Transparent Breast Cancer Diagnosis through Causality, Explainability and Visualization

Yi-Jui Huang[1*], Cheng-Yu Wen[2]

[1]Department of Computer Science and Information Engineering, National Taitung University, Taitung, Taiwan; chris910910s@gmail.com

[2]Department of Computer Science and Information Engineering, National Taitung University, Taitung, Taiwan; littlewen6@gmail.com

*Corresponding Author: chris910910s@gmail.com

DOI: https://doi.org/10.30211/JIC.202503.006

Submitted: Mar. 20, 2025 Accepted: Jun. 03, 2025

ABSTRACT

Breast cancer diagnosis is crucial for improving patient survival rates, yet the explainability of machine learning models remains a significant challenge in clinical applications. This study focuses on feature importance analysis and model explainability in breast cancer diagnosis, highlighting the importance of transparency in medical feature interpretation. By combining FreeViz visualization, SHAP analysis, and LiNGAM causal inference, this research explores key features influencing tumor classification and enhances interpretability in the decision-making process. The results show high consistency across methods, confirming that tumor size, shape irregularity, and boundary morphology are essential in distinguishing malignant from benign tumors. Furthermore, integrating causal inference provides insight into feature interactions and clinical relevance. These findings underscore the value of explainable AI in medical diagnostics, enhancing clinical trust, supporting early detection, and enabling personalized treatment planning. The study contributes to evidence supporting the deployment of interpretable machine learning models in critical healthcare domains.

Keywords: Breast cancer, Causal inference, SHAP explanations, Feature visualization

1. Introduction

1.1 Breast Cancer

Breast cancer is one of the most common malignant tumors among women worldwide [1]. Early diagnosis and accurate prediction are crucial for improving patient survival rates and optimizing clinical decision-making. To facilitate early detection and precise differentiation between benign and malignant tumors, researchers have extensively applied machine learning and data mining techniques in breast cancer diagnosis [2]. Breast cancer arises from the uncontrolled growth of epithelial cells in the ducts or lobules of the breast tissue, and its progression is often associated with complex interactions among genetic, hormonal, and environmental factors [3]. Clinical features commonly used in diagnosis include tumor size, shape, margins, and tissue texture, which can be observed through imaging modalities such as mammography, ultrasound, and magnetic resonance imaging (MRI) [4]. Given the heterogeneity of breast cancer—such as differences among molecular subtypes

like HER2-positive, luminal A/B, and triple-negative—there is an urgent need to build interpretable and robust artificial intelligence (AI) models [5]. These models should not only improve diagnostic accuracy but also provide insights into the underlying biological mechanisms, thus supporting personalized treatment strategies and clinical decision-making [6].

1.2 Medical Imaging and Deep Learning

In recent years, with advancements in medical imaging technology and artificial intelligence (AI), machine learning methods have become essential tools for assisting physicians in diagnosing breast cancer. Among these methods, deep learning techniques, particularly Convolutional Neural Networks (CNNs), have demonstrated outstanding performance in analyzing mammography, ultrasound, magnetic resonance imaging (MRI), and histopathology images [7,8]. These approaches can automatically learn image features, enhancing tumor classification accuracy and reducing the likelihood of human misdiagnosis. Studies have shown that CNN-based models can achieve diagnostic accuracy comparable to that of radiologists in detecting abnormalities in mammograms[9].

In addition, machine learning techniques have been widely applied in breast cancer diagnosis as well as in the analysis of clinical and genomic data. Commonly used models include Support Vector Machines (SVM), Convolutional Neural Networks (CNN), Random Forest, XGBoost, and Deep Neural Networks (DNNs). These methods have been employed to analyze patient history, biomarkers, and gene expression data, demonstrating excellent classification performance and predictive capabilities [10,11]. Furthermore, the integration of multi-omics data—such as genomic, transcriptomic, and proteomic information—further enhances the diagnostic accuracy of these models [12]. However, several practical challenges remain, including data imbalance, lack of interpretability, and difficulties in integrating these models into clinical workflows, all of which limit their scalability and widespread deployment in healthcare systems.

1.3 Importance of Model Explainability

Despite machine learning's success in breast cancer diagnosis, model explainability remains a critical challenge, especially in high-stakes clinical environments. Many state-of-the-art models, particularly deep learning architectures, function as “black boxes” with limited interpretability, which can hinder their integration into routine clinical workflows. The emergence of Explainable AI (XAI) has improved transparency, enabling physicians to better understand model predictions and build trust in clinical practice [13]. Techniques such as Feature Importance Analysis, Local Interpretable Model-agnostic Explanations (LIME), SHapley Additive exPlanations (SHAP), and Grad-CAM are increasingly used to visualize and interpret model outputs [14,15]. These methods help reveal which input features or image regions contribute most to classification, enhancing diagnostic confidence and enabling error analysis.

Moreover, explainability methods can aid in discovering novel imaging biomarkers or confirming known clinical indicators, bridging the gap between data-driven findings and medical knowledge [16]. In breast cancer, where heterogeneous subtypes require tailored treatments, interpretable models help identify subtype-specific features, contributing to more precise treatment

planning and improved patient outcomes. Additionally, regulatory bodies and ethical frameworks increasingly demand AI transparency, particularly in medical applications where explainability supports accountability, fairness, and patient safety [17].

1.4 Goal of the Study

This study aims to evaluate the application of machine learning models in breast cancer diagnosis, with a particular focus on feature importance and model explainability. To achieve this, the study integrates three techniques: FreeViz, SHAP, and LiNGAM. These methods are used to analyze the feature distribution visualization (FreeViz), feature contribution (SHAP), and underlying causal structure (LiNGAM), respectively. Each technique provides complementary information on different levels of model interpretation—namely, visual presentation, local feature importance, and global causal explanation.

This multi-layered explainability framework helps build an AI model with enhanced interpretability, increasing the trustworthiness and transparency of the results in clinical applications. Through this integrative approach, the study aims to identify key factors influencing tumor classification and further improve the accuracy, reliability, and interpretability of breast cancer diagnostic models.

2. Methodology

2.1 Datasets

This study utilizes the publicly available Wisconsin Breast Cancer Dataset (WBCD)[18], a high-quality dataset for breast cancer diagnosis. It contains 569 samples and 30 numerical features, describing the morphological characteristics of cell nuclei, as shown in Table 1.

Table 1. Summary of Breast Cancer Diagnostic Data

Attribute	Count
Total Samples (Cases)	569
Malignant Tumors (M)	212 (37.3%)
Benign Tumors (B)	357 (62.7%)
Number of Features	30
Target Variable	1 (Binary Classification: M / B)
Missing Values	None

2.2 Feature Categories

The features in the Wisconsin Breast Cancer Dataset (WBCD) are derived from the morphological analysis of cell nuclei, obtained through image processing techniques applied to digitized fine needle aspiration (FNA) biopsy images. These features serve as numerical representations of the tumor's physical and structural characteristics, which are critical in distinguishing between benign and malignant cases.

These features can be categorized into three groups:

● Mean Features: Represent the average shape and structure of the cell nucleus across all detected cells in a sample. These features provide a general morphological profile, such as the average radius, texture, and compactness, which helps indicate typical tissue appearance.

● Standard Error (SE) Features: Measure the variability of each morphological characteristic, reflecting how much the features fluctuate across the cell nuclei in a given sample. Higher standard error values may imply structural inconsistency or heterogeneity in the tumor, which is often associated with malignant behavior.

● Worst (Maximum) Features: Capture the most extreme (maximum) values for each morphological attribute within the sample. These features emphasize the most irregular or aggressive cell patterns observed, and are particularly useful in identifying malignancy due to the prominence of atypical nuclei.

Each of the three feature categories contains 10 specific morphological characteristics, resulting in a total of 30 features in the dataset. These core features include measurements such as radius, texture, perimeter, area, smoothness, compactness, concavity, concave points, symmetry, and fractal dimension. These 10 core features are listed in Table 2.

These well-structured categories not only enhance interpretability but also support machine learning models in capturing both the average and the extremes of cellular abnormalities—facilitating more accurate classification and diagnosis. By analyzing features across these three dimensions, clinicians and models alike can better understand both typical and atypical tumor behavior, supporting early detection and personalized decision-making.

Table 2. Description of the 10 morphological features

Feature	Description
radius	radius of the cell nucleus
texture	variation in cell texture
perimeter	perimeter of the cell nucleus
area	area of the cell nucleus
smoothness	smoothness of the cell boundary
compactness	compactness of the cell $(\frac{\text{perimeter}^2}{\text{area}} - 1.0)$
concavity	concavity of the cell nucleus boundary
concave points	number of concave points on the cell nucleus boundary
symmetry	symmetry of the cell nucleus
fractal dimension	fractal dimension of the cell nucleus boundary

2.3 FreeViz

FreeViz (Free Visualization) [19] is a method specifically designed for exploring classified data and is widely used in machine learning and data analysis. Unlike traditional dimensionality reduction techniques such as Principal Component Analysis (PCA) or t-Distributed Stochastic Neighbor Embedding (t-SNE), which aim to preserve the overall variance or local structures in the data, FreeViz focuses on maximizing class separability by optimizing the placement of feature vectors in a 2D space. It treats each feature as a vector in a circular layout, and the position of each sample is projected according to a weighted combination of its feature values. This approach provides a more interpretable visualization of how each feature contributes to class discrimination, which is particularly useful in biomedical domains where interpretability is critical [20].

In Figure 1, the color of the data points represents different sample categories, with red points possibly corresponding to malignant tumors and blue points representing benign tumors. From the distribution of the data, the projection of samples in space exhibits a certain degree of separability, indicating that some feature variables can effectively distinguish between these two types of samples. Furthermore, the direction and clustering of samples in the FreeViz plot can hint at which features are most influential, providing valuable insights for feature selection and model interpretation.

Compared to other visualization methods, FreeViz offers a balance between visualization intuitiveness and class-oriented interpretability, making it well-suited for exploratory analysis in medical datasets.

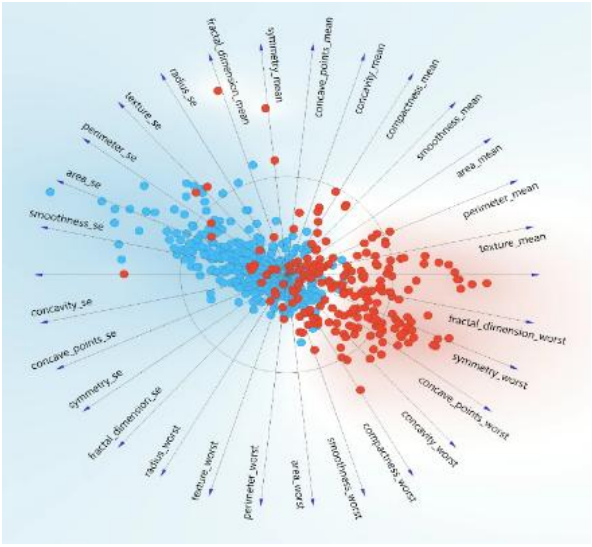


Figure 1. Datasets visualized with FreeViz

2.4 LiNGAM in Causal Inference

This experiment uses LiNGAM [21] for causal inference. LiNGAM (Linear Non-Gaussian Acyclic Model) is a powerful causal discovery method that identifies causal directions between variables based on the assumption of non-Gaussianity, thereby determining direct causal relationships between variables. LiNGAM assumes that the observed variables X follow the following structural equation model (SEM), as shown in Formula 1:

$$X = BX + e \quad \text{[Formula 1]}$$

X: The vector of observed variables.
B: The adjacency matrix representing causal relationships (a directed matrix)
e: The non-Gaussian error term (noise).

The causal structure is reconstructed by estimating the *B* matrix using the Linear Non-Gaussian Acyclic Model (LiNGAM). Once the causal relationships are derived, a Directed Acyclic Graph (DAG) is generated to visually represent the causal structure among variables, as illustrated in Figure 2. The DAG explicitly describes the causal influences between features, offering a clear depiction of the underlying relationships. Moreover, it serves as a crucial interpretability tool, facilitating a deeper understanding of the causal mechanisms within the dataset.

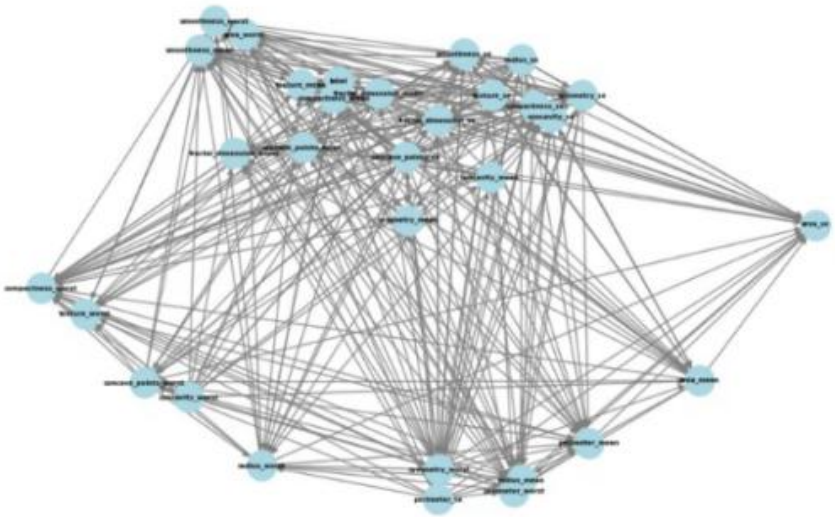


Figure 2. Directed Acyclic Graph

2.5 Model Training

In this study, multiple interpretable machine learning models were trained to ensure that the models maintained high accuracy while also exhibiting good interpretability and transparency. The selected models included Decision Trees [22], Random Forest [23], XGBoost (Extreme Gradient Boosting) [24], LightGBM (Light Gradient Boosting Machine) [25], GDBT (Gradient Boosting Decision Tree) [26], and CatBoost (Categorical Boosting) [27]. These tree-based boosting methods effectively handle high-dimensional data and nonlinear relationships while enabling model decision interpretation through feature importance analysis, thereby enhancing model transparency and reliability.

To evaluate model performance, the dataset was split into training set and testing set at an 8:2 ratio, where the training set was used for model learning and the testing set was utilized to assess the model's generalization ability on unseen data. Several evaluation metrics were employed, including accuracy, specificity, and sensitivity, to comprehensively measure the model's classification performance and its ability to distinguish between different classes. All experimental results are summarized in Table 3.

Table 3. Model Performance Evaluation

Model	Accuracy	Sensitivity	Specificity
Decision Tree	0.947	0.890	0.980
Random Forest	0.964	0.906	1.000
XGBoost	0.964	0.906	1.000
LightGBM	0.964	0.906	1.000
GDBT	0.953	0.890	0.990
CatBoost	0.970	0.921	1.000

2.6 SHAP

SHAP (Shapley Additive Explanations) [28] is employed to analyze and interpret the trained model, assessing the impact of each feature on model decisions. SHAP, grounded in game theory, quantifies feature importance by computing its marginal contribution to different predictions. This approach offers both global interpretability, which provides an overview of feature influence across the entire model, and local interpretability, which explains individual predictions by attributing contributions to specific features.

To further illustrate feature contributions, SHAP values are visualized using summary plots, dependence plots, and force plots, enabling a comprehensive understanding of the model's decision-making process. These visualizations not only enhance model transparency but also help identify potential biases or misleading patterns, thereby ensuring the robustness and reliability of the model in clinical applications.

3. Results

3.1 FreeViz Analysis

Further observation of Figure 1 reveals that malignant tumor samples (red dots) are more concentrated in certain directions, while benign tumor samples (blue dots) are more dispersed, indicating that some features have stronger discriminative power for identifying malignant tumors. This phenomenon suggests that features such as concave_points_worst, radius_worst, and perimeter_mean play a crucial role in the classification of malignant tumors, whereas fractal_dimension_mean and symmetry_se may only provide auxiliary information.

3.2 Causal Analysis

The causal inference derived from the DAG generated by LiNGAM, as shown in Table 4, indicates that the maximum tumor area (area_worst) has the strongest causal effect on the standard error of fractal dimension (fractal_dimension_se), with a causal strength of 5254.00028. This highlights a significant relationship between tumor size and boundary complexity.

Similarly, the mean tumor area (area_mean) has a highly significant impact on the mean fractal

dimension (fractal_dimension_mean), with a causal strength of 5188.94672. This suggests that larger tumors tend to have more complex and irregular boundaries, reinforcing the relationship between tumor size and morphological irregularities.Additionally, the mean tumor area (area_mean) significantly influences the standard error of smoothness (smoothness_se), with a causal strength of 4210.03306. This indicates a correlation between tumor size and variations in boundary smoothness, suggesting that larger tumors may exhibit more fluctuations in smoothness.The maximum tumor area (area_worst) also affects the most severe number of concave points (concave_points_worst), with a causal strength of 1522.23201, suggesting that larger tumors may have more concave edges. Furthermore, the mean tumor area (area_mean) has a certain causal effect on the mean number of concave points (concave_points_mean), with a causal strength of 678.36314. This further supports the relationship between tumor size and boundary morphology characteristics.

Table 4. Causal Strength Analysis of Features

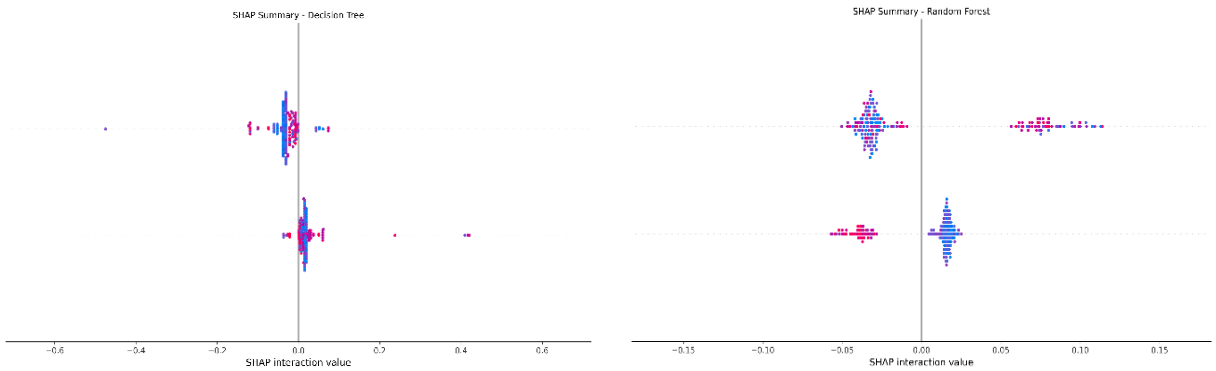
Cause	Effect	Causal Strength
area_worst	fractal_dimension_se	5254.00028
area_mean	fractal_dimension_mean	5188.94672
area_mean	smoothness_se	4210.03306
area_worst	concave_points_worst	1522.23201
area_mean	concave_points_mean	678.36314

3.3 SHAP Interpretation

Due to differences in model architectures and learning strategies, decision trees and random forests primarily rely on SHAP interaction values for model interpretation, whereas gradient boosting decision tree models (such as CatBoost, GBDT, LightGBM, and XGBoost) utilize a stepwise learning mechanism (Boosting) and typically focus on SHAP feature values to evaluate the impact of features on prediction outcomes. In these models, each tree is trained on the residuals of the previous tree, allowing feature interactions to be naturally embedded through iterative learning. Therefore, SHAP feature values alone can effectively reflect both the importance and the directional influence of individual features.

In contrast, a single decision tree makes split decisions based on one feature at a time and lacks the mechanism to capture feature interactions. As a result, its SHAP interaction values are generally close to zero, as illustrated in Figure 3(a). Random forests, composed of an ensemble of independent decision trees, are capable of partially capturing more complex relationships through aggregation, leading to slightly higher SHAP interaction values, as shown in Figure 3(b). However, the level of interaction captured is still less pronounced than that of Boosting-based models.

This analysis helps clarify the appropriate strategies and limitations when interpreting different tree-based models and provides practical guidance on selecting models and matching them with suitable explanation techniques.



(a) (b)

Figure 3. SHAP Interaction Values (a) Decision trees (b) Random forest

Through SHAP analysis, the importance of different features in the prediction results varies across different machine learning models.

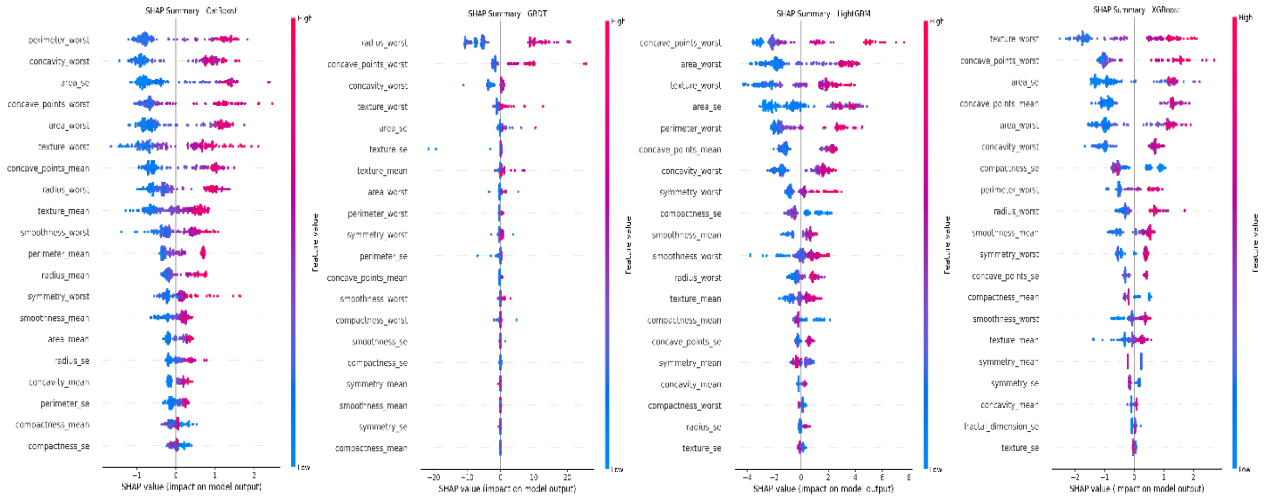
For the CatBoost model, the most significant feature influencing the prediction result is the maximum perimeter value (perimeter_worst), as shown in Figure 4(a). This feature is directly related to the size of the tumor, and a higher perimeter generally correlates with a more malignant tumor. In addition to the perimeter, the maximum concavity value (concavity_worst) also shows significant impact, indicating that the degree of concavity at the tumor's edge plays an important role in the prediction outcome. The standard error of the area (area_se) also influences the prediction results, with higher values typically leading to more severe predicted outcomes.

In the GBDT model, the largest radius (radius_worst) is considered the most influential feature, directly and significantly affecting the prediction results. This feature is closely related to the size of the tumor, as shown in Figure 4(b). The most severe concave points (concave_points_worst) and concavity (concavity_worst) also show considerable impact. The SHAP value range for these features is wide, indicating that the model is particularly sensitive to extreme values, especially the maximum radius, which has a significant effect on the prediction results.

The feature importance in the LightGBM model is focused on the most severe concave points (concave_points_worst) and the maximum area (area_worst), as shown in Figure 4(c). These features have a significant positive impact on the prediction results. Compared to other models, the maximum texture value (texture_worst) has a greater influence on LightGBM, indicating that the model relies more heavily on texture features. The SHAP value range for this feature is wide, showing that LightGBM is more sensitive to extreme values.

In the XGBoost model, the maximum texture value (texture_worst) is the most influential feature among all, as shown in Figure 4(d). This is not commonly observed in other models. The learning approach of XGBoost makes the model more reliant on texture features. Additionally, the most severe concave points (concave_points_worst) also play an important role in XGBoost, while the standard error of the area (area_se) ranks as one of the most influential features.

In summary, although different models place varying levels of importance on different features, some features consistently show high importance across all models. For example, the most severe concave points (concave_points_worst) and concavity (concavity_worst) are considered key predictive factors in all models. Additionally, tumor size-related features, such as the maximum area (area_worst) and maximum perimeter (perimeter_worst), also have significant influence in most models. These results suggest that the structure and size of the tumor are the most influential factors in predictions, while the models differ in their reliance on these features.



(a) (b) (c) (d)

Figure 4. SHAP Values (a) CatBoost (b) GBDT (c) LightGBM (d) XGBoost

3.4 Feature Importance Comparison

Combining the results of FreeViz visualization, LiNGAM causal inference, and SHAP analysis, these three methods consistently highlight similar feature importance rankings, indicating that tumor size and boundary morphology are the most influential factors in breast cancer classification. FreeViz visualization reveals a significant distinction between malignant and benign tumor samples along certain feature dimensions, demonstrating that features such as concave_points_worst, radius_worst, and area_worst are crucial for classification decisions.

LiNGAM causal inference further establishes the causal relationships among these variables, showing that tumor area and boundary morphology are closely related, with larger tumors often exhibiting more concave boundary points. This insight reinforces the idea that these variables are central to predicting malignant tumors. The causal links also suggest that the tumor's structural characteristics directly influence its classification.

SHAP analysis quantifies the impact of these features on different machine learning models and finds that regardless of the model, concave_points_worst and concavity_worst consistently emerge as the most influential features. This underscores the central role of boundary concavity characteristics in diagnosis, demonstrating their key contribution to model predictions across all models.

4. Conclusions

This study demonstrates a high level of consistency in feature importance across three methods—FreeViz visualization, SHAP analysis, and LiNGAM causal inference which not only enhances the reliability of the diagnostic results but also improves the transparency of model decisions. By integrating interpretability techniques, the model's decision-making process is transformed from a traditional "black-box" approach into a more explainable framework rooted in clinically meaningful features.The convergence of results across diverse techniques reinforces the robustness of the identified key predictors, particularly tumor size, shape irregularity, and boundary morphology. These characteristics consistently emerged as critical factors in distinguishing malignant from benign tumors, confirming their clinical relevance and diagnostic significance.

Moreover, the varying sensitivity of machine learning models to specific features reflects the inherent biases and priorities of each algorithm, offering insight into model behavior. Understanding these tendencies supports model selection and tuning while guiding clinicians in integrating AI tools into diagnostic workflows. More broadly, this research underscores the potential of explainable AI to bridge the gap between computational models and clinical practice. By enhancing interpretability and transparency, medical professionals can build greater trust in AI systems, encouraging their adoption in real-world healthcare. Additionally, the identified feature patterns lay a foundation for future research in biomarker discovery, personalized medicine, and hybrid diagnostic systems that combine data-driven insights with clinical expertise.

Ultimately, this work contributes to the advancement of trustworthy machine learning in oncology, and offers both theoretical and practical implications for improving diagnostic accuracy, model accountability, and clinical decision-making.

5. Future Work

AWhile this study demonstrated strong performance on the Wisconsin Breast Cancer Dataset (WBCD), the dataset is relatively small and consists of structured numerical data, lacking imaging information and diverse patient characteristics. This may limit the model's applicability and generalizability in real clinical settings. Therefore, future work should extend to unstructured and multimodal clinical data—such as mammography, ultrasound, and histopathology slides—to evaluate the model's stability and interpretability across different data types.

In addition, incorporating strategies such as self-supervised learning, transfer learning, and multi-center data analysis can further enhance the model's adaptability and robustness under various clinical conditions. Finally, collaboration with clinicians for real-world testing will help assess the feasibility and practical utility of the proposed framework within clinical decision support systems.

These extensions will help validate the model's transferability and clinical value, promote trustworthy AI applications in breast cancer diagnosis, and lay a foundation for future interdisciplinary integration.

References

[1] Sung, H., Ferlay, J., Siegel, R.L., Laversanne, M., Soerjomataram, I., Jemal, A. and Bray, F. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: A Cancer Journal for Clinicians, 2021, 71(3), 209-249.

[2] Esteva, A., Robicquet, A., Ramsundar, B. and others. A guide to deep learning in healthcare. Nature Medicine, 2019, 25(1), 24-29.

[3] Polyak, K. Breast cancer: origins and evolution. The Journal of Clinical Investigation, 2007, 117(11), 3155-3163.

[4] D'Orsi, C.J., Sickles, E.A., Mendelson, E.B. and Morris, E.A. ACR BI-RADS® Atlas, Breast Imaging Reporting and Data System. American College of Radiology, 2013.

[5] Jiang, Y., Yang, M., Wang, S., Li, X., Sun, Y. and He, Z. Interpretable deep learning framework for breast cancer classification from multi-modal data. Information Fusion, 2021, 67, 132-144.

[6] Lundberg, S.M. and Lee, S.-I. A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 2017, 30.

[7] Litjens, G., Kooi, T., Bejnordi, B.E. and others. A survey on deep learning in medical image analysis. Medical Image Analysis, 2017, 42, 60-88.

[8] Leibig, C., Brehmer, M., Bunk, S., Byng, D., Pinker, K. and Umutlu, L. Combining the strengths of radiologists and AI for breast cancer screening: a retrospective analysis. The Lancet Digital Health, 2022, 4(7), e507-e519.

[9] Nassif, A.B., Talib, M.A., Nasir, Q., Afadar, Y. and Elgendy, O. Breast cancer detection using artificial intelligence techniques: A systematic literature review. Artificial Intelligence in Medicine, 2022, 127, 102276.

[10] Tarighati, E., Keivan, H. and Mahani, H. A review of prognostic and predictive biomarkers in breast cancer. Clinical and Experimental Medicine, 2023, 23(1), 1-16.

[11] Hou, Y., Peng, Y. and Li, Z. Update on prognostic and predictive biomarkers of breast cancer. In: Seminars in Diagnostic Pathology, 2022, 39(5), 322-332. WB Saunders.

[12] Huang, S., Chaudhary, K. and Garmire, L.X. More than meets the eye: Biomarker discovery using multi-omics data. Frontiers in Genetics, 2017, 8, 200.

[13] Hulsen, T. Explainable artificial intelligence (XAI): concepts and challenges in healthcare. AI, 2023, 4(3), 652-666.

[14] Ribeiro, M.T., Singh, S. and Guestrin, C. "Why should I trust you?" Explaining the predictions of any classifier. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, 1135-1144.

[15] Selvaraju, R.R., Cogswell, M., Das, A. and others. Grad-CAM: Visual explanations from deep networks via gradient-based localization. Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2017, 618-626.

[16] Holzinger, A., Langs, G., Denk, H., Zatloukal, K. and Müller, H. Causability and explainability of artificial intelligence in medicine. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2019, 9(4), e1312.

[17] European Commission. Proposal for a regulation laying down harmonized rules on artificial intelligence (Artificial Intelligence Act). Brussels: European Commission, 2021.

[18] Wolberg, W., Mangasarian, O., Street, N. and Street, W. Breast Cancer Wisconsin (Diagnostic) [Dataset]. UCI Machine Learning Repository, 1993. DOI: 10.24432/C5DW2B.

[19] Demšar, J., Leban, G. and Zupan, B. FreeViz—An intelligent multivariate visualization approach to explorative analysis of biomedical data. Journal of Biomedical Informatics, 2007, 40(6), 661-671.

[20] Orange Data Mining. FreeViz — Orange Visual Programming 3 documentation. Available online: https://orange3.readthedocs.io/projects/orange-visual-programming/en/latest/widgets/visualize/freeviz.html

[21] Shimizu, S. LiNGAM: Non-Gaussian methods for estimating causal structures. Behaviormetrika, 2014, 41(1), 65-98.

[22] De Ville, B. Decision trees. Wiley Interdisciplinary Reviews: Computational Statistics, 2013, 5(6), 448-455.

[23] Rigatti, S.J. Random forest. Journal of Insurance Medicine, 2017, 47(1), 31-39.

[24] Chen, T. and Guestrin, C. XGBoost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, 785-794.

[25] Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q. and Liu, T.Y. LightGBM: A highly efficient

gradient boosting decision tree. Advances in Neural Information Processing Systems, 2017, 30.

[26] Feng, J., Yu, Y. and Zhou, Z.H. Multi-layered gradient boosting decision trees. Advances in Neural Information Processing Systems, 2018, 31.

[27] Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A.V. and Gulin, A. CatBoost: Unbiased boosting with categorical features. Advances in Neural Information Processing Systems, 2018, 31.

[28] Nohara, Y., Matsumoto, K., Soejima, H. and Nakashima, N. Explanation of machine learning models using SHAPley additive explanation and application for real data in hospital. Computer Methods and Programs in Biomedicine, 2022, 214, 106584.