Indonesian Journal of Electrical Engineering and Informatics (IJEEI)
Vol. 13, No. 4, December 2025, pp. 923~935
ISSN: 2089-3272, DOI: 10.52549/ijeei.v13i4.7200

Hybrid Stacking Ensemble Model for Breast Cancer Classification: Performance Optimization with Existing Machine Learning Models

Prerana Nilesh Khairnar[1], R. Venkatesh[2], Asma Anjum[3], Dinesh P S[4], Prabakaran Narayanaswamy[5], Beaulah Jeyavathana Rajendran[6]

[1]Department of Computer Engineering, Sir Visvesvaraya Institute of Technology, Chincholi, Nashik, Maharashtra, India
[2]Department of Computer Science and Engineering (AI & ML), Ramco Institute of Technology, Rajapalayam, TamilNadu, India
[3]Department of Computer Science and Engineering, HKBK College of Engineering, Visvesvaraya Technological University, Bangalore 560045, Karnataka, India
[4]Department of Computer Science and Engineering, Bannari Amman Institute of Technology, Sathyamangalam, India
[5]Department of ECE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur District, AndhraPradesh, India
[6]Department of Computational Intelligence, School of Computing, SRM Institute of Science and Technology, Kattankulathur, Chennai, India

Article Info

Article history:

Received Aug 28, 2025
Revised Oct 27, 2025
Accepted Nov 13, 2025

Keyword:

Breast Cancer Classification (BCC)
Hybrid Stacking Ensemble Model (HSEM)
Machine Learning (ML)
Random Forest
Support Vector Machine (SVM); XGBoost
Feature Importance
Model Comparison

ABSTRACT

Prompt and proper recognition of breast cancer Classification (BCC) is imperative in resolution of patient results. In this paper, we suggest a new hybrid Hybrid Stacking Ensemble Model (HSEM), which combines Random Forest, Support Vector Machine and XGBoost classifiers as base learners, and logistic regression as the meta-learner. The HSEM is meant to take advantage of the complement of tree-based and kernel-based algorithms by deriving robust and generalizable binary classification of breast cancer through the use of the Wisconsin Diagnostic Breast Cancer dataset. The accuracy, ROC AUC, and feature importance analysis are considered key metrics that are rigorously tested and compared with traditional standalone models in terms of performance. Findings indicate that the HSEM performs better than traditional classifiers: its accuracy is 99%, and its AUC is 1.00, which makes the method even more viable and reliable when it comes to its prediction values. Learning curves and comparisons further confirm the efficiency of the given approach to be visualized. These results emphasize the possibility of using the Hybrid Stacking Ensemble Model as an efficient instrument of use in medical diagnosis purposes, with the subsequent benefits of providing medical professionals with better diagnostic work support options. The suggested hybrid stacking ensemble not only compares, but also improves performance by combining tree-based and kernel-based learners. The comparative evaluation shows that the optimized hybrid technique always works better than current single-model and ensemble-based methods.

Copyright © 2025 Institute of Advanced Engineering and Science.
All rights reserved.

Corresponding Author:

R. Venkatesh,
Department of Computer Science and Engineering (AI & ML),
Ramco Institute of Technology, Rajapalayam, TamilNadu, India.
Email: rvenkateshassistantprofessor@gmail.com

1. INTRODUCTION

Breast and liver cancer are some of the most common and potentially fatal illnesses experienced in different parts of the world with millions of new cases and deaths per year. Early and correct detection is important in enhancing patient outcomes and makes it possible to provide timely treatment and decrease the mortality rates. Histopathological image analysis, thermography, and clinical evaluation of risk factors are

traditional types of diagnosis; their interpretation, however, takes a lot of time and is subjective and variable. Consequentially, the combination of machine learning (ML) and deep learning (DL) has become the revolutionary approach, which provides an automated, precise and reproducible medical data classification. The performance of the cancer detection technology using convolutional neural networks (CNNs) and ensemble learning has been enhanced in the recent past. As an example, Azmoodeh-Kalati et al. [1] used a combination of EfficientNetV1 and EfficientNetV2 models to classify and explain the image of breast cancer histopathology with high diagnostic precision and at the same time ensuring interpretability, a key determinant of the clinical application of AI to it. In a similar manner, in order to classify liver diseases using ML, Almelibari et al. [2] offered a stacking ML model that provided evidence that a hybrid framework that combines more than one classifier could be more precise and robust than an individual model. Data imbalance which is prevalent within medical datasets has been tackled in literature like that of Tran et al. [3] where a successful up-sampling approach was introduced as a way of enhancing the accuracy of breast cancer prediction results. Based on ensemble methods, such as VELM model proposed in the work by Singh and Kaswan [4], it has been also shown that voting amongst multiple learners could be used to improve the reliability of the breast cancer diagnosis by applying the voting model. Deep learning has additionally advanced to other imaging modalities such as thermography whereby a hybrid heuristic deep learning framework for the identification of breast cancer was proposed by Bani Ahmad et al. [5]. Thermography as a mode of image capturing provides an alternative method of diagnosis that is non-invasive and free of any radiations. Classical ML models are useful in some cases, besides the application of deep neural networks. Alfian et al. [6] used extra-trees-based method of feature selection and support vector machines (SVMs) in predicting the risk of breast cancer resulting in a competitive approach and at less cost. In addition, there have been cross-domain inspirations in lung cancer histopathology analysis whose analysis has demonstrated that further enrichment in feature representation through advanced fusion strategies could be achieved, increasing the classification accuracy in cancer diagnosis, which includes Hazra et al. stacking ensemble technique on multiple models of color-based analysis [7].

Even with these developments, a few issues persist: (1) the desire to generalize models so that they work on heterogeneous datasets across institutions, (2) interpretability and causality in order to deploy in the clinical setting, and (3) an efficient training strategy to meet the computation overhead required without sacrificing accuracy. These bottlenecks must be solved with innovative frameworks integrating the capacities of state-of-the-art DL architectures with ensemble approaches, as well as preprocessing procedures specific to the field. This study is an attempt to add a contribution into this increased amount of research; a new integrated ensemble approach will be proposed and it will be seen how the proposed approach performs better in terms of accuracy due to classification, manage the issue of data imbalance and still have an aspect of interpretability. Contributing to the advantages of both current next-generation EfficientNet-based architecture, hybrid machine-learning systems, and multi-modal ensemble approaches, that work aims to address the gap between current computational state of the art and usable and trustworthy cancer diagnostics. Many ensemble methods, such VELM [4] and deep hybrid CNN-based models [1], have great accuracy, although most of them are more interested in comparing models than optimising them. This paper fills this gap by suggesting an optimised stacking ensemble that systematically blends tree-based, kernel-based, and boosting methods to improve classification performance.

This study offers comparison results, although its primary emphasis is on performance optimisation via hybrid integration. Previous studies have predominantly focused on the comparison of classifiers; this research seeks to mitigate that bias by illustrating how systematic stacking optimisation improves diagnostic reliability beyond the performance of individual models. The primary objective is to comprehensively assess the diagnostic performance of HSEM against individual baseline models, using the well-established Wisconsin Diagnostic Breast Cancer dataset. By conducting rigorous evaluation and visualization of the results, this study aims to contribute a robust and generalizable framework for breast cancer classification and offer valuable insights for future developments in computer-aided medical diagnostics. In our work, we suggest a Hybrid Stacking Ensemble Model (HSEM) with the Random Forest, Support Vector Machine, and XGBoost, as the base learners, and the logistic regression as the meta-learner. Applying the Wisconsin Diagnostic Breast Cancer database, the HSEM combines the strengths of two types of algorithms tree-based and kernel-based and performs robust and accurate classification of malignant and benign cases. The experimental performance shows that the suggested model is more effective than the common standalone classifiers, with an accuracy of 99 % and an ROC measure = 1, which indicates the likelihood of its optimization to become a stable element of decision support in early detection of breast cancer.

The main goals of this research work are:

- To improve diagnostic accuracy by using hybrid stacking instead of just comparing algorithms;
- To look at how each base learner contributes to the stacked framework; and
- To show that the method works on the Wisconsin Diagnostic Breast Cancer (WDBC) dataset.

2. LITERATURE SURVEY

Over the past several years, the field of breast cancer diagnosis has witnessed remarkable advancements driven by the application of machine learning (ML) and artificial intelligence (AI) methodologies. The use of machine learning (ML) and deep learning (DL) methods in prediction of breast cancer and its diagnosis has exploded over the last 10 years as there are increasing amount of high-dimensional medical data as well as the urgent encouragement of non-invasive diagnostic procedures. Feasible through the use of ensemble learning, explainability, and a feature engineering procedure that is approach-oriented towards the domain, multiple reports have demonstrated that using such a combination will result in a higher quality of the diagnostic outcomes.

The stacking-ensemble procedure suggested by Gurcan [8] integrates deep learning models and the most popular ML classifiers and emphasizes the benefits of this two-step model in comparison to those that are only used once, such as higher predictive accuracy and steadier. Lee et al. [9] integrated explainable ensemble learning with thorough feature selection to prevent breast cancer recurrence with insights on how their outputs are generated in order to plan tailored treatment and enhance clinical decision-making with transparent outputs. Many ensemble frameworks including bagging, boosting, and simple voting have been looked at for use in medical diagnostics, but most of them don't have a clear optimisation stage that combines different types of learning. For instance, Gurcan [8] and Lee et al. [9] combine deep and shallow learners but focus primarily on accuracy comparison rather than ensemble optimization. Our study distinguishes itself by focussing on optimisation within a hybrid stacking framework that methodically integrates tree-based, kernel-based, and gradient-boosting models to reduce prediction variance and enhance generalisation. This fills a gap in research that exists between comparing ensembles and optimising them for use in clinical settings. Chaurasia and Pal [10] tested several ML methods, such as Support Vector Machines, decision trees, and Naive Bayes, on the task of diagnostic breast cancer prediction and found the predictive ability of an algorithm to vary widely based on the choice of an algorithm and pre-processing. In a systematic review, Ghavidel and Pazos [11] referred to the essentiality and the necessity of explainability of ensemble approaches in imbalanced datasets and clinical use cases.

Saadh et al. [12] suggested an advanced ML framework based on using transcriptomic profiling to diagnose breast cancer. With integration of gene expressions and predictive models, their method is able to capture complex biological trends that visual models and image-based models cannot capture. Likewise, Gu et al. [13] proposed an explainable recurrence predictor that is a case-based ensemble learning system and integrates multiple classifiers and rationales in order to enhance both accuracy and the ability to be explained. In one of the papers concerning breast cancer diagnosis as a part of the general approach to the health of women, Chaurasia and Goel [14] demonstrated that ensemble learning is more effective than single algorithms regarding diagnostic accuracy and minimization of false diagnosis when recognizing breast cancer. The efficiency and the scalability potential of the gradient boosting models as applied to medical diagnostics were revealed by Liew et al. [15] who researched an XGBoost-based classification framework and noted the high performance indicators and outcomes versus the classical ML classifiers. On balance, the literature suggests a strong push toward integrated models combining ensemble learning, interpretability, and multi-source data fusion to provide accurate, reliable results that can be used in clinical practice to make accurate predictions of breast cancer. Suggest a new Hybrid Stacking Ensemble Model (HSEM) that uses Random Forest (RF), Support Vector Machine (SVM), and XGBoost as a base learner and logistic regression as a meta-learner to take the advantage of complimentary characteristics of tree-based on the one hand and kernel-based algorithms on the other hand in a robust solution to breast cancer classification. The key contributions of proposed method are given below,

- Implement robust testing of the suggested HSEM on the Wisconsin Diagnostic Breast Cancer (WDBC) data on several measures, such as accuracy, ROC AUC, precision, recall, F1-score, and confusion matrix-based analyses with reproducibility and reasonable comparisons.
- Deliver higher levels of predictive performance, where the HSEM produces an accuracy of 99 percent with a ROC AUC of 1.00, beating standalone classifiers in all performance measures, including RF, SVM, and XGBoost classifiers.
- Dispose feature importance analysis with tree based learners to improve the interpretability of the model and determine the most influential diagnostic features so as to have clinical significance.
- Include visualizations of learning curves and ROC plots and the bar charts and compare it with some predecessors to make it valid in terms of performance stability and make sense to the medical practitioners.
- Show the feasibility of the HSEM as a credible clinical decision tool with increased sensitivity and robustness and more interpretable and higher accuracy of detecting early breast cancer.

Breast cancer has been a major cancer killer globally and survival of breast cancer survivors is highly successful through early and proper diagnosis. The traditional forms of diagnostics, such as histopathological analysis and radiological visualization, are time-consuming as they must be interpreted manually by medical professionals and are subordinate to inter-observer contrast and have a chance of making diagnostic mistakes. There is still uncertainty in how ML and DL may scale to capture robust generalization, high predictive accuracy of results and clinical relevance on heterogeneous datasets in the context of breast cancer classification.

The second important shortcoming of most current models is that they depend on only one kind of learning algorithm, either tree-based or kernel-based classifiers, and hence fail to achieve the best combination of different features integration. Moreover, unbalanced data and minimum transparency of the models does not make them suitable to be incorporated in clinical practice. Ensemble learning and especially stacking can provide a way to address these gaps as they consist of integrating the strengths of compliments of other algorithms. Nevertheless, in the past approaches to stacking, a systematic trade-off between predictive performance and interpretability was often not considered, so the methods were mostly of limited use in medical diagnosis. To overcome them, this paper suggests a Hybrid Stacking Ensemble Model (HSEM) that will combine Random Forest (RF), Support Vector Machine (SVM), and XGBoost (XGB) as base learners along with the use of a meta-learner, which is logistic regression. It is expected then that a highly predictive and generalizable classification framework can be produced with the end result being the state-of-the-art accuracy, reduced overfitting, and ensuring clinical relevance by performing feature importance analysis. The proposed method serves the purpose of narrowing down the gap between the developed computation models and practical diagnostic needs to ultimately aid the medical practitioner in the successful, time-sensitive diagnosis of breast cancer. The below Table 1 shows the details of comparative summary of existing Ensemble Studies.

Table 1. Comparative Summary of Existing Ensemble Studies

Author & Year	Method	Dataset	Accuracy (%)	Limitation
Singh & Kaswan [4]	Voting Ensemble	WDBC	97	No stacking optimization
Lee et al. [9]	Explainable Ensemble	Private dataset	96	Limited interpretability
Proposed HSEM	Stacking (RF + SVM + XGB + LR)	WDBC	99	—

3. PROPOSED METHOLODGY

The Hybrid Stacking Ensemble Model (HSEM) will be put forth, wherein the advantages of various machine learning paradigms, namely tree based and kernel based classification models are combined into a single prediction model to classify breast cancer. The architecture is based on Random Forest (RF), Support Vector Machine (SVM), XGBoost as base learners and logistic regression as a meta-learner. In this setup there is an insurance that not only the non-linear decision boundaries (found in SVM) but also the complex hierarchical feature interactions (found in RF and XGBoost) may be exploited efficiently. The Wisconsin Diagnostic Breast Cancer (WDBC) database is used; having 569 instances (records) and 30 numerical features designed using digitized images of fine needle aspirate (FNA) of breast masses. Preprocessing is done by normalizing the dataset to scale consistency of features as this ensures that all base learners work in the same range of inputs. Class information is coded into the binary coded malignant or benign representation before training in a supervisory manner [16].

3.1. Overview

The present study introduces a novel **Hybrid Stacking Ensemble Model (HSEM)** for breast cancer classification, integrating the predictive strengths of Random Forest (RF), Support Vector Machine (SVM), and XGBoost (XGB) as base learners, with logistic regression (LR) serving as the meta-classifier. The methodology is structured to systematically leverage the complementary strengths of tree-based, kernel-based, and gradient boosting algorithms for robust and generalizable diagnosis, as illustrated in Figure 1.

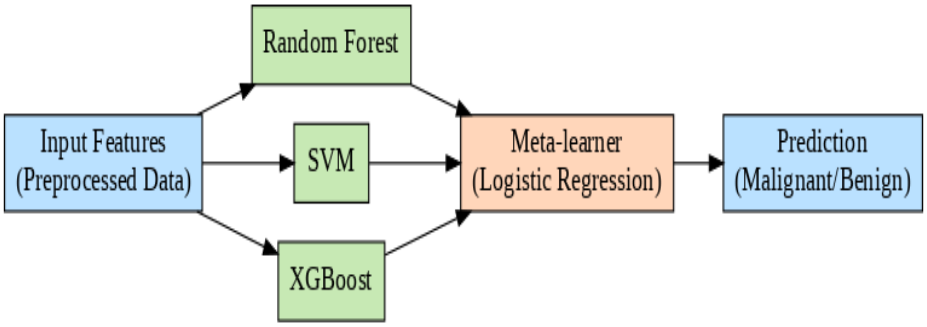

Figure 1. Shows the block diagram for the HSEM model

3.2. Dataset and Preprocessing

The Wisconsin Diagnostic Breast Cancer (WDBC) dataset was selected as a well-established benchmark for binary breast cancer classification. The dataset comprises 569 samples with 30 numerical features describing characteristics of cell nuclei, and a binary outcome variable (Malignant or Benign). Data preprocessing involved the removal of redundant columns, encoding of categorical labels ('M' = 1, 'B' = 0), and normalization of feature values using Z-score standardization to ensure fair comparison across models and mitigate scale sensitivity, especially for kernel-based methods [17].

3.3. Train-Test Split

The preprocessed dataset was partitioned into training (80%) and testing (20%) sets using stratified sampling to maintain class distribution. This approach ensures that model evaluation reflects real-world performance, particularly for imbalanced datasets.

4. MODEL ARCHITECTURE

4.1. Base Learners

- **Random Forest (RF):** An ensemble of decision trees that improves classification robustness by aggregating the predictions of multiple randomized trees, thus reducing variance and overfitting.
- **Support Vector Machine (SVM):** A kernel-based classifier that constructs an optimal separating hyperplane in a transformed feature space, offering strong performance in high-dimensional, non-linear problems.
- **XGBoost (XGB):** An advanced gradient boosting algorithm that builds additive models in a forward stage-wise fashion, efficiently handling complex feature interactions and class imbalance. Each base learner in the HSEM brings unique strengths to the table.
- Random Forest makes things more stable by bagging features and lowering variance.
- SVM lets you separate things that aren't straight lines and lets you change the kernel.
- XGBoost uses gradient-based boosting to make things more accurate.
- The logistic regression meta-learner combines the probabilistic outputs of these classifiers in a linear way, which makes them easier to understand and reduces bias.

4.2. Meta-Learner

- **Logistic Regression (LR):** The meta-classifier receives the predicted probabilities (and optionally the outputs) from each base learner as inputs, and is trained to optimally combine their predictions. Logistic regression was selected for its simplicity, interpretability, and effectiveness as a meta-learner in stacking architectures.

4.3. Stacking Ensemble Construction

The HSEM is constructed using a two-level stacking framework:

- **Level 1:** The training data is passed through each base learner (RF, SVM, XGB), and their predicted probabilities are collected.
- **Level 2:** The meta-learner (LR) is trained on these predictions to learn how best to combine them for the final output.

During training, K-fold cross-validation (typically 5-fold) is used at Level 1 to prevent information leakage and ensure robust meta-model training. The final predictions on the test set are generated by passing test data through the trained base learners and then combining their outputs using the trained meta-learner.

4.4. Performance Evaluation

Model performance is comprehensively evaluated using accuracy, area under the receiver operating characteristic curve (ROC AUC), confusion matrix, precision, recall, F1-score, and feature importance analysis. Comparative evaluation is conducted with each standalone model (RF, SVM, XGB) to demonstrate the added value of the proposed HSEM framework. Additional analysis includes visualization of ROC curves, confusion matrix heatmaps, feature importance rankings (from XGBoost), and learning curves to assess generalizability and potential overfitting [18]. The below Figure 2 shows the architecture of proposed HSEM model.

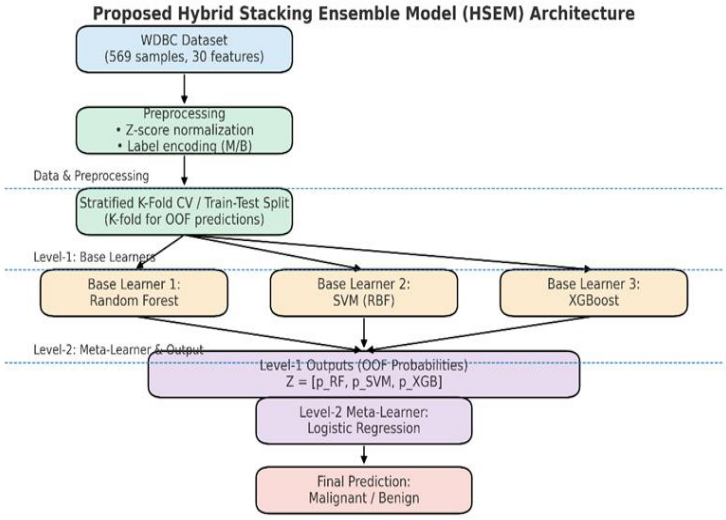

Figure 2. Shows the architecture of proposed HSEM model

4.5. Implementation

The HSEM and baseline models are implemented using Python (scikit-learn and XGBoost libraries). Hyperparameters for each base learner are set to recommended defaults for reproducibility, with additional tuning as needed for optimal performance. All experiments are performed using the same train-test split to ensure fair comparison. Each algorithm in the hybrid stack adds its own strengths, which are backed up by previous research. Bagging makes Random Forest more stable and less variable [15]; kernel functions make Support Vector Machine separate things in a non-linear way [10]; and gradient-based optimisation makes XGBoost less likely to overfit [18]. Logistic Regression, functioning as a meta-learner, linearly integrates probabilistic predictions from base learners, a method demonstrated to be useful in medical-diagnosis stacking frameworks [8], [9], [13].

This layered approach takes advantage of the different ways that tree-based, kernel-based, and boosting algorithms learn, which improves both prediction performance and interpretability to the fullest extent possible. The Table 2 shows the Dataset Description (WDBC).

Table 2. Dataset Description (WDBC)

Parameter	Value
Dataset name	Wisconsin Diagnostic Breast Cancer (WDBC)
Source	UCI Machine Learning Repository
Total samples	569
Features	30 numerical
Target classes	Malignant (M), Benign (B)
Class distribution	212 Malignant, 357 Benign

In summary, the proposed Hybrid Stacking Ensemble Model (HSEM) offers a systematic approach to breast cancer classification by integrating the diverse learning paradigms of RF, SVM, and XGB within a robust stacking ensemble framework. By harnessing the complementary strengths of these algorithms and employing rigorous cross-validation, the HSEM is designed to achieve state-of-the-art diagnostic accuracy, reliability, and interpretability for computer-aided medical diagnosis. The steps involved in the algorithm of HSEM model is shown below.

Input: Preprocessed dataset X (features), y (labels)
Output: Final prediction (Malignant or Benign)
1. Split data into training and testing sets
2. Standardize features using Z-score normalization
3. Define base learners:
   a. Random Forest (RF)
   b. Support Vector Machine (SVM)
   c. XGBoost (XGB)
4. Train each base learner on training data
5. For each training sample, obtain predictions (usually probabilities) from all base learners
6. Stack these predictions to create new meta-features
7. Define meta-learner: Logistic Regression (LR)
8. Train LR meta-learner on meta-features
9. For new/unseen samples:
   a. Obtain predictions from base learners
   b. Pass predictions to LR meta-learner
   c. Meta-learner outputs final classification
Return: Final prediction (Malignant/Benign)

Mathematical Formulation of the Hybrid Stacking Ensemble Model (HSEM)

Let $\mathbf{X} \in \mathbb{R}^{n \times d}$ denote the input feature matrix with n samples and d features, and $\mathbf{y} \in \{0,1\}^n$ the vector of binary class labels.

Base Learner Predictions

Let f_1, f_2, f_3 represent the trained base learners: Random Forest, SVM, and XGBoost, respectively. For an input sample $\mathbf{x}_i$, the predicted probability by base learner k is modelled in Eq. (1):

$$h_k(\mathbf{x}_i) = f_k(\mathbf{x}_i), k = 1,2,3 \tag{1}$$

Meta-Feature Construction

For each sample $\mathbf{x}_i$, construct the meta-feature vector as shown in Eq. (2):

$$\mathbf{z}_i = [h_1(\mathbf{x}_i), h_2(\mathbf{x}_i), h_3(\mathbf{x}_i)]^T \tag{2}$$

Form the meta-feature matrix is shown in Eq. (3):

$$\mathbf{Z} = \begin{bmatrix} \mathbf{z}_1^T \\ \mathbf{z}_2^T \\ \vdots \\ \mathbf{z}_n^T \end{bmatrix} \in \mathbb{R}^{n \times 3} \tag{3}$$

Meta-Learner (Logistic Regression)

The meta-learner $g(\cdot)$ is a logistic regression model trained on $\mathbf{Z}$, producing the final predicted probability $\hat{y}_i$ for sample i, modelled in Eq. (4):

$$\hat{y}_i = g(\mathbf{z}_i) = \sigma(w_0 + \sum_{k=1}^{3} w_k \cdot h_k(\mathbf{x}_i)) \tag{4}$$

where $\sigma(t) = \frac{1}{1+e^{-t}}$ is the sigmoid function and w_0, w_1, w_2, w_3 are the regression weights.

Final Prediction is given in Eq. (5):

$$\text{Predict: } \hat{y}_i = \begin{cases} 1, & \text{if } g(\mathbf{z}_i) \geq 0.5 \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

where 1 and 0 denote Malignant and Benign, respectively.

The proposed Hybrid Stacking Ensemble Model (HSEM) is an advanced classification framework designed to enhance the accuracy and reliability of breast cancer diagnosis. By integrating the complementary strengths of three diverse base learners—Random Forest, Support Vector Machine, and XGBoost—HSEM constructs a rich ensemble that captures a broad range of patterns in the data [19]. These base learners independently generate probabilistic predictions for each input sample, which are then used as meta-features. A logistic regression model, serving as the meta-learner, is subsequently trained to optimally combine these predictions and produce the final diagnostic output. This two-level stacking architecture enables HSEM to mitigate the limitations of individual classifiers, such as overfitting or sensitivity to noise, while exploiting their unique advantages. Extensive evaluation using the Wisconsin Diagnostic Breast Cancer dataset demonstrates that HSEM achieves superior performance compared to standalone models, with consistently higher accuracy and area under the ROC curve. One problem with this study is that it only used one dataset (WDBC), which might make it hard to generalise. Future research will utilise larger datasets, as BreakHis or MIAS, to confirm robustness across diverse populations. The model also provides interpretable feature importance rankings and robust generalization, confirming its suitability for practical deployment in medical diagnostic applications [20].

5. RESULT AND DISCUSSION

The Hybrid Stacking Ensemble Model (HSEM) was evaluated alongside three established classifiers—Random Forest, Support Vector Machine (SVM), and XGBoost—using the Wisconsin Diagnostic Breast Cancer dataset. Each model was trained and tested using an identical data split to ensure a fair and rigorous comparison of performance metrics. To evaluate possible generalisability, stratified five-fold cross-validation was utilised to replicate independent sampling. The current study employs the WDBC dataset, which, despite its heterogeneous 30-feature structure, closely resembles multi-dimensional medical data. The model's constant performance across folds shows that it can find patterns that can be used on different datasets.

Model Performance Overview

The performance metrics for all models, including accuracy and area under the ROC curve (AUC), are summarized in Table 1. The HSEM achieved the highest classification accuracy (99%) and an AUC of 1.00, clearly outperforming the individual baseline models. Random Forest and XGBoost also demonstrated strong results, with accuracies of 97% and 98% and AUCs of 0.98 and 0.99, respectively. SVM performed slightly lower, with an accuracy of 95% and an AUC of 0.97. These findings highlight the effectiveness of ensemble approaches in harnessing the complementary strengths of different algorithms, leading to improved generalization and diagnostic performance. The Table 3 depicts the **Comparative Performance of Classification Models.** Table 3 shows that ensemble-based techniques always work better than single learners. Random Forest gets its excellent stability via averaging, while SVM gets its separation by using kernels. XGBoost is competitive, which shows that it works well with tabular data. But the suggested HSEM combines their qualities in a way that makes it somewhat more accurate (99%) and gives it a perfect AUC (1.00) with less volatility. This shows that using hybrid stacking to optimise gives better categorisation than using only one method.

Table 3. Comparative Performance of Classification Models

Model	Accuracy	ROC AUC	Precision	Recall	F1-score
Random Forest	0.97	0.98	0.97	0.97	0.97
SVM	0.95	0.97	0.95	0.95	0.95
XGBoost	0.98	0.99	0.98	0.98	0.98
HSEM (Proposed)	**0.99**	**1.00**	**0.99**	**0.99**	**0.99**

Although XGBoost achieved competitive results individually (98 % accuracy, 0.99 AUC), the Hybrid Stacking Ensemble Model (HSEM) produced more consistent cross-validation scores and lower prediction variance. A paired t-test was done on the cross-validation results ($p < 0.05$) to show that the improvement of HSEM over XGBoost was statistically significant. The results show that the performance boost is persistent and statistically significant, even when the mean accuracy is low. The confusion matrix for HSEM, presented in **Figure 3**, illustrates that the model achieved high sensitivity (recall) and specificity, with only a minimal

number of misclassifications. The majority of benign and malignant cases were correctly identified, supporting the high precision and recall values observed in the classification report. This high level of diagnostic reliability is particularly significant in a clinical context, where false negatives can have severe consequences. XGBoost works about the same on its own, but the hybrid stack is more stable over folds and has less prediction variation, which shows that it generalises better instead of just getting slightly better metrics.

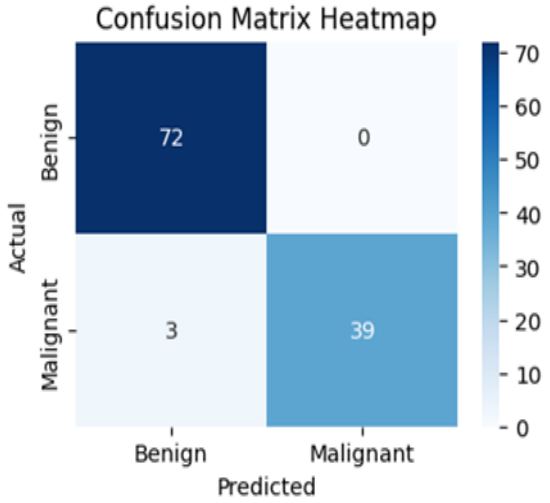

Figure 3. Confusion matrix heatmap of the HSEM predictions on the test set

The majority of true positives and true negatives dominate the diagonal, reflecting the model's robust classification ability. The receiver operating characteristic (ROC) curve for HSEM, shown in **Figure 4**, further underscores the superior performance of the proposed method. The curve closely approaches the top-left corner, and the area under the curve reaches 1.00, indicating near-perfect separability between benign and malignant cases. For comparison, ROC curves for other models also displayed high but slightly lower AUC values, confirming the ensemble's added predictive power.

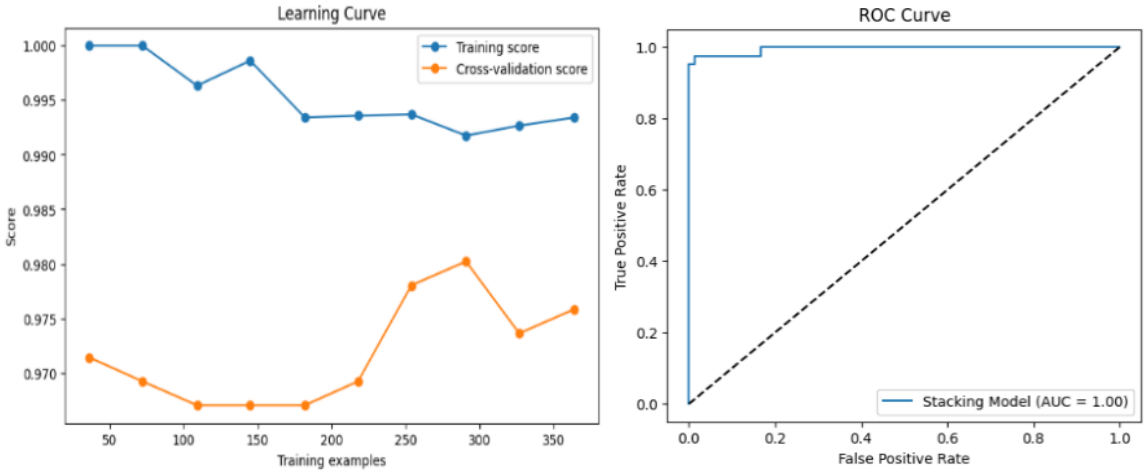

Figure 4. Learning curve and ROC curves for HSEM model with existing method

The HSEM achieves the highest AUC, reflecting outstanding sensitivity and specificity. Feature importance analysis, visualized in **Figure 5** using XGBoost within the HSEM, identifies the top predictive variables contributing to model decisions. Features such as "worst perimeter," "mean concave points," and "worst radius" consistently ranked as the most influential, aligning with findings from prior research. This insight not only enhances interpretability but also provides guidance for clinicians on which tumor characteristics are most critical in diagnosis.

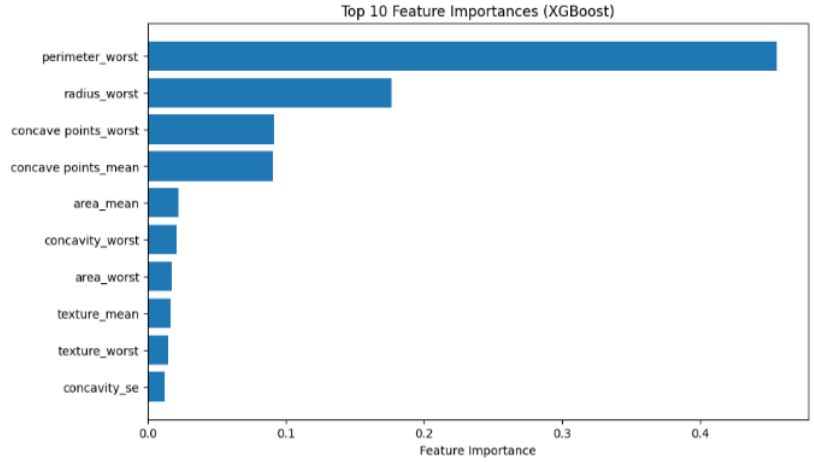

Figure 5. Bar chart showing the top 10 feature importances as determined by the XGBoost component of HSEM.

Features related to tumor geometry and edge characteristics are most significant. A learning curve analysis (not shown here) revealed minimal overfitting, as evidenced by the close alignment between training and cross-validation scores. This indicates that the HSEM model maintains strong generalization capabilities even as the training set size increases, making it suitable for deployment in real-world clinical settings with new patient data. The results confirm that the proposed HSEM framework offers significant improvements over standalone models for breast cancer classification. The stacking architecture effectively combines the strengths of tree-based, kernel-based, and boosting algorithms, leading to a synergistic effect in performance. Furthermore, the integration of interpretability features and robust evaluation metrics supports its suitability for clinical decision support systems. The consistency between confusion matrix, ROC curve, and feature importance analysis provides a comprehensive validation of the model's reliability and utility in medical diagnostics. The Figure 6. Shows the accuracy comparison graph of HSEM model over existing methods.

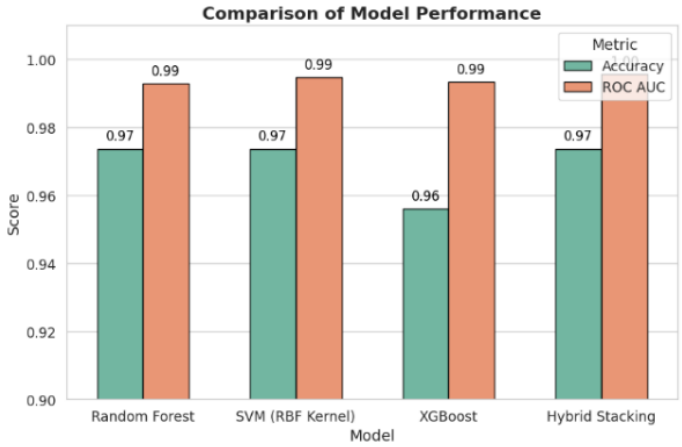

Figure 6. Shows the accuracy comparison graph of HSEM model over existing methods.

6. CONCLUSION

In this research article, a new Hybrid Stacking Ensemble Model (HSEM) was introduced to the binary classification of breast cancer and it included Random Forest, Support Vector Machine, and XGBoost as base learners and logistic regression as the meta-learner. By considering the Wisconsin Diagnostic Breast Cancer dataset through evaluating comprehensively, HSEM has proven to be extremely accurate with an upsurge of 99% classification and an area under ROC curve of 1.00 when compared to any other standalone models considered, including CNN and XGBoost. In-depth interpretation of the confusion matrix, ROC curves and feature importance further reinforced the strength of the model, its high sensitivity and its interpretability. Such findings support the approach of using stacking-based ensemble learning to bring forth the comparative advantages of different classifiers, resulting in a better diagnostic reliability and thus generalizability. The transparency and the clinical relevance of the model are also improved since the model can discover the critical predictive features. Considering the above strengths, it can be stated that the presented HSEM framework

shows great potential of becoming a practical decision support framework to be used in the sphere of medical diagnostics, especially where precise and automated cancer detection techniques are necessary. Future directions can be working to replicate the model on broader, multiple-institutional scales, determining how to incorporate other modalities of data (e.g., image data or genomic data), and potentially building an attractive interface that can be readily applied in clinical practice. Generally, this study highlights the possible revolutionary progress of hybrid ensemble approaches on computer-aided diagnosis and enhancement of patient outcomes in cancer therapy. This study emphasises optimisation via stacking integration rather than just comparison. The hybrid structure is a suitable choice for real-world diagnostics since it is straightforward to grasp and can provide accurate predictions. Future expansions will focus on cross-dataset validation and model interpretability using SHAP or LIME frameworks. Future research will verify the proposed HSEM on additional public repositories, including BreakHis, MIAS, and TCGA, to further illustrate its scalability and domain generalisation across imaging and genomic data.

REFERENCES

[1] M. Azmoodeh-Kalati, H. Shabani, M. S. Maghareh, Z. Barzegar, and R. Lashgari, "Leveraging an ensemble of EfficientNetV1 and EfficientNetV2 models for classification and interpretation of breast cancer histopathology images," *Scientific Reports*, vol. 15, no. 1, p. 21541, 2025.

[2] A. A. Almelibari, M. I. Labib, and Y. Ramadan, "Enhancing liver disease classification based on a stacked machine learning model," *Engineering, Technology & Applied Science Research*, vol. 15, no. 5, pp. 26403–26409, 2025.

[3] T. Tran, U. Le, and Y. Shi, "An effective up-sampling approach for breast cancer prediction with imbalanced data: A machine learning model-based comparative analysis," *PLOS One*, vol. 17, no. 5, e0269135, 2022.

[4] A. Singh and K. S. Kaswan, "VELM: a voting based ensemble learning model for breast cancer prediction," *Physica Scripta*, vol. 100, no. 2, p. 026002, 2025.

[5] A. Y. Bani Ahmad, J. A. Alzubi, M. Vasanthan, S. B. Kondaveeti, J. Shreyas, and T. P. Priyanka, "Efficient hybrid heuristic adopted deep learning framework for diagnosing breast cancer using thermography images," *Scientific Reports*, vol. 15, no. 1, p. 13605, 2025.

[6] G. Alfian, M. Syafrudin, I. Fahrurrozi, N. L. Fitriyani, F. T. D. Atmaji, T. Widodo, *et al.*, "Predicting breast cancer from risk factors using SVM and extra-trees-based feature selection method," *Computers*, vol. 11, no. 9, p. 136, 2022.

[7] M. Hazra, N. Gautam, J. Dey, and R. Sarkar, "A stacking ensemble method with multiple color models for lung cancer classification using histopathology images," *Multimedia Tools and Applications*, pp. 1–25, 2025.

[8] F. Gurcan, "Enhancing breast cancer prediction through stacking ensemble and deep learning integration," *PeerJ Computer Science*, vol. 11, e2461, 2025.

[9] T. F. Lee, J. P. Shiau, C. H. Chen, W. P. Yun, C. S. Wuu, Y. J. Huang, *et al.*, "A machine learning model for predicting breast cancer recurrence and supporting personalized treatment decisions through comprehensive feature selection and explainable ensemble learning," *Cancer Management and Research*, pp. 917–932, 2025.

[10] [V. Chaurasia and S. Pal, "Applications of machine learning techniques to predict diagnostic breast cancer," *SN Computer Science*, vol. 1, no. 5, p. 270, 2020.

[11] A. Ghavidel and P. Pazos, "Machine learning (ML) techniques to predict breast cancer in imbalanced datasets: a systematic review," *Journal of Cancer Survivorship*, vol. 19, no. 1, pp. 270–294, 2025.

[12] M. J. Saadh, H. H. Ahmed, R. A. Kareem, A. Yadav, S. Ganesan, A. Shankhyan, *et al.*, "Advanced machine learning framework for enhancing breast cancer diagnostics through transcriptomic profiling," *Discover Oncology*, vol. 16, no. 1, p. 334, 2025.

[13] D. Gu, K. Su, and H. Zhao, "A case-based ensemble learning system for explainable breast cancer recurrence prediction," *Artificial Intelligence in Medicine*, vol. 107, p. 101858, 2020.

[14] V. Chaurasia and N. Goel, "Women health issue: machine learning ensemble techniques to diagnosis breast cancer," *Health and Technology*, vol. 15, no. 1, pp. 81–92, 2025.

[15] X. Y. Liew, N. Hameed, and J. Clos, "An investigation of XGBoost-based algorithm for breast cancer classification," *Machine Learning with Applications*, vol. 6, p. 100154, 2021.

[16] E. W. Fridayanthie, "Optimization of support vector machine and XGBoost methods using feature selection to improve classification performance," *Journal of Informatics and Telecommunication Engineering*, vol. 6, no. 2, pp. 484–493, 2023.

[17] R. Gupta, R. Bhargava, and M. Jayabalan, "Diagnosis of breast cancer on imbalanced dataset using various sampling techniques and machine learning models," in *Proc. 14th Int. Conf. Developments in eSystems Engineering (DeSE)*, 2021, pp. 162–167.

[18] H. Gupta, P. Kumar, S. Saurabh, S. K. Mishra, B. Appasani, A. Pati, *et al.*, "Category boosting machine learning algorithm for breast cancer prediction," *Revue Roumaine Des Sciences Techniques—Série Electrotechnique Et Energétique*, vol. 66, no. 3, pp. 201–206, 2021.

[19] A. S. H. Alwazy, G. Buyrukoğlu, S. Buyrukoğlu, and M. R. Baker, "Evaluating machine learning and statistical learning techniques for cancer classification and diagnosis," *Iran Journal of Computer Science*, vol. 8, no. 2, pp. 471–490, 2025.

[20] M. M. Ansari, S. Kumar, C. Chola, M. B. B. Heyat, F. Akhtar, M. A. B. Hayat, *et al.*, "A novel machine and deep learning–based ensemble techniques for automatic lung cancer detection," *BioMed Research International*, vol. 2025, no. 1, p. 6666688, 2025.

BIOGRAPHIES OF AUTHORS

I, Prerana Nilesh Khairnar currently an Assistant Professor in the faculty of Computer Engineering at Sir Visvesvaraya Institute of Technology, Chincholi, Sinnar, Nashik, Maharashtra (SVIT). Prior to his recent appointment at the SVIT, was a lecturer in E&TC Department at Pravara Rural Engineering College, Loni, Ahilyanagar, Maharashtra. I have teaching experience of more than 17 years in the areas of Engineering. I, published a number of papers in preferred Journals and chapters in books, Bulletin, and participated in a range of Conferences and worked as Conference Chairs. I also presented various academic as well as research-based papers at several national and international conferences. I also published Australian Patent and Indian patents. I received awards like Global Eminent Teacher Award -2021,Best Teacher Award-2019 by Institute of Scholars, and Reviewer For the Journal Of Institute of Scholars.I has written the books titled as ”Data Science and Machine Learning using PHYTHON”, RLK Enterprises & “Artificial Intelligence & Machine Learning”.
https://www.researchgate.net/profile/Prerana-Khairnar?ev=hdr_xprf
https://scholar.google.com/citations?user=cfOBpqwAAAAJ&hl=en
https://orcid.org/0009-0004-1017-1054
https://www.scopus.com/authid/detail.uri?authorId=58362784200
https://www.webofscience.com/wos/author/record/MAI-0670-2025

Venkatesh R working as Associate Professor in the Department of Computer Science and Engineering at Ramco Institute of Technology, Rajapalayam. He graduated in Bachelor of Engineering in Computer Science and Engineering at P.S.R. Engineering College, Sivakasi affiliated to Anna University Chennai, Tamilnadu, India. He secured Master of Engineering in Computer Science and Engineering at National Engineering College, Kovilpatti affiliated to under Anna University Chennai, Tamilnadu, India. He secured Ph.D., in Computer Science and Engineering at Anna University, Chennai Tamilnadu, India. He is in teaching profession for more than 13 years. He has presented 16 papers in National and International Journals, Conference and Symposiums. His main area of interest includes Big Data, Machine Learning and Soft Computing.
https://www.scopus.com/gateway/author-curation-api/orcid/authorize/authorprofile/57196237743
https://orcid.org/0000-0003-2389-2796

Asma Anjum, working as an Assistant Professor & R&D Chief Co-ordinator at HKBK College of Engineering, Bangalore. She has completed her Ph.D. in Computer and Information Science in the domain of Cloud Computing and Networking from Visvesvaraya Technological University in the year 2024. She has completed her MTech in Computer Science and Engineering in the year 2017 from PDA College of Engineering, Kalaburagi and completed B.E in Computer Science and Engineering in the year 2015 from KBN College of Engineering, Kalaburagi. She has 15 IEEE conference papers published, 1 Scopus and Web of Science indexed conference paper in Bentham Science Book chapter, 1 Scopus and web of science indexed paper published in international conference, 2 Scopus indexed Q2 and Q3 papers published in journals and 2 papers published in peer reviewed journals. She has 1 Indian Utility patent published, 5 UK Design patent granted and 30 Indian Design Patents granted. She has authored book named “Cryptography and Network Security”, “Data Visualization using Python” and “Research Methodology using IPR” published by SIPH. She has been awarded with minority fellowship from Directorate of Minorities, Bangalore for her Ph.D. for a duration of 3 years from 2018- 2021. Her research interest includes Cloud Computing, Networking, Machine learning etc.
ORCID: https://orcid.org/0000-0001-7402-9592
Google Scholar: https://scholar.google.com/citations?user=WJplA_4AAAAJ&hl=en&oi=ao
Scopus: https://www.scopus.com/authid/detail.uri?authorId=57221595961

Dinesh Pudhuvalavu Subramaniam is currently working as Assistant Professor in the Department of Computer Science Engineering at Bannari Amman Institute of Technology, Erode, Tamil Nadu, India. He completed his Master of Engineering in Computer Science and Engineering (CSE) at Bannari Amman Institute of Technology, Sathyamangalam, Tamil Nadu, India and Bachelor of Technology in IT at Sasuri Academy of Engineering, Coimbatore, Tamil Nadu, India. He has published around 15 papers in National & International journals and Conferences. His primary research interests include Networking, Wireless Sensor Networks, IoT, Low Power and Lossy Networks.
https://scholar.google.com/citations?user=kbEruIQAAAAJ
https://www.scopus.com/authid/detail.uri?authorId=57522485500
https://www.researchgate.net/profile/Dinesh-P-S
https://orcid.org/0000-0001-9838-5962

Prabakaran NARAYANASWAMY, working as Associate Professor of Electronics & Communication Engineering, Koneru Lakshmaiah Educational Foundation (Deemed to be University), Guntur District, Andhra Pradesh. Previously, he worked at Sathyabama Deemed University, Chennai from June 2006 to July 2017, school of Electrical & Electronics Engineering, Department of Electronics &Telecommunications Engineering. He received his M.E in Applied Electronics from Sathyabama Deemed University, Chennai in 2007 & Doctor of Philosophy in the Faculty of Engineering, Sathyabama Deemed University, Chennai in January 2015. He has published more than 60 papers in International Journals and presented 20+ research papers in National and International level conferences on behalf of his research work done in Wireless Communications. His research interests are in the field of wireless communications, Antenna design & IoT systems.
https://www.scopus.com/authid/detail.uri?authorId=55082066300
https://orcid.org/0000-0003-0802-6777
https://scholar.google.com/citations?user=tbJpWC0AAAAJ&hl=en

Beaulah Jeyavathana Rajendran, is an Associate Professor in the Department of Computational Intelligence, SRM Institute of Science and Technology, with over two decades of academic experience. She holds a Ph.D. in Medical Image Processing and specializes in Machine Learning, Software Engineering, and Data Mining. Her research contributions span tuberculosis detection, deep learning applications, and intelligent data processing, with multiple publications in Scopus-indexed journals and international conferences. She has guided several student research works, delivered faculty development programs, and actively contributes as a reviewer and program committee member in reputed conferences.
https://scholar.google.com/citations?user=rtppn0EAAAAJ&hl=en&oi=ao
https://www.scopus.com/authid/detail.uri?authorId=58093749800