



A Hybrid Explainable Logistic Regression model with Bayesian Feature Selection, SHAP Explainability, Stability Analysis, and Uncertainty Quantification for Breast Cancer Diagnosis

Khamrunissa Hussain Sheikh¹, Sudheer Kumar Sharma², Narendra Kumar³

¹Research Scholar, Department of Mathematics & Statistics, NIET, NIMS University Rajasthan, Jaipur INDIA.

E-mail: khamarnoo2021@gmail.com,

²Department of Mathematics & Statistics, NIET, NIMS University Rajasthan, Jaipur INDIA. E-mail: sudheersharma.math@gmail.com

³Department of Mathematics & Statistics, IILM University Gr. Noida (U.P.) INDIA. E-mail: drnk.cse@gmail.com

Abstract

Breast Cancer (BC) diagnosis requires predictive models with uncertainty, stability, and interpretability. Predictive models have a much better performance than traditional BC methods. The proposed framework is based on prediction, feature selection, explanation, stability, and uncertainty. This approach has provided us with both probabilistic risk estimation and an explanation of what is behind it. The present research was developed to predict uncertainty and improve clinical decision-making for further evaluation by linking Bayesian modeling with XAI, stability analysis, and uncertainty quantification. The current study mathematically presents BC classification. The framework applied is a hybrid explainable logistic regression model used to classify malignant and benign breast tumors. Moreover, Bayesian sparse feature selection was used with logistic regression, SHAP-based explanation, bootstrap stability analysis, and uncertainty on a single platform. Useful predictors were selected owing to Bayesian sparse priors, which perform model simplification and exclusion of irrelevant features. The SHAP value method was used to determine what global and local features mean in regard to prediction. Finally, bootstrap resampling was applied to carefully assess the strong and stable predictors and regression coefficients.

Keywords: Breast cancer diagnosis, logistic regression, Bayesian selection, SHAP, explainable AI, stability analysis, WDBC dataset, posterior inference, probabilistic classification.

1. Introduction

Breast cancer (BC) is one of the most common cancer types and has become an important global health issue. Early and accurate diagnosis is critical for the treatment of the disease and its prevention from progression and possible mortality. Recently, artificial intelligence (AI) and machine learning (ML) have been used in automatic BC identification, diagnosis, and classification because computational models can reveal complicated relationships between data [9-12]. In particular, there has been an increase in the number of ML models for BC classification using structured data and quantitative characteristics [13-16]. The UCI ML Repository houses the WDBC dataset, which has established itself as one of the most commonly referenced databases for benchmarking various approaches for BC detection [3]. The WDBC dataset includes information and quantitative features obtained from digital images of fine needle aspiration from breast tumors. The main features of the cells are determined based on characteristics of their nuclei like texture, radius, area, perimeter, smoothness, symmetry, concavity, compactness, and fractal dimension [15, 22]. Because the dataset comprises precisely defined predictors and binary outcomes, it is easy to conduct experiments in areas such as feature selection, classification, model interpretability, and uncertainty estimation. The dataset has been successfully utilized for traditional and advanced ML approaches, including vector support machines, logistic regression, and other ensemble learning approaches [23, 24]. Among these methods, logistic regression presents a professional probabilistic model for the medical classification of simple structures, probabilistic outcomes, and interpretable coefficients. The model, which results in only two possible class outputs, logistic regression, allows probability of malignancy predictions instead of binary classifications and adjusts its decision thresholds according to clinical needs. Such a probabilistic model can be used in different situations in terms of false-negative and false-positive outcomes. In addition, conventional logistic regression can lose its effectiveness, especially when the predictors are highly correlated with datasets that include many measurements of similar morphological features from many subjects. Thus, this provokes researchers to apply feature-selection techniques to find those features that provide information and reduce number of predictor variables used in the model. Bayesian selection provides a natural scheme to solve the problem that arises because of the uncertainties related to the selection of factors to be included in the model. Bayesian techniques for variable selection are

old and were developed using the idea of using the posterior probabilities of alternative models [1]. Subsequently, more advanced methods, such as stochastic variable selection techniques (e.g., Gibbs sampling), have been suggested [2]. Recently, new methods related to spike-and-slab techniques and variation Bayesian approaches for high-dimensional regression/classification have emerged, offering computationally effective solutions for the recognition of wide models employed in the analysis [7, 8]. These methods are also meaningful in the breast cancer classification context, as they help achieve parsimony and provide insights into the uncertainty of the inclusion of specific features.

Predictive performance is important, but accuracy alone is inadequate for many healthcare applications. Clinical users must understand the reasons for the prediction. This need for understanding has triggered the acceptance and development of explainable AI (XAI) with a focus on the interpretation of complex and statistical ML results. A recent systematic scoping review indicated that the use of XAI in BC diagnosis and risk prediction has dramatically increased [4]. Among the popular XAI techniques is SHapley Additive exPlanations (SHAP), which provides additive contribution of different feature layers to be analyzed at the population and patient levels. This feature of SHAP makes it possible to identify morphological characteristics that are most important in terms of clinical diagnosis and serve as a complement to traditional methods of interpretation.

The application of AI in medicine has accentuated the significance of uncertainty. A device can yield an extremely precise forecast even if the available information does not support it or if the input varies from that used for training. The latest healthcare advances using AI technology conclude that quantifying uncertainty has become increasingly and fervently emphasized in the context of credible prediction systems [5]. Probability calibration is vital because it considers the degree to which the predicted probabilities of a device relate to the occurrence of the result. Studies in this area have also shown that the efficacy of prediction and probability calibration aspects are closely related, although they remain a separate area of research [6]. Therefore, the information provided in this part of the research emphasizes that sound estimates of uncertainty can help differentiate forecasts that are more trustworthy from those that are less feasible.

Another principle worth stressing is the interpretability and stability in the process of developing models for clinical use. Even if certain features are deemed important in one fitted model, they tend to lose their role when the training data change slightly. Therefore, resampling and bootstrap methods can be employed to assess the reproducibility of the predictors under consideration and the parameters of the fitted models. Consequently, the problem is also strongly correlated to the well-known issue of biased error estimation during the model selection and validation processes [28]. Thus, model evaluation should not only include predictive performance but also the stability of the model features used, reliability of probability estimates, and explanation consistency [12–16]. The effectiveness of ML approaches for BC diagnosis and classification has been previously highlighted in existing studies [22–24]. However, most methods focus solely on predictive accuracy, while lacking the integration of sparse feature selection, explainability, stability analysis, and uncertainty quantification into a single interpretable framework. This raises the urgent need for models that remain transparent about their nature and implement feature selection and trustworthy AI. Bayesian variable selection has a long history in statistical modelling. The spike-and-slab family proposed using a sparse prior that would put probability mass close to 0 for the coefficients that are not important while allowing a wider distribution for the relevant variables. The key contributions made in terms of the advancement of Bayesian variable selection and stochastic search variable selection have laid the statistical groundwork for sparse model selection and posterior inclusion probability.

More recent research extends these ideas to sparse logistic regression and variational approximations, showing that Bayesian sparsity remains effective in modern classification tasks. In breast cancer diagnosis, many studies report strong accuracy for machine learning classifiers, relatively combine interpretability, feature stability, and uncertainty quantification in the statistical framework. A large part of the literature still focuses on point estimates and static feature importance, which can vary substantially across resampled data splits. Recent breast cancer explainability work confirms that XAI is increasingly used, but it also shows that most of the studies emphasize explanation generation rather than explanation stability or confidence-aware decision support. SHAP provides a principled explanation mechanism by itself, it does not solve feature instability or predictive uncertainty. Logistic regression is interpretable, may not adequately quantify parameter uncertainty unless Bayesian inference or resampling is used. Therefore, an integrated framework is needed in which feature selection, explanation, stability, and uncertainty are treated as mathematically linked components rather than independent post-processing steps.

This study proposes a hybrid Bayesian and explainable framework for BC classification using the WDBC dataset. In this framework we combine Bayesian sparse feature selection with logistic regression to identify a compact and informative set of morphological predictors. SHAP-based analysis is employed to visualise and quantify the contribution of individual features to model predictions. Bootstrap-based stability analysis is incorporated to evaluate the reproducibility of selected features and model coefficients across resampled datasets. Finally, uncertainty is assessed using Bayesian posterior predictive inference and probability-calibration analysis. The overall objective is to develop a classification framework with predictive, interpretable, stable, and uncertainty-aware.

This study proposes a hybrid framework that addresses these requirements simultaneously. The main contributions are (i) a Bayesian sparse feature-selection layer is placed on top of logistic regression to identify a compact subset of informative WDBC features. (ii) SHAP explainability is used to give both global and patient-specific attributions. (iii) bootstrap-based stability analysis is incorporated to assess how reproducible the selected features and coefficients are across resampled training sets. (iv) Uncertainty quantification is integrated through Bayesian posterior predictive inference and calibration-based confidence assessment.

2. Proposed Hybrid Explainable Logistic Regression Framework

Let the dataset be

$$[D = (x_i, y_i)_{i=1}^n,] \quad \dots(1)$$

where $(x_i \in R^p)$ is the feature vector and $(y_i \in \{0,1\})$ denotes benign ($\{0\}$) or malignant ($\{1\}$) class. For the WDBC setting, $(p=30)$. The goal is to learn a parsimonious probabilistic classifier with interpretable coefficients, stable feature selection, and explicit uncertainty quantification.

2.1 Logistic Regression Likelihood

The conditional probability of malignancy is modelled by

$$[P(y_i = 1 | x_i, \beta) = \sigma(\eta_i), \quad \eta_i = \beta_0 + x_i^T \beta,] \quad \dots(2)$$

where $(\sigma(z) = \frac{1}{1+e^{-z}})$ is the sigmoid function. The likelihood for the full dataset is

$$\prod_{i=1}^n \sigma(\eta_i)^{y_i} (1 - \sigma(\eta_i))^{1-y_i} \quad \dots(3)$$

Equivalently, the negative log-likelihood is

$$-\sum_{i=1}^n y_i \log \sigma(\eta_i) + (1 - y_i) \log (1 - \sigma(\eta_i)) \quad \dots(4)$$

2.2 Bayesian Feature Selection

To select informative predictors, introduce binary inclusion indicators

$$[\gamma_j \in \{0,1\}, \quad j = 1, \dots, p] \quad \dots(5)$$

The coefficient prior is specified as a spike-and-slab mixture:

$$[\beta_j | \gamma_j \sim (1 - \gamma_j) N(0, \tau_0^2) + \gamma_j N(0, \tau_1^2), \quad \tau_0^2 \ll \tau_1^2] \quad \dots(6)$$

The indicator prior is Bernoulli:

$$[\gamma_j \sim \text{Bernoulli}(\pi_j).] \quad \dots(7)$$

A simple noninformative choice is $(\pi_j = \pi)$, while a hierarchical prior can be used when prior clinical knowledge is available. The posterior is then

$$[p(\beta, \gamma | D) \propto p(y | X, \beta) p(\beta | \gamma) p(\gamma).] \quad \dots(8)$$

The posterior inclusion probability (PIP) of feature (j) is

$$[PIP_j = P(\gamma_j = 1 | D).] \quad \dots(9)$$

Features with large (PIP_j) are retained in the final sparse model. A practical rule is

$$[\gamma_j^* = I(PIP_j > \kappa),] \quad \dots(10)$$

where $(\kappa \in [0.5, 0.95])$ is selection threshold.

2.3 MAP Formulation

For computational efficiency, the Bayesian model can be approximated by a maximum a posteriori objective:

$$\arg \max \{ \log p(y | X, \beta) + \log p(\beta | \gamma) + \log p(\gamma) \} \quad \dots(11)$$

This is equivalent to minimizing

$$\sum_{j=1}^p \gamma_j \log \pi + \sum_{j=1}^p (1 - \gamma_j) \log (1 - \pi), \quad \dots(12)$$

with $(\lambda_0 = 1/(2\tau_0^2))$ and $(\lambda_1 = 1/(2\tau_1^2))$. This formulation preserves the interpretability of logistic regression while embedding a sparse Bayesian prior.

3. SHAP Explainability

SHAP interprets a prediction by breaking it down into its additive contributions. In a given instance (x), its output is represented as follows

$$[f(x) = \phi_0 + \sum_{j=1}^p \phi_j,] \quad \dots(13)$$

where (ϕ_0) is the baseline output and (ϕ_j) is the SHAP value of feature (j). In SHAP, the values attributed to various features are derived using the concepts of cooperative game theory.

For logistics regression, the particular form of log-odds is very suitable

$$\beta_0 + x^T \beta \quad \dots(14)$$

This indicates that there is a contribution of natural features at the level of log-odds. The formula is as follows:

$$[\phi_j \approx \beta_j (x_j - \underline{x}_j),] \quad \dots(15)$$

where (x_j) is the empirical mean of feature (j). Global feature importance is then obtained by averaging absolute SHAP values:

$$[I_j = \frac{1}{n} \sum_{i=1}^n |\phi_{ij}| \cdot] \quad \dots(16)$$

In the case of WDBC, the feature importance of the tumour concavity, radius, area, perimeter, and concave points is expected, as these features are related to the geometric irregularities linked to malignancy. This conforms to the view expressed in the medical XAI literature that the methods of explanation should provide the attribution of features that are meaningful from the clinical viewpoint rather than only providing the number.

4. Stability Analysis

A model must be able to deliver a consistent feature set along with feature coefficient parameters across numerous resampled datasets as a measure of its clinical usefulness. Bootstrap stability analysis was used for this purpose.

Let $(D^{(b)})$ represent the (b)-th bootstrap sample, $(b = 1, \dots, B)$. For each sample, estimate $(\beta^{(b)})$ and $(\gamma^{(b)})$. The selection frequency of feature (j) is

$$[S_j = \frac{1}{B} \sum_{b=1}^B I(\gamma_j^{(b)} = 1) \cdot] \quad \dots(17)$$

A stable characteristic must have (S_j) close to 1. The coefficient stability index can be defined as a formula

$$[1 - \frac{!(\beta_j^{(1)}, \dots, \beta_j^{(B)})}{|[\beta_j^*]| + \epsilon}] \quad \dots(18)$$

In this formula, $(\epsilon > 0)$ is a term that avoids the division by zero issue. A high value of (CSI_j) is a sign of the low sensitivity of the coefficient to the changes in the training data. The overall score of stability is defined as the following formula

$$[\frac{1}{p} \sum_{j=1}^p S_j] \quad \dots(19)$$

This metric calculates the variability of the sparse representation in case of a sampling change. The significance of stability in the process of making medical decisions is clear; without some level of consistency from sample to sample, physicians cannot trust the outcome.

5. Uncertainty Quantification

In breast cancer diagnosis, point predictions alone are insufficient, as a model might be correct on average but oversimplistic for borderline patients. Recent reports on healthcare uncertainty state that calibrated uncertainty increases robustness, makes deployment safer, and helps physicians detect cases of uncertainty for further evaluation by experts.

5.1 Posterior Predictive Probability

Under the Bayesian model, the posterior predictive probability for a new observation (x_*) is

$$\int \sigma(\beta_0 + x_*^T \beta), p(\beta, \gamma | D), d\beta, d\gamma \quad \dots(20)$$

This integral can be approximated through posterior samples, the Laplace approximation, or variational inference. The predictive uncertainty is then summarized by the Bernoulli variance:

$$p_*(1 - p_*), \quad p_* = P(y_* = 1 | x_*, D).] \quad \dots(21)$$

Maximum uncertainty is obtained at $(p_* = 0.5)$, corresponding to borderline cases.

5.2 Credible Intervals

For each coefficient (β_j) , a $(100(1 - \alpha))$ credible interval is obtained from the posterior:

$$[\beta_j \in [q_{\alpha/2}, q_{1-\alpha/2}],] \quad \dots(22)$$

where $(q_{\alpha/2})$ and $(q_{1-\alpha/2})$ are posterior quantiles. When there is no overlap with zero in the credible interval, the relevant feature has high relevance to the outcome, which serves as an alternative, more interpretable, statistically based method to apply instead of running a simple p-screen.

5.3 Calibration-Based Confidence

We will evaluate the predictive probabilities based on the observed frequencies using calibration, reliability charts, and the Brier score.

$$\frac{1}{n} \sum_{i=1}^n (p_i - y_i)^2] \quad \dots(23)$$

Lower Brier scores indicate better probabilistic calibration. This is especially useful because clinical decision support depends on well-calibrated risk estimates, not merely ranking performance. General calibration methods are widely used in probabilistic classification, and probability calibration is recognized as an important step when model outputs are to be interpreted as confidence estimates.

6. Algorithm

Hybrid Explainable Bayesian Logistic Regression for Breast Cancer Diagnosis

Input: Training data (D) , bootstrap size (B), inclusion threshold (κ) , confidence level (α)

Output: Sparse Bayesian classifier, SHAP explanations, stability scores, uncertainty estimates

1. Standardize all features in(D).
2. Initialize Bayesian logistic regression with spike-and-slab priors.
3. Estimate the posterior ($p(\beta, \gamma \mid D)$) using MAP, Laplace, or variational approximation.
4. Compute posterior inclusion probabilities (PIP_j).
5. Select features satisfying ($PIP_j > \kappa$).
6. Refit the sparse logistic regression model on the selected subset.
7. Compute SHAP values for global and local explanations.
8. Generate bootstrap resamples($D^{(1)}, \dots, D^{(B)}$).
9. Estimate feature selection frequencies(S_j), coefficient stability indices, and overall model stability.
10. Compute posterior predictive probabilities and uncertainty scores for new cases.
11. Report calibrated probabilities, confidence intervals, and explanation plots.

7. Statistical Evaluation

The proposed framework should be evaluated using stratified cross-validation and bootstrap validation. The following metrics are recommended:

$$\left[\begin{aligned} &Accuracy = \frac{TN + TP}{TN + TP + FN + FP} \\ &\left[Sensitivity = \frac{TP}{FN + TP}, Specificity = \frac{TN}{FP + TN} \right] \\ &\left[Precision = \frac{TP}{TP + FP}, F1 = \frac{2 \cdot Sensitivity \cdot Precision}{Sensitivity + Precision} \right] \\ &AUC = \int_0^1 TPR(u), du, \end{aligned} \right]$$

Along with Brier score and calibration error. Because it focuses on diagnosis, sensitivity and calibration are of utmost importance. To establish statistical significance and compare the proposed framework with baseline models, techniques such as the Wilcoxon signed-rank test and paired t-test, which depend on whether the problem conforms to the normal distribution, can be applied. Feature stability is typically reported in terms of selection frequency distributions and coefficient sign consistency. The quality of the explanation can be assessed through SHAP stability measurements and interpretability evaluations by clinicians.

8. Discussion

This model is good because it takes advantage of logistic regression, which is interpretable. Bayesian feature selection allows redundancy to be removed and provides resistance to related variables in the context of breast cancer. The next stage was to apply SHAP to change the final sparse model into an interpretable medical diagnostic tool by identifying the features that lead to a possible diagnosis of cancer. Stability analysis is an important element of the framework because a model can be trusted when the features are stable during resampling. Uncertainty quantification can also indicate the reliability of the prediction. This combination of methods is ideal for the WDBC dataset because its features are useful and can be easily calculated.

9. Conclusion

The current study introduces a hybrid model of explanatory logistic regression for identifying pathological breast tumours that integrates Bayesian variable selection and SHAP interpretation. The arithmetic footprint of the proposed model is represented in the mathematical world of sparse Bayesian inference and classification, while the model itself is interpretable to help medical professionals in their work. By allowing the user to obtain the posterior probabilities of inclusion, the model helps identify the most significant features for tumor characterization. The SHAP coefficients make it possible to understand both the general and individual predictions. Bootstrap stability analysis is used to assess the stability of the model results. Finally, posterior predictive uncertainty provides calibrated confidence intervals for predicted probabilities. Thus, this methodology can be applied in a medical context, where proper predictive performance, interpretability, and reliability are required.

References

1. Mitchell T J and Beauchamp J J. "Bayesian variable selection in linear regression", J Am Stat Assoc. (1988), Vol. 83, pp. 1023–1036.
2. George E I and McCulloch R E. "Variable selection via Gibbs sampling", J Am Stat Assoc. (1993), Vol. 88, pp. 881–889.
3. UCI Machine Learning Repository. "Breast Cancer Wisconsin (Diagnostic) Data Set", UCI Machine Learning Repository.

4. Ghasemi A, Hashtarkhani S, Schwartz D L and Shaban-Nejad A. "Explainable artificial intelligence in breast cancer detection and risk prediction: A systematic scoping review", *Cancer Innov.* (2024), Vol. 3, Article e136.
5. Seoni S et al. "Application of uncertainty quantification to artificial intelligence in healthcare", (2023).
6. Ojeda F M et al. "Calibrating machine learning approaches for probability estimation: A comprehensive comparison", (2023).
7. Ray K et al. "Spike and slab variational Bayes for high dimensional logistic regression", (2020).
8. Zhang J et al. "Variational Bayesian variable selection in logistic regression based on spike-and-slab Lasso", (2025).
9. Dileep G and Gyani S G G. "Artificial intelligence in breast cancer screening and diagnosis", *Cureus.* (2022), Vol. 14.
10. Esteva A, Robicquet A, Ramsundar B, Kuleshov V, DePristo M, Chou K, Cui C, Corrado G, Thrun S and Dean J. "A guide to deep learning in healthcare", *Nat Med.* (2019), Vol. 25, pp. 24–29. doi: 10.1038/s41591-018-0316-z.
11. Topol E J. "High-performance medicine: The convergence of human and artificial intelligence", *Nat Med.* (2019), Vol. 25, pp. 44–56. doi: 10.1038/s41591-018-0300-7.
12. Ravi S and Saranya A. "Breast cancer detection using machine learning in medical imaging—A survey", *Proc Comput Sci.* (2024), Vol. 239, pp. 2235–2242.
13. Kapila R and Saleti S. "An efficient ensemble-based machine learning for breast cancer detection", *Biomed Signal Process Control.* (2023), Vol. 86, Article 105269. doi: 10.1016/j.bspc.2023.105269.
14. Das A K, Biswas S K, Mandal A, Bhattacharya A and Sanyal S. "Machine learning based intelligent system for breast cancer prediction (MLISBCP)", *Expert Syst Appl.* (2024), Vol. 242, Article 122673. doi: 10.1016/j.eswa.2023.122673.
15. Yadav R K, Singh P and Kashtriya P. "Diagnosis of breast cancer using machine learning techniques—A survey", *Proc Comput Sci.* (2023), Vol. 218, pp. 1434–1443. doi: 10.1016/j.procs.2023.01.122.
16. Oztekin P S et al. "Comparison of explainable artificial intelligence model and radiologist review performances to detect breast cancer in 752 patients", *J Ultrasound Med.* (2024), Vol. 43, pp. 2051–2068. doi: 10.1002/jum.16535.
17. Sharma S, Aggarwal A and Choudhury T. "Breast cancer detection using machine learning algorithms", *Proc Int Conf Computational Techniques Electronics and Mechanical Systems (CTEMS), IEEE.* (2018), pp. 114–118.
18. Silva-Aravena F, Núñez Delafuente H, Gutiérrez-Bahamondes J H and Morales J. "A hybrid algorithm of ML and XAI to prevent breast cancer: A strategy to support decision making", *Cancers.* (2023), Vol. 15, Article 2443. doi: 10.3390/cancers15092443.
19. Munshi R M et al. "A novel approach for breast cancer detection using optimized ensemble learning framework and XAI", *Image Vis Comput.* (2024), Vol. 142, Article 104910. doi: 10.1016/j.imavis.2024.104910.
20. Arravalli T, Chadaga K, Muralikrishna H, Sampathila N, Cenitta D, Chadaga R and Swathi K S. "Detection of breast cancer using machine learning and explainable artificial intelligence", *Sci Rep.* (2025), Vol. 15, Article 26931.
21. Ali A, Alghamdi M, Marzuki S S, Din T A T, Yamin M S, Alrashidi M, Alkhazi I S and Ahmed N. "Exploring AI approaches for breast cancer detection and diagnosis: A review article", *Breast Cancer Targets Ther.* (2025), Vol. 17, pp. 927–947.
22. Wolberg W H, Street W and Mangasarian O. "Machine learning techniques to diagnose breast cancer from image-processed nuclear features of fine needle aspirates", *Cancer Lett.* (1994), Vol. 77, pp. 163–171.
23. Akay M F. "Support vector machines combined with feature selection for breast cancer diagnosis", *Expert Syst Appl.* (2009), Vol. 36, pp. 3240–3247.
24. Cakmak Y and Pacal I. "Enhancing breast cancer diagnosis: A comparative evaluation of machine learning algorithms using the Wisconsin dataset", *J Oper Intell.* (2025), Vol. 3, pp. 175–196.
25. Mienye I D, Obaido G, Jere N, Mienye E, Aruleba K, Emmanuel I D and Ogbuokiri B. "A survey of explainable artificial intelligence in healthcare: Concepts, applications, and challenges", *Inform Med Unlocked.* (2024), Vol. 51, Article 101587.
26. Doshi-Velez F and Kim B. "Towards a rigorous science of interpretable machine learning", *arXiv.* (2017), Article arXiv:1702.08608.
27. Ghasemi A, Hashtarkhani S, Schwartz D L and Shaban-Nejad A. "Explainable artificial intelligence in breast cancer detection and risk prediction: A systematic scoping review", *Cancer Innov.* (2024), Vol. 3, Article e136.
28. Varma S and Simon R. "Bias in error estimation when using cross-validation for model selection", *BMC Bioinform.* (2006), Vol. 7, Article 91.