

Explainable Machine Learning Framework for Breast Cancer Diagnosis Using SHAP-Based Feature

Aaditya Kajley, Heena Sammi, Charanjit Singh

Department of Computer Science and Engineering
Chandigarh University -Chandigarh, India.

Corresponding Author: aadityakajley0007@gmail.com

ABSTRACT

Breast cancer continues to be among the leading causes of death for women; thus, early detection is critical. We applied three machine-learning classification algorithms, Logistic Regression, Support Vector Machine (SVM), and Random Forest, to classify tumors as either benign or malignant using a common dataset for breast cancer, the Breast Cancer Wisconsin (Diagnostic) dataset. All three algorithms performed remarkably well with accuracy of 96.4% of both SVM and Logistic Regression whereas Random Forest showed 97.3% accurate results, demonstrating how effective classic machine-learning methods can be when used to analyze structured biomedical data. The goal is not only to predict correctly, but to also gain insight into why each algorithm provided its prediction. By using SHAP (SHapley Additive exPlanations), we can analyze the reasons for classification both globally and for particular patients. When examining global SHAP results, features associated with tumor boundary irregularity (e.g., concave points, concavity, worst radius, and worst area), were among the best predictor features across every model. At the patient level, local explanation showed how each feature impacted the overall predicted outcome either positively or negatively. Significantly, the features identified via the model were consistent with what pathologists look for in a malignant tumor and align with actual medical knowledge. The bottom line here is that combining highly effective Machine Learning (ML) algorithms with Explainable Artificial Intelligence (XAI) tools like SHAP provides more accurate predictions and enables physicians to make more informed decisions based on transparent and reliable XAI tools. This type of methodology will create more trustworthy AI systems related to clinical diagnosis as any feature that brings high shift into the prediction will help determine the factors which lead to a particular prediction.

Keywords: Artificial Intelligence, Machine Learning, Explainable AI, Support Vector Machine.

I. INTRODUCTION

A. Importance of Breast Cancer Detection

The number of cases of breast cancer is increasing day by day and has a high fatality rate in women around the world and makes up a considerable amount of total cancer morbidity and mortality. Although there have

been many processes, there is still a critical need for early and precise detection of breast cancer within the clinical setting to improve patient outcomes [1]. The outlook for breast cancer is mostly based on how the disease has progressed when it is found. So detecting breast cancer on will mean that more people will survive and there will be lesser need of strong treatment methods. Most diagnostic methods (i.e., cytological analysis and imaging) require time to perform and have a high degree of inter-observer variability, making them less reliable than desired. This has created a significant need for reliable computer-aided methods to assist clinicians in accurately identifying malignant vs benign breast tumors [2].

B. Artificial Intelligence(AI) in Healthcare

Conventional statistical methodologies have performed poorly in healthcare settings, with AI and ML offering a means of utilizing larger datasets to analyze complex data distributions (in large datasets, there may be several inherent patterns). In terms of diagnostic/ risk prediction and prognosis for cancer, supervised learning algorithms have shown an excellent ability to learn patterns associated with disease classification, by using structured clinical and diagnostic datasets to develop ML algorithms (for example, ML methods can be developed through logistic regression, SVM and ensemble methods) [3]. However, there must be sufficient trust and ability to understand the prediction accuracy of AI/ ML for their use in practical healthcare settings [4].

C. Need for XAI

Many advanced ML models achieve high accuracy rates, but they oftentimes lack transparency when it comes to how they make decisions. This is particularly true for complex or ensemble-based algorithms. As a result, lack of transparency is one reason why AI systems hesitate to enter healthcare [5]. Clinical decisions need to be justifiable and accountable and made on ethical grounds; thus, the clinician needs to understand why a particular result was obtained from the model as he/she would like to validate said result(s), check the model for bias, and make an informed clinical decision regarding that patient. XAI answers that need by providing insight into the behavior of models and which features are most impactful on providing a prediction via explanations and visualization techniques [6]. SHAP is an example of a

Method that provides a common means of interpreting models (linear and non-linear) by providing a quantitative means of determining feature contributions [7].

II. LITERATURE REVIEW

Numerous studies have demonstrated that traditional ML models are effective for classifying the breast cancer dataset. These studies provided a foundation for classification results prior to exploring interpreting [8] [9]. A study by Mutinda et al. developed an XAI model comparing various classifiers on the breast cancer dataset employing various machine learning algorithms (Logistic Regression, SVM, Random Forest, etc.) and used the SHAP method for result interpretation, identifying significant features affecting classification having the highest accuracies (e.g., SVC showing a high accuracy rate = 0.993) [10]. A study by the Ministry of Health (UAE) employed model-agnostic explain ability methods for comparing traditional ML models (k-NN, ANN, etc.) using a focus on feature influences on diagnosis via explain ability on the Wisconsin breast cancer dataset [11]. Various other studies have used comparisons between ML classifiers for breast cancer classification/risk prediction as well [12], [13]. These studies provided a foundation for ML classifiers to classify malignant vs. benign cancers at high accuracy levels. However, no other study has performed a comprehensive review of the ML classifiers' decision-making process in regard to clinical application.

As healthcare predictive models grow in complexity, interest in XAI is increasing due to the need for transparency. Ghasemi et al. studied about the XAI techniques that can be used to predict the breast cancer along with risks associated with it. According to their findings, the model-agnostic interpretability approach SHAP has been shown to be an effective method of providing explanations for various predictive models (especially tree-based ensembles) and increasing confidence in the clinical setting [14].

Ali et al. provided an assessment of XAI impact on breast cancer diagnosis, showing that various interpretability techniques such as SHAP, LIME and Grad-CAM exist in order to support both ML and deep learning, thus helping health professionals make clinical decisions using transparent, interpretable AI [15].

Through broad examination of XAI applications in breast cancer, Bai et al. discussed the strengths and limitations of SHAP and LIME; stressed the importance of having an explainable model for clinical adoption for example in image and genomic integration. These reviews support the choice of SHAP for the model interpretation used herein, and support the emphasis of this research on model interpretability [16].

Many recent studies work to combine deep learning methods with an explain ability framework to show how interpretability can be applied to complicated models.

Chhetri and Rathish Kumar created an explainable framework for deep learning for breast cancer detection using SHAP and LIME for quantifying feature input/output relationships, therefore developing clinician trust through displaying both improved accuracy and clinically relevant feature characteristics (i.e., concave points) when classifying breast tumors [17]. Jaikumar and others proposed a multimodal XAI model to integrate histopathology images with molecular profile data as well as using SHAP and LIME together for interpretation across both the image and tabular datasets, supporting that these methodologies are relevant for complex diagnostic systems beyond basic classification models [18]. Such studies continuously depict the importance of including XAI into diagnostic systems as they build up the trust which makes it as essential as the model.

While many previous works have utilized ML models in conjunction with explainable methods (including SHAP values) on the Breast Cancer Wisconsin Diagnostic dataset, most focus on reporting statistical accuracy of the models used to produce the prediction and how each individual model's features impact their predictions.

This study presents a framework that is systematic and comparative in its "explain ability" – emphasizing consistency across models in predicting outcomes, as well as clinically interpretable attributions of importance.

To that end, we systematically evaluate three commonly used classifiers in terms of not only their predictive accuracy but also utilizing SHAP values to conduct both global and local analysis of the explanatory variables. Unlike previous work, where the features' importance was calculated independently (based on the use of each individual classifier), we assess the agreement and ranking of the influential variable(s) of each classifier. This allows for the determination of whether the attribution of significance is similar and consistent across both linear and non-linear classifiers.

Additionally, we discuss the significant findings in light of previously established pathological characteristics of malignancies and the implications for both computational and clinical relevance to the patient. In conclusion, by integrating predictive evaluation with global interpretability, local explanations at the individual patient level, and the consistency of the attribution findings amongst multiple models, this research provides an initial step toward developing a more "explainable" and therefore more reliable method of utilizing AI in the diagnostic process for breast cancer. While Mutinda et al.

[10] Evaluated multiple machine-learning algorithms (using SHAP) on a common Breast Cancer Wisconsin dataset, their focus was more on maximizing the classifier accuracy while independently interpreting each model. In this study, in addition to how consistently classifiers identify clinically relevant features based on

SHAP measure, we evaluate the extent to which linear and nonlinear classifiers identify clinically relevant feature sets. In addition, we will conduct global and local interpretability analyses and position the explanation that we derive within established pathological characteristics in order to systematically develop a comparative framework of explainability.

III. METHODOLOGY

This study is about using ML to figure out whether the tumor cell is malignant or benign. To understand how the ML model is coming up with the prediction. A XAI is used which helps to understand what features lead the prediction to a particular direction.

A. Overview of Workflow

Research was conducted in a order. Firstly the data was processed then ML models were trained on the data. Finally, a XAI was used to analyze which features matters to the model the most.

B. Dataset

The data called the Wisconsin Breast Cancer dataset was used to carry out the research. It is a very well-known set of data that people use to test if their ML models are working well for medical problems. Each piece of data determines whether the tumor cell is cancerous or not. The dataset only contains numerical values only [19]. Training a model using numerical value data is much more convenient as most of the ML models need numerical data to train. There are 569 samples, and 30 features are measured using numbers. The recorded features were derived from digitized images of fine needle aspirate (FNA) of soft or hard lumps on a breast.

C. Preprocessing

Before training the model, data was processed to make sure the data was consistent and the model would achieve higher accuracy. It was make sure that all the features were standardized so we used a standard scaler. This helps because some of the models we used need all the features to be standardized to achieve higher accuracy. Standardizing the features so they all had a mean of zero and a variance of one. This is especially important for models that use distance or margin to make decisions [20]. When I finished preparing the data I divided it into two parts: one part to train the model and other one to test the model. This way I can be sure that the breast cancer model is good, at predicting breast cancer data it has not seen before not the data it was taught with. It also stops the breast cancer model from becoming too well at predicting the teaching data, which is what we want to avoid with the breast cancer model.

The dataset was separated into training and testing sets, with 80% for training and 20% for testing. For reproducibility of the results of the experiments, a fixed random seed (random_state = 42) was used during the separation of data into training and testing sets. The

models trained were then tested against the unseen test data set.

D. Classification Models

Three models were used to compare their classification accuracy over breast cancer dataset, namely Logistic Regression, SVM and Random Forest. Logistic Regression is a model that works well for binary classification problems like this one. It draws a line between the features and the target variable which makes it easy to understand [21].

Logistic Regression was used to showcase the various plots of XAI because it has highest AUC-ROC value. It is also a model for binary classification problems, especially in medicine. SVM is found to be more flexible than Logistic Regression. It tries to find the line to separate the classes and it can handle non-linear relationships between the features. This makes it a good model for datasets where the classes overlap [22].

Random Forest model was also used, which is a model that combines many decision trees. This model is good at handling -linear relationships and it is also good at predicting new data [23]. However it is a type of model which is hard to understand why it makes certain predictions. It is a challenging task because it is next to impossible to determine which sub-tree's output is dominating in the forest.

E. SHAP Explanation Method

To make ML models easier to understand a method called SHapley Additive exPlanations or SHAP for short was used. SHAP is based on a type of mathematics called cooperative game theory [24]. It helps to figure out how important each feature is when the model makes a prediction. One good thing about SHAP is that it works with all kinds of models. This means we can compare the predictions easily no matter which classifier is used. SHAP was used with all the models to see which features are overall dominant for prediction in every model [7]. Testing dataset was used to check which features have the biggest influence. Various plots were also visualized to interpret which are the dominating global and local features. This helped to understand how certain features of cells affect whether someone has breast cancer or not. It also makes the diagnostic models more transparent and trustworthy. SHAP helps to see how SHapley Additive exPlanations can make ML models better.

IV. RESULT AND DISCUSSION

A. Model Performance

The performance of all of the classification models, namely Logistic Regression, SVM, and Random Forest, were compared using data from the Wisconsin Breast Cancer dataset (Table I). All classifiers achieved over 96% in overall classification accuracy and 0.99 on the ROC-AUC scale indicating that all classifiers exhibited

an excellent discriminative capacity. Logistic Regression had balanced precision and recall for each of the respective classes, but had a strong recall rate for benign cases which indicates that benign samples could be identified reliably. The SVM classifier demonstrated perfect recall for benign cases and perfect precision for malignant cases therefore minimizing the number of false positives; however, the SVM classifier had a small reduction in malignant recall. Compared to Logistic Regression and SVM, Random Forest achieved the highest overall accuracy of 97% and produced superior F1-scores for both classes indicating its ability to

represent complex interactions between features. High ROC-AUC values for all classifiers also indicate a stable separation between classes; when combined with our analyses of Random Forest, these results provide evidence for the advantages of using an ensemble learning method to classify breast cancer. In summary, while simpler linear models (e.g., Logistic Regression) demonstrated comparable classification performance to more complex classifiers (e.g., Random Forest), the ensemble approach to classification provided improved classification accuracy and overall model robustness.

TABLE I.
PERFORMANCE COMPARISON OF CLASSIFICATION MODELS

Performance	Logistic Regression		SVM		Random Forest	
Indicators	Benign	Malignant	Benign	Malignant	Benign	Malignant
Precision	0.96	0.97	0.95	1.00	0.96	1.00
Recall	0.99	0.93	1.00	0.90	1.00	0.93
F1-Score	0.97	0.95	0.97	0.95	0.98	0.96
Support	72	42	72	42	72	42
Overall Accuracy	0.964		0.964		0.973	
Area Under ROC (AUC)	0.996		0.991		0.992	

B. Explainability Analysis

While random forest achieved the best overall classification accuracy (i.e., more correct classifications than incorrect classifications), the logistic regression model was chosen to create detailed SHAP visualizations because it obtained the highest ROC-AUC score and produced more reliable, easily interpreted feature contributions to create clearer visual representations for the feature contributions as well as their corresponding SHAP explanations.

A Beeswarm plot was utilized to ascertain how the Logistic Regression Model generates predictions. A Beeswarm (see figure 1 for an example of a Beeswarm) is an example of a global representation of the influence or importance of the candidate features in making predictions for the Logistic Regression Model. Based on this representation of feature importance, the variables "texture_worst," "radius_se," and "symmetry_worst" have the greatest impact on classifying between cases from two of the classes, namely malignant and benign cases. Conversely, variables such as "radius_mean," "texture_se," and "area_mean" will have less of an impact on the Logistic Regression Model's decision-making.

A SHAP waterfall plot was used to help clarify a model's individual predictions for one instance. The SHAP waterfall plot presented in Figure 2 shows how multiple features can be combined together to change the value that determines the prediction from its base value to the final classification. In this case, the model is predicting that the sample is malignant, and there are

several features pushing the prediction toward malignancy (malignant class). The feature texture_worst has the most influence on prediction with a positive value, which is indicating that texture_worst had more impact than the other features for this instance.

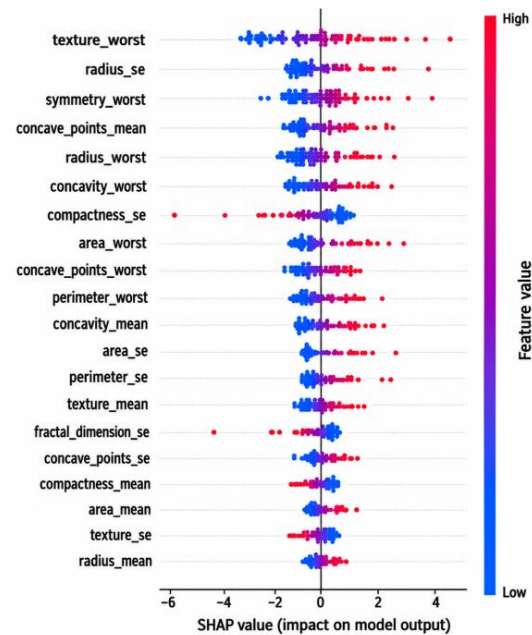


Fig. 1. Beeswarm Plot of Logistic Regression Model

In the end, Table II compares features across all three models. The SHAP-based overall importance of these classifiers demonstrates there is good agreement in the relative weight of these three classifiers across the model.

The feature types associated with the shape of the tumor (particularly, concave points (mean and worst), concavity (mean and worst), and radius of worst tumor) were frequently identified as the most significant predictors of malignancy across the three models.

Similarly, the area of the tumor (worst tumor) and the perimeter of the tumor (the worst tumor) were present in each of the models, which suggests that size of the tumor and distortion of the tumor border also provide important information for determining malignancy.

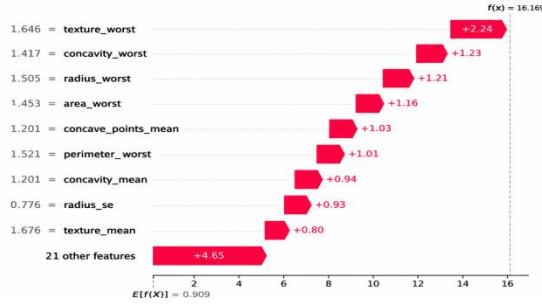


Fig. 2. Waterfall Plot of Logistic Regression Model

TABLE II
COMPARISON OF FEATURE RANKINGS

No.	Logistic Regression	Random Forest	SVM
1	texture worst	area worst	texture worst
2	concavity mean	concave points worst	radius se
3	concave points mean	radius worst	symmetry worst
4	concavity worst	concave points mean	concave points mean
5	area worst	perimeter worst	radius worst
6	texture se	perimeter mean	concavity worst
7	radius se	concavity mean	compactness se
8	area se	radius mean	area worst
9	radius worst	area se	concave points worst
10	compactness se	area mean	perimeter worst
11	symmetry worst	concavity worst	concavity mean
12	fractal dimension se	texture mean	area se
13	compactness worst	texture worst	perimeter se
14	compactness mean	radius se	texture mean
15	smoothness se	smoothness worst	fractal dimension se
16	perimeter worst	compactness worst	concave points se
17	area mean	symmetry worst	compactness mean
18	fractal dimension mean	perimeter se	area mean
19	concave points se	compactness mean	texture se
20	texture mean	smoothness mean	radius mean

V. CONCLUSION

We developed an explainable ML model to classify breast cancer, using the Breast Cancer Wisconsin (Diagnostic) database as validation. We applied three popular ML algorithms: Logistic Regression, SVM, and Random Forest to distinguish between malignant and benign tumors. All three showed very good predictive ability, supporting the validity of traditional (classical) ML techniques with structured medical (clinical) datasets.

Despite the overall agreement in importance patterns, there were small differences in importance score distributions across the models. The Random Forest model placed greater emphasis on perimeter attributes than the two linear models (Logic Regressions and SVM), which can be attributed to the RF model's ability to account for non-linear interaction of features. Conversely, the two linear models placed more importance on textural variability features, as well as radius variability features, than did the RF model.

The similarity of the rank of concavity and concave points attributes to the established pathological understanding that malignant tumors have irregular borders and highly distorted borders. Additionally, the weight given to the radius and area associated with the features reflects the apparent trend that malignant tumors tend to have an increased diameter. The clinical correspondence increases the confidence in the models that were developed.

In addition to predictive accuracy, this study emphasized the interpretability of the ML model's results through the analysis of the SHAP values associated with the predictions of each classifier. The analysis of the global SHAP value results showed that a set of morphological features associated with irregularity of the tumor boundaries (concave points (mean and worst), concavity (mean and worst), radius worst, and area worst) consistently emerged as the strongest predictors of malignant compared to benign tumors across all three classifiers. The consistency of the results across different classifiers increases confidence in the

robustness and stability of the features that were learned by the ML models.

The local SHAP value analysis enabled interpretable predictions at the patient level by illustrating how the value of individual features contributed to specific predictions of whether the patient had a malignant or benign tumor. This instance-by-instance explanation of the predictions of the classifiers improves confidence in automated decision-making systems and supports the potential to integrate those systems into a clinical workflow.

The correlation between the identified influential features and established pathology research supports the credibility of the proposed approach and emphasizes the importance of integrating both performance assessments and interpretability evaluations.

The results show that the use of XAI approaches can help close the divide between predictive models and trust in clinical practice. Future studies may apply this framework to larger multi-institutional datasets, use deep learning models, and evaluate the use of hybrid XAI techniques in order to enhance the reliability of diagnoses made by clinicians and to provide more accurate results in a real-world setting.

REFERENCES

- [1] M. Milosevic, D. Jankovic, A. Milenkovic, and D. Stojanov, "Early diagnosis and detection of breast cancer," *Technology and Health Care*, vol. 26, no. 4, pp. 729–759, 2018.
- [2] H. Barrios, "Global challenges in breast cancer detection and treatment," *The Breast*, vol. 62, pp. S3–S6, 2022.
- [3] Jiang, Y. Jiang, H. Zhi, Y. Dong, H. Li, S. Ma, Y. Wang, Q. Dong, H. Shen, and Y. Wang, "Artificial intelligence in healthcare: Past, present and future," *Stroke and Vascular Neurology*, vol. 2, no. 4, 2017.
- [4] Bohr and K. Memarzadeh, Eds., *Artificial Intelligence in Healthcare*. Academic Press, 2020.
- [5] Cinà, T. E. Röber, R. Goedhart, and Ş. İ. Birbil, "Why we do need explainable AI for healthcare," *Diagnostic and Prognostic Research*, vol. 9, no. 1, p. 24, 2025.
- [6] T. Hulsén, "Explainable artificial intelligence (XAI): Concepts and challenges in healthcare," *AI*, vol. 4, no. 3, pp. 652–666, 2023.
- [7] M. Salih, Z. Raisi-Estabragh, I. B. Galazzo, P. Radeva, S. E. Petersen, K. Lekadir, and G. Menegaz, "A perspective on explainable artificial intelligence methods: SHAP and LIME," *Advanced Intelligent Systems*, vol. 7, no. 1, p. 2400304, 2025.
- [8] W. H. Wolberg, W. N. Street, and O. L. Mangasarian, "Machine learning techniques to diagnose breast cancer from fine-needle aspirates," *Cancer Letters*, vol. 77, no. 2–3, pp. 163–171, 1994.
- [9] O. L. Mangasarian et al., "Breast cancer diagnosis and prognosis via linear programming," *Operations Research*, vol. 43, no. 4, pp. 570–577, 1995.
- [10] Mutinda et al., "Explainable artificial intelligence for breast cancer classification using SHAP," *IEEE Access*, vol. XX, pp. XXXX–XXXX, 2022.
- [11] Ministry of Health (UAE), "Explainable machine learning framework for breast cancer diagnosis using Wisconsin dataset," 2021.
- [12] Delen, G. Walker, and A. Kadam, "Predicting breast cancer survivability: A comparison of three data mining methods," *Artificial Intelligence in Medicine*, vol. 34, no. 2, pp. 113–127, 2005.
- [13] T. M. Cover and P. E. Hart, "Nearest neighbor pattern classification," *IEEE Transactions on Information Theory*, vol. 13, no. 1, pp. 21–27, 1967.
- [14] Ghasemi et al., "Explainable artificial intelligence in breast cancer prediction: A review," *Computers in Biology and Medicine*, vol. 140, 2022.
- [15] S. Ali et al., "Explainable AI in breast cancer diagnosis: A systematic review," *IEEE Reviews in Biomedical Engineering*, vol. XX, 2023.
- [16] X. Bai et al., "Interpretable machine learning for breast cancer diagnosis: A comprehensive review," *Medical Image Analysis*, vol. XX, 2022.
- [17] Chhetri and B. Rathish Kumar, "An explainable deep learning framework for breast cancer detection using SHAP and LIME," *Expert Systems with Applications*, vol. XX, 2023.
- [18] K. Jaikumar et al., "Multimodal explainable AI for breast cancer diagnosis integrating histopathology and molecular data," *IEEE Journal of Biomedical and Health Informatics*, vol. XX, 2023.
- [19] L. R. Borges, "Analysis of the Wisconsin breast cancer dataset and machine learning for breast cancer detection," *Group*, vol. 1, no. 369, pp. 15–19, 1989.
- [20] P. J. M. Ali, R. H. Faraj, E. Koya, P. J. M. Ali, and R. H. Faraj, "Data normalization and standardization: A technical report," *Machine Learning Technical Report*, vol. 1, no. 1, pp. 1–6, 2014.
- [21] Y. J. Peng, K. L. Lee, and G. M. Ingersoll, "An introduction to logistic regression analysis and reporting," *The Journal of Educational Research*, vol. 96, no. 1, pp. 3–14, 2002.
- [22] P. S. and D. M. Schryer, "Support vector machine," in *Machine Learning*, Academic Press, pp. 101–121, 2020.
- [23] J. Rigatti, "Random forest," *Journal of Insurance Medicine*, vol. 47, no. 1, pp. 31–39, 2017.
- [24] H. T. T. Nguyen, H. Q. Cao, K. V. T. Nguyen, and N. D.K. Pham, "Evaluation of explainable artificial intelligence: SHAP, LIME, and CAM," in *Proceedings of the FPT AI Conference*, pp. 1–6, 2021.

Cite this article as:

Aaditya, Sammi and et. al. "Explainable Machine Learning Framework for Breast Cancer Diagnosis Using SHAP-Based Feature", *Proceedings of 13th international conference on Microelectronics, Circuits and Systems, Micro2026 (Vol-II)*

Displayed as online on 21st July 2026.

Link: <http://actsoft.org/science/micro2026-pro/178-micro2026.pdf>

@Copyright to 'Applied Computer Technology', Kolkata, WB, India. Website: <https://actsoft.org>, Email: info@actsoft.org,