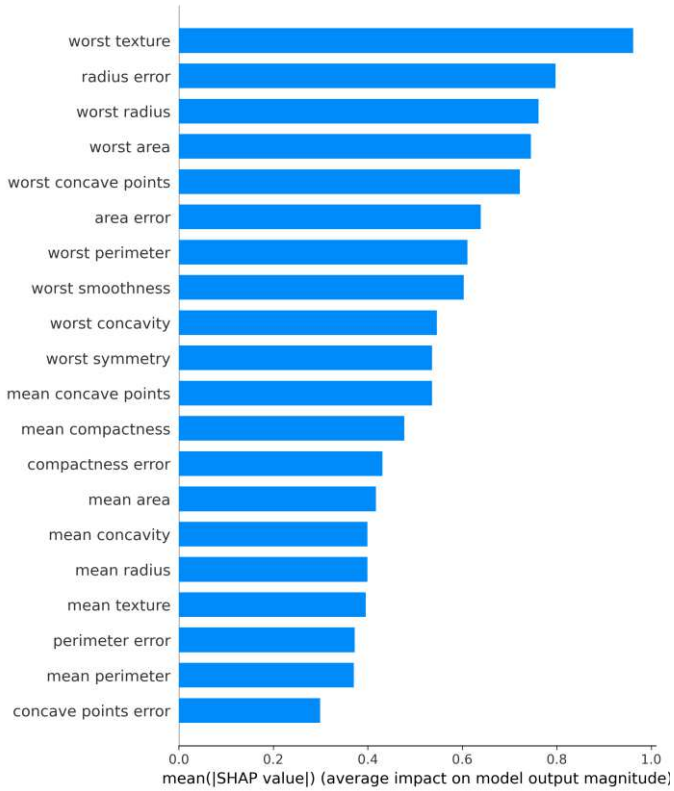


Fig. 6. SHAP bar plot for Logistic Regression model.

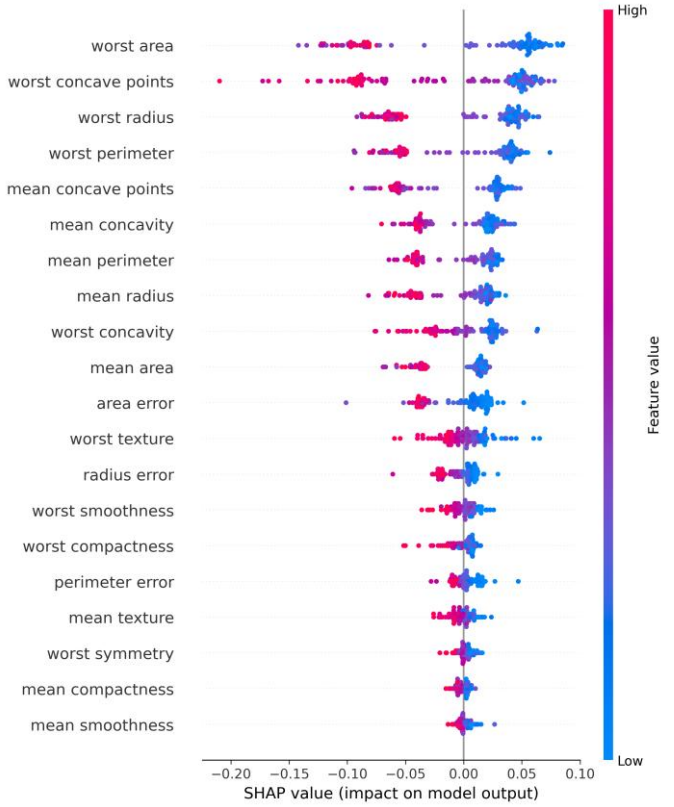


Fig. 7. SHAP beeswarm plot for Random Forest model.

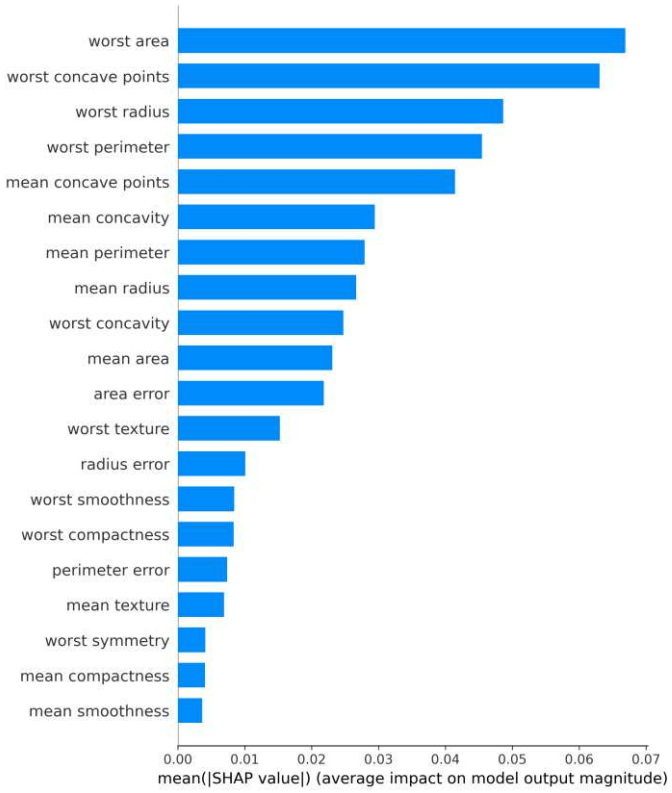


Fig. 8. SHAP bar plot for Random Forest model.

E. Lime Analysis

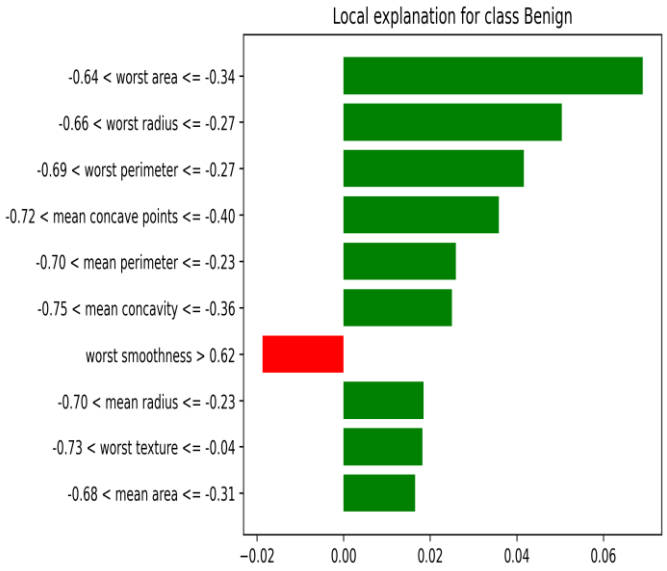


Fig. 9. LIME explanation for a sample instance using Random Forest model.

F. Aggrement Analysis

The agreement between SHAP and LIME explanations was evaluated by comparing the top 10 contributing features for each model.

The average agreement scores were found to be 0.7632 for Logistic Regression and 0.7825 for Random Forest, indicating a high level of consistency between the two explainability methods.

V. Discussion

The results of the experiment confirm that SHAP and LIME have a significantly strong level of consistency in the explanations given by the experimented models, with the agreement scores being above 75. This finding is consistent with previous literature that propose that various explainability methods tend to arrive at similar patterns of feature importance especially on well-organized tabular data [12], [13].

The agreement of the random Forest is a bit higher than that of Logistic Regression, which can be explained by the fact that it is capable of capturing a complex interaction between features, which are always detected by SHAP and LIME. As a linear model, the contribution of features in Logistic Regression is stable, but due to local fluctuations in explanations, it may be slightly reduced.

SHAP is also a global view as it considers the contribution of features to the dataset which is consistent and interpretable. On the contrary, LIME targets local approximations of single predictions, which can be different depending on the sampled neighborhood. This inherent variance is the reason behind the slight differences that are evident between the two approaches.

The results indicate that SHAP and LIME are complementary methods, and the integration of the two can provide a more detailed picture of the model behavior.

VI. Conclusion

This research paper provided a comparative evaluation of SHAP and LIME to detect breast cancer with the help of Logistic Regression and Random Forest models. The experimental findings reveal that both of the explainability methods are reliable and consistent in their accountability with agreement scores of more than 75% in both models.

Random Forest had a little more agreement (0.7825) than Logistic Regression (0.7632), meaning that SHAP and LIME give stronger patterns of feature importance in non-linear models with intricate interaction between features. All in all, the results prove that both approaches are efficient to interpret machine learning models in healthcare use.

The given proposed agreement measure offers a quantitative representation of consistency in the explainability methods and can be applied to other fields and architecture of models. In further studies, it might be possible to include more XAI techniques and test the method on more extensive and real-world clinical data.

VII. Limitations and Future Work

Limitations are that;first the analysis was conducted on a single benchmark dataset, which may limit the generalizability of the results and second is LIME relies on local

approximations, which may introduce variability in feature importance explanations

Future Research can explore deep learning models to capture complex patterns. Additionaly, advanced explainability methods such as Grad-CAM and integrated Gradients can be incoprporated. Furthermore evaluating the framework in clinical deployment scenarios is an important direction for future work.

VIII. REFERENCES

[1] American Cancer Society, "Breast Cancer Facts & Figures," 2023.

[2] W. H. Wolberg, W. N. Street, and O. L. Mangasarian, "Machine learning techniques to diagnose breast cancer," Cancer Letters, vol. 77, no. 2–3, pp. 163–171, 1994.

[3] F. Doshi-Velez and B. Kim, "Towards a rigorous science of interpretable machine learning," arXiv preprint arXiv:1702.08608, 2017.

[4] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in Proc. Advances in Neural Information Processing Systems (NeurIPS), 2017, pp. 4765–4774.

[5] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why should I trust you? Explaining the predictions of any classifier," in Proc. ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining (KDD), 2016, pp. 1135–1144.

[6] UCI Machine Learning Repository, "Breast Cancer Wisconsin (Diagnostic) Dataset." [Online]. Available: https://archive.ics.uci.edu/

[7] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.

[8] E. Tjoa and C. Guan, "A survey on explainable artificial intelligence (XAI): Toward medical XAI," IEEE Transactions on Neural Networks and Learning Systems, vol. 32, no. 11, pp. 4793–4813, 2021.

[9] A. Holzinger, C. Biemann, C. Pattichis, and D. Kell, "What do we need to build explainable AI systems for the medical domain?" arXiv preprint arXiv:1712.09923, 2017.

[10] T. Fawcett, "An introduction to ROC analysis," Pattern Recognition Letters, vol. 27, no. 8, pp. 861–874, 2006.

[11] L. Breiman, "Random forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.

[12] C. Molnar, Interpretable Machine Learning. Lulu.com, 2020.

[13] D. V. Carvalho, E. M. Pereira, and J. S. Cardoso, "Machine learning interpretability: A survey on methods and metrics," Electronics, vol. 8, no. 8, p. 832, 2019.

[14] D. Slack, S. Hilgard, E. Jia, S. Singh, and H. Lakkaraju, "Fooling LIME and SHAP: Adversarial attacks on post hoc explanation methods," in Proc. AAAI Conf. Artificial Intelligence, 2020.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- shaplrbar.png
- shaprfbar.png