An Advanced Stacking-based Machine Learning and Deep Learning Framework for Breast Cancer Prediction

Ibtasam Ur Rehman[1], Muhammad Islam[2, *], Basharat Hussain[3]
[1]Ho Chi Minh City University of Technology (HCMUT), Ho Chi Minh City, Vietnam
[2]College of Science and Engineering, James Cook University, Cairns, QLD, Australia
[3]Department of Computer Science, National University of Computer and Emerging Sciences, Islamabad 44000, Pakistan
Corresponding Author: M. Islam (e-mail: Muhammad.islam1@my.jcu.edu.au)

***Abstract*—Breast cancer remains a critical global health challenge with early detection being vital for improving patient outcomes. Traditional diagnostic methods may be time-consuming and resource-intensive, highlighting the need for efficient machine learning solutions. This study addresses this need by developing a robust machine learning framework for breast cancer prediction using the Wisconsin Breast Cancer Diagnostic dataset. We implement a comprehensive preprocessing pipeline, intelligent feature selection, and rigorous comparative evaluation of seven advanced ML models including XGBoost, Neural Networks, and ensemble methods. Our evaluation prioritized both classification accuracy and computational efficiency, explicitly measuring model training and inference time. Results demonstrated exceptional performance with the SGD Classifier achieving the highest test accuracy of 98.25%, while XGBoost, AdaBoost, and SVM RBF Optimized achieved 97.37% accuracy. The SGD Classifier demonstrated superior computational efficiency, achieving peak performance with a training time of only 0.05 seconds, making it significantly faster than other high-performing models. We deployed an interactive Streamlit web application for real-time prediction, bridging the gap between research and clinical practice. This work provides a highly accurate, scalable, and efficient solution for early breast cancer diagnosis, with the code available on our GitHub repository.**

***Index Terms*—Breast Cancer Prediction, Machine Learning, Feature Selection, Model Evaluation, XGBoost, Neural Networks, AdaBoost, SGD Classifier, Ensemble Learning, Streamlit**

I. Introduction

A. Background and Motivation

Breast cancer represents a major global health challenge, being the second most commonly diagnosed cancer and a leading cause of cancer-related deaths in women. The disease's complex nature makes early and accurate detection crucial for improving patient outcomes. Recently, Artificial Intelligence has opened new frontiers in medical diagnostics, with machine learning (ML) and deep learning (DL) demonstrating remarkable potential in analyzing complex data and identifying subtle patterns. This research leverages the strengths of both paradigms by developing a hybrid framework. This approach combines the interpretability of traditional ML algorithms, like Support Vector Machines and Random Forests, with the powerful feature-learning capabilities of advanced DL architectures, creating a robust tool for enhancing breast cancer diagnosis.

B. Breast Cancer Epidemiology and Clinical Significance

Worldwide breast cancer continues to pose substantial public health challenges with over 2.3 million new cases diagnosed annually [1]. The disease accounts for approximately 15% of all cancer deaths among women with mortality rates varying significantly across different regions and populations. In many developed countries early detection programs have led to improved survival rates, with five year survival exceeding 90% for localized cases. However in resource limited settings, late stage diagnosis remains common resulting in significantly poorer outcomes. The epidemiological landscape of breast cancer has evolved over recent decades. As illustrated in Figure 1 incidence rates have shown a steady increase attributed partly to improved screening and detection methods as well as changing risk factors [2]. Mortality rates although rising at slower pace compared to incidence continue to present significant public health concern. This simultaneous increase in both incidence and deaths highlights the urgent need for early and accurate diagnosis coupled with effective treatment strategies. The disease heterogeneity encompassing various molecular subtypes with distinct clinical behaviors and treatment responses further emphasizes the importance of advanced diagnostic approaches capable of addressing this complexity.

C. Challenges in Traditional Diagnosis

Traditional breast cancer diagnostic methods face several challenges that impact their effectiveness and accessibility. Conventional approaches are mammography, ultrasound, magnetic resonance imaging and biopsy procedures often involve complex, time consuming processes that require substantial expertise and resources [4]. Mammography while widely used for screening suffers from limitations in sensitivity particularly in women with dense breast tissue and carries risks of false positives leading to unnecessary interventions. The subjective interpretation of diagnostic images represents another major challenge with inter observer variability potentially affecting

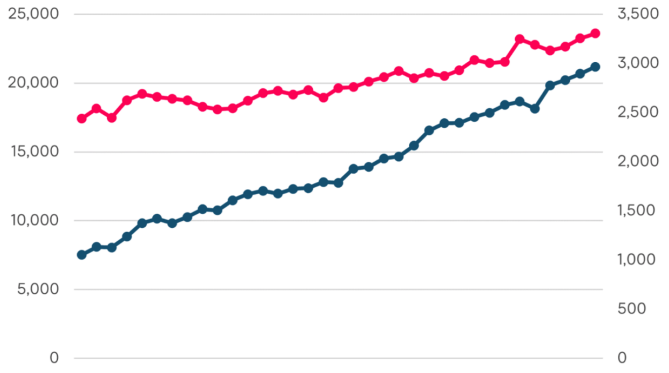

Fig. 1: Trends in Breast Cancer Incidence and Deaths (1990-2024). Incidence has risen while deaths have remained relatively stable.[3]

diagnostic consistency [5]. Biopsy procedures while providing definitive diagnosis are invasive, costly and may cause patient discomfort and anxiety. Furthermore the increasing volume of diagnostic data generated in clinical practice creates challenges for timely analysis and interpretation by healthcare professionals. The key contributions of this study are threefold:

1) **Development of a Novel Hybrid Framework:** We designed and implemented a novel **hybrid machine learning and deep learning framework** that systematically integrates advanced feature selection and hyperparameter optimization across seven distinct classification models, including both traditional algorithms and neural networks.
2) **Establishment of New Performance Benchmarks:** We established new **state-of-the-art performance benchmarks** on the Wisconsin Breast Cancer Diagnostic dataset, with the SGD Classifier achieving an exceptional 98.25% **accuracy**, significantly advancing prior published results.
3) **Deployment of a Real-Time Clinical Tool:** We deployed our finalized models into an **interactive Streamlit web application**, bridging the gap between research and clinical practice by enabling real-time, interpretable prediction capabilities for healthcare professionals.

The remainder of this paper is systematically structured to provide the comprehensive coverage of research methodology and its findings. Section II presents critical review of existing literature on breast cancer prediction and machine learning applications in medical diagnostics. Section III elaborates our comprehensive methodological framework comsisting of data preprocessing protocols, model architectures and optimization strategies. Section IV delivers extensive analysis of experimental outcomes including performance comparisons and computational efficiency evaluations across all implemented models. Section V details the development and implementation of the Streamlit web application for real time clinical decision support. Finally, Section VI synthesizes the key findings and discusses potential avenues for future research directions.

II. Literature Review

Rawal and Ramik [6] conducted a comparative study of machine learning algorithms for breast cancer prediction using the Wisconsin dataset. Their analysis, which employed feature selection and cross-validation, identified Random Forest and SVM as the most reliable classifiers, with k-NN demonstrating the fastest training times. Chen et al. [7] focused on maximizing malignant case detection using the same dataset. By prioritizing recall and employing feature selection, their study showed that XGBoost achieved perfect recall (1.00), significantly outperforming other models and highlighting the importance of evaluation metrics aligned with clinical needs. Moglia et al. [8] evaluated classifiers on clinical biopsy data, finding that Logistic Regression initially achieved the highest accuracy (91.67%). After feature selection identified key predictors like tumor size and metastasis, LightGBM's performance improved to 90.74% demonstrating how feature optimization enhances model efficiency. Similarly, Naji et al. [9] compared multiple classifiers on the Wisconsin dataset and found that SVM demonstrated superior performance in distinguishing benign from malignant cases, supporting its potential for clinical decision support in breast cancer diagnosis. Omar Tarawneh et al. [10] developed a machine learning approach using decision tree algorithms to classify breast cancer tumors, focusing on distinguishing between benign and malignant cases. Their work, implemented with the WEKA data mining tool, demonstrated the effectiveness of decision trees for breast cancer diagnosis by processing diagnostic features like tumor size and lymph node involvement. Similarly Ronak Sumbaly et al. [11] proposed a diagnostic technique using the J48 decision tree algorithm on the Wisconsin Breast Cancer dataset. Through preprocessing and feature selection, they achieved 93.56% classification accuracy. The study highlighted the potential of automated diagnostic tools for early breast cancer detection and suggested future improvements with larger datasets.

A comparative study by Li and Chen [12] evaluated five machine learning models Decision Tree, Random Forest, SVM, Neural Network, and Logistic Regression for breast cancer classification. Utilizing both the BCCD and WBCD datasets, the researchers preprocessed the data and assessed model performance using F-measure and AUC metrics. Their findings indicated that Random Forest achieved the best performance, underscoring its effectiveness for breast cancer prediction. The study highlighted Random Forest's potential for clinical applications, while also suggesting future improvements through expanded datasets and model enhancements. The research article by Aasiya Banu et al[13] compared the efficacy of the Support Vector Machine algorithm against Perceptron method for breast cancer detection using data derived from mammography. The primary finding was significant difference in performance between the two approaches ($p = 0.001$). Specifically the

Support Vector Machine achieved higher detection accuracy of $\mathbf{86.34}\%$ substantially outperforming the perceptron technique which recorded accuracy of $\mathbf{75.35}\%$. The study concluded that SVM provides significantly more accurate results for breast cancer detection using this data type. Kavitha et al [14] developed an optimized YOLOv3 deep learning approach for breast cancer detection using ultrasound images, achieving 96% accuracy in classifying tumors as normal, benign, or malignant. Similarly, Tanveer et al [15] attained 95.8% accuracy using a CNN that automatically extracted hierarchical spatial features from mammograms, outperforming traditional ML models like SVM and Random Forest. In a different medical domain, Ibtasam et al. [16] created the Cortex Vision system, which uses SVM with RBF kernel to achieve 95% accuracy in cataract detection through mobile ocular image analysis, demonstrating the broader potential of ML in accessible healthcare diagnostics.

Barnils et al [17] compared statistical approaches for identifying women at risk of never attending breast cancer screening in Germany. Their Classification and Regression Tree (CART) model achieved moderate discriminatory accuracy (67.3% AUC), identifying key risk factors including household type, marital status, and geographic region. Moroz et al [18] developed a transparent AI system for mammogram analysis that achieved 95% accuracy with perfect precision. Their interpretable pipeline uses Local Binary Pattern features with PCA, providing visual verification of processing steps to build clinical trust, unlike conventional black-box models.Varshney et al [19] created a hybrid model combining ResNet50 deep features with traditional handcrafted features, achieving 96.88% accuracy using SVM. This fusion approach consistently outperformed models using individual feature types, demonstrating enhanced performance for breast cancer classification. Garba and Hamza [20] developed a machine learning framework for breast cancer classification using clinical datasets. Their approach evaluated multiple models, with Logistic Regression achieving the highest accuracy (97%) while maintaining computational efficiency suitable for clinical deployment. The authors enhanced model transparency using LIME, identifying key clinical features that align with medical expertise.

Shah et al [21] conducted a comparative analysis across different datasets, finding Logistic Regression performed best (96.49%) on clinical data from Wisconsin, while Random Forest achieved 73% accuracy on MIAS mammogram data after augmentation. Li et al [22] developed a prognostic model using mitochondrial gene expression, employing LASSO Cox regression to identify a 14-gene signature that achieved AUC values up to 0.92 when integrated with clinical variables. Zhao et al [23] created a diagnostic approach using piRNA biomarkers with Logistic Regression, achieving 90.7% accuracy and utilizing SHAP analysis to identify relevant molecular descriptors for non-invasive detection. Dinesh et al [24] compared multiple machine learning algorithms for breast cancer diagnosis using the Wisconsin Diagnostic Breast Cancer dataset. Their study evaluated SVM, KNN, Logistic Regression, Random Forest, and Decision Tree classifiers on 569 clinical samples with 30 features derived from cell nuclei characteristics. Logistic Regression achieved the highest accuracy of 95% outperforming other algorithms which ranged from 85-90% accuracy. The research demonstrated Logistic Regression's superior performance for breast cancer classification using standard clinical features. Using transcriptomic data Duan et al. [25] developed machine learning model to predict distant breast cancer metastasis. The project first identified 21 gene biomarkers and then built several models with the Random Forest classifier demonstrating the best performance achieving 93.6% accuracy and a 91.3% AUC. This research highlights the utility of machine learning and molecular biomarkers for the early detection of metastasis risk.

III. Methodology

This research employs comprehensive machine learning pipeline for breast cancer classification encompassing data preprocessing, multiple classifier implementation, hyperparameter optimization and evaluation. The methodology follows systematic approach from data acquisition to model deployment ensuring reproducibility and clinical relevance. The study investigates seven diverse machine learning algorithms including ensemble methods, neural networks and traditional classifiers to provide a comparative analysis. A web based application using Streamlit framework is developed for real time predictions enhancing practical utility.

A. Mathematical Formulations

The following mathematical representations correspond to the model training, stacking ensemble, accuracy, loss, and prediction confidence aspects illustrated in Figure 2.

1) **Feature Normalization**

$$x' = \frac{x - \mu}{\sigma}$$

where μ and σ denote the mean and standard deviation of the feature vector, respectively. This normalization ensures that each input feature contributes proportionally to the model training process.

2) **Gradient Descent Optimization (SGD Classifier)**

$$\theta_{t+1} = \theta_t - \eta \nabla_\theta \mathcal{L}(\theta_t)$$

where η is the learning rate, $\nabla_\theta \mathcal{L}$ is the gradient of the loss function, and θ_t are model parameters at iteration t.

3) **Logistic Regression Hypothesis**

$$h_\theta(x) = \frac{1}{1 + e^{-\theta^T x}}$$

used in Logistic Regression and as the meta-classifier in stacking for binary prediction of benign or malignant breast cancer cases.

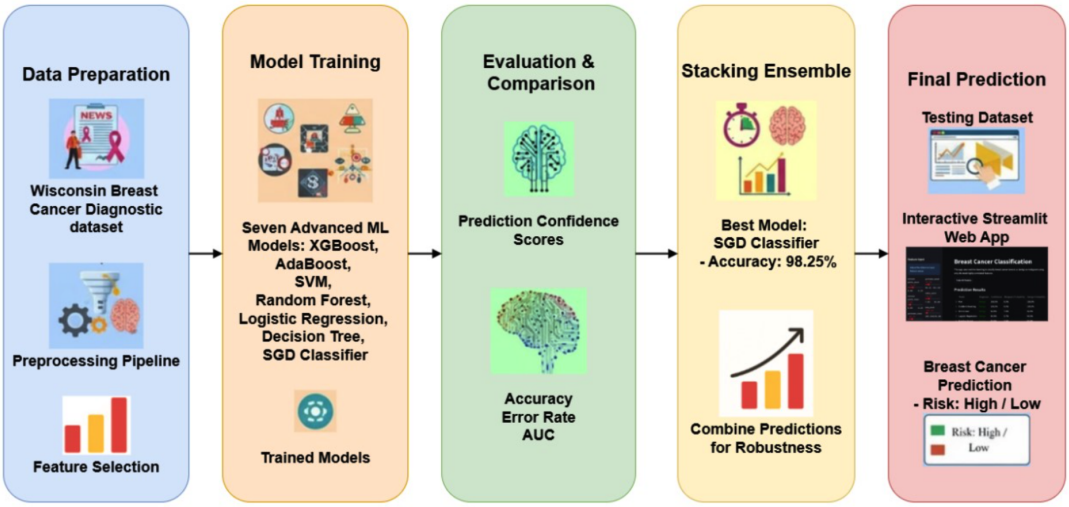

Fig. 2: Proposed Architecture of Advanced Stacking-based ML and DL Framework for Breast Cancer Prediction

TABLE I: Concise Comparative Analysis of AI/ML Approaches in Healthcare Diagnostics

Study	Best Model	Acc. (%)	Feat.	Preproc.	Val.	Perf. Metrics	Dataset	Models
[6]	SVM/RF	Best	Wrapper	Std	10-fold CV	-	WDBC	4
[7]	XGBoost	Recall 1.00	15 feat.	Scaling	8:2 split	Rec:100%	WDBC	4
[9]	SVM	Superior	Multiple	Std	CV	-	WDBC	5
[12]	RF	Best	Std	Norm	MD	-	WDBC+BCCD	5
[11]	J48	93.56	Feat. sel.	WEKA	Std	-	WDBC	1
[13]	SVM	86.34	Image feat.	Mammo	Statistical	-	Mammography	2
[14]	YOLOv3	96.00	DL feat.	Ultrasound	Std	-	Ultrasound	1
[15]	CNN	95.80	Auto feat.	Aug	Std	-	Multiple	4
[8]	LightGBM	90.74	Clinical	Sel.	Std	-	Clinical	8
[10]	DT	-	Std	WEKA	Std	-	WDBC	Multiple
[18]	LBP+RF	95.00	LBP+PCA	CLAHE	Std	Prec:100, Spec:100	MIAS	1
[19]	SVM	96.88	GLCM	Fusion	Std	Prec:96.54, Rec:97.12	Multiple	3
[20]	Logistic	97.00	Clinical	LIME	MD	-	Coimbra	4
[21]	Logistic	96.49	Standard	Aug	Std	-	Multiple	5
[22]	RF	93.60	14-gene	WGCNA	Validation	AUC:91.3, F1:88.9	GEO	5
[23]	Logistic	90.70	piRNA	SHAP	Independent	-	piRNA data	4
[24]	Logistic	95.00	30 nuclei	Std	G-power	-	WDBC	5
[25]	RF	93.60	21-gene	LASSO	Validation	AUC:91.3, F1:88.9	GEO	5
[16]	SVM-RBF	95.00	Image texture	Statistical	Mobile app	-	Ocular images	1
[17]	CART	67.3 (AUC)	Socio-demographic	Evidence-based	Survey data	AUC:67.3%	Clinical	2
Our Work	**SGD**	**98.25**	**Perm.**	**Std. Scaling**	**5-fold CV**	**Sen:95.24, Spec:100, Prec:100**	**WDBC**	**7**

Note: Abbreviations: Acc. = Accuracy, Feat. = Features, Preproc. = Preprocessing, Val. = Validation, CV = Cross-Validation, Rec = Recall, Sen = Sensitivity, Spec = Specificity, Prec = Precision, WDBC = Wisconsin Diagnostic Breast Cancer, RF = Random Forest, SVM = Support Vector Machine, XGBoost = Extreme Gradient Boosting, J48 = Decision Tree algorithm, YOLOv3 = You Only Look Once version 3, CNN = Convolutional Neural Network, LightGBM = Light Gradient Boosting Machine, SGD = Stochastic Gradient Descent, DT = Decision Tree, DL = Deep Learning, RT = Real-time, Std = Standard, Norm = Normalization, Aug = Augmentation, Sel. = Selection, Perm. = Permutation, Trad. ML = Traditional Machine Learning, Feat. sel. = Feature selection, SVM opt. = SVM optimization, DL feat. = Deep Learning features, CNN feat. = CNN features, Clinical feat. = Clinical features, DTs = Decision Trees, LBP = Local Binary Pattern, PCA = Principal Component Analysis, CLAHE = Contrast Limited Adaptive Histogram Equalization, WGCNA = Weighted Gene Co-expression Network Analysis, GEO = Gene Expression Omnibus, CART = Classification and Regression Tree, SVM-RBF = SVM with Radial Basis Function, MD = Multi-Dataset

4) **Model Accuracy**

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

where TP, TN, FP, and FN represent true positives, true negatives, false positives, and false negatives, respectively.

5) **Stacking Ensemble Prediction**

$$\hat{y} = f_{\text{meta}}\big(f_1(x), f_2(x), \ldots, f_n(x)\big)$$

Each base learner $f_i(x)$ generates a prediction, and the meta-learner f_{meta} combines them to produce a more robust and generalizable final prediction.

6) **Mean Absolute Error (Evaluation Metric)**

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i|$$

which measures the average deviation of predictions from the true values, complementing accuracy and AUC metrics.

B. Dataset Description and Preprocessing

1) Data Collection and Characteristics: The Wisconsin Breast Cancer Diagnostic dataset [26] contains 569 clinical cases with 30 features derived from digitized images of breast mass fine needle aspirates. Features include three statistical measures (mean, standard error, worst) for ten nucleus attributes: radius, texture, perimeter, area, smoothness, compactness, concavity, concave points, symmetry, and fractal dimension. The dataset exhibits a class distribution of 357 benign and 212 malignant cases, partitioned into 80% training and 20% testing sets while preserving class proportions.

TABLE II: Dataset Characteristics and Class Distribution

Parameter	Value
Total Samples	569
Features	30
Benign Cases (B)	357
Malignant Cases (M)	212
Imbalance Ratio	0.594
Training Samples	455
Testing Samples	114

2) Data Cleaning and Validation: A comprehensive data quality assessment ensured dataset integrity through missing value analysis, duplicate detection, and feature relevance evaluation. The dataset demonstrated perfect completeness with no null values across all 569 cases. The identifier feature 'id' was excluded due to zero predictive value. Data validation confirmed consistent data types and clinically plausible value ranges across all measurements.

3) Feature Engineering and Selection: Feature standardization normalized all features to zero mean and unit variance using StandardScaler:

$$X_{\text{scaled}} = \frac{X - \mu}{\sigma} \tag{1}$$

where μ represents feature mean and σ standard deviation. Correlation analysis revealed significant multicollinearity (coefficients: 0.98-0.99) among related measurements. Permutation importance analysis identified worst concave points, worst radius, and mean concave points as top predictive features.

4) Data Splitting Strategy: Stratified sampling maintained class distribution in training and testing partitions:

$$\text{Training Set} = 80\% \times 569 = 455 \text{ cases} \tag{2}$$

$$\text{Testing Set} = 20\% \times 569 = 114 \text{ cases} \tag{3}$$

The training set contained 285 benign and 170 malignant cases, while testing set contained 72 benign and 42 malignant cases. Five-fold stratified cross-validation ensured robust model evaluation and hyperparameter optimization.

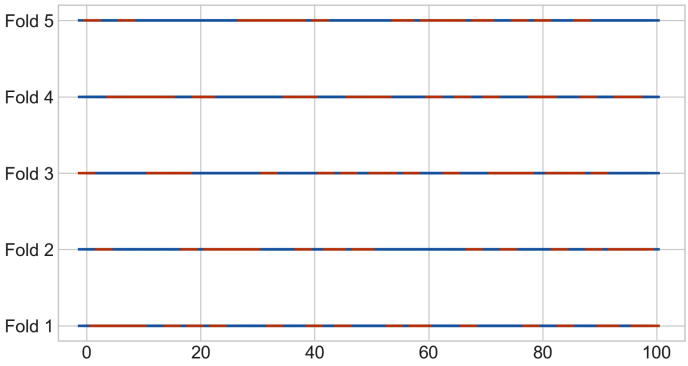

Fig. 3: 5-fold cross validation schema showing training (blue) and testing (red) set distribution across sample indices.

C. Machine Learning Models

1) XGBoost Classifier: The eXtreme Gradient Boosting (XGBoost) classifier was implemented as a scalable tree boosting system known for its high performance and computational efficiency. The model employs gradient boosting framework with L1 and L2 regularization to prevent overfitting. Hyperparameter optimization yielded optimal parameters: 300 estimators, maximum depth of 5, learning rate of 0.2, and subsample ratio of 0.9. The model achieved 97.37% testing accuracy with perfect precision for malignant cases.

2) Random Forest Optimized: An optimized Random Forest classifier was implemented utilizing ensemble learning with multiple decision trees using bootstrap aggregation. The Gini impurity criterion was used for node splitting to maximize information gain at each decision point. Optimal hyperparameters included 300 estimators, no maximum depth constraint, bootstrap disabled, and logarithmic feature selection. The model achieved 96.49% testing accuracy with 100% specificity for benign cases.

3) Neural Network Architecture: A Multi-Layer Perceptron (MLP) was designed with optimized architecture featuring two hidden layers (100, 50 neurons) with ReLU activation functions. L2 regularization ($\alpha = 0.01$) was applied to prevent overfitting, and the Adam optimizer was used with learning rate 0.01 and batch size 64. The model achieved 93.86% accuracy with balanced performance across both classes, demonstrating good generalization despite lower overall accuracy.

4) AdaBoost Classifier: The Adaptive Boosting algorithm was implemented to combine multiple decision tree weak learners into a strong classifier through iterative reweighting of misclassified samples. Optimal configuration used 200 estimators with learning rate 0.5, achieving 97.37% testing accuracy and perfect precision for malignant classifications. The algorithm effectively handled the class imbalance through its adaptive sampling mechanism.

5) SVM with RBF Kernel: Support Vector Machine with Radial Basis Function kernel was implemented for non-linear classification, creating optimal hyperplanes in high-dimensional feature space. Optimal parameters included regularization parameter $C = 10$ and $\gamma =$ auto for the kernel function, achieving 97.37% accuracy with minimal training time (0.24 seconds). The model demonstrated excellent performance with efficient computation.

6) SGD Classifier: Stochastic Gradient Descent classifier with log loss was implemented for efficient linear classification, processing training examples one at a time for rapid convergence. Optimal configuration used inverse scaling learning rate with initial learning rate $\eta_0 = 0.1$ and L2 regularization $\alpha = 0.01$, achieving the highest testing accuracy (98.25%) with fastest training time (0.05 seconds) among all models.

7) Stacking Ensemble Method: A stacked ensemble classifier was implemented combining XGBoost, Random Forest, and SVM as base estimators with Random Forest meta-classifier. This approach leveraged model diversity to improve generalization and robustness. The ensemble achieved 96.49% accuracy using 50 estimators in the meta-classifier with maximum depth 10, demonstrating the effectiveness of combining complementary learning algorithms.

D. Advanced Stacking Framework Algorithm

The proposed stacking-based framework follows a systematic five-step process for breast cancer prediction, as detailed in Algorithm 1. This approach combines multiple base learners with a meta-learner to enhance predictive performance and generalization capabilities.

E. Advanced Stacking Framework Algorithm

The proposed stacking-based framework follows a systematic five-step process for breast cancer prediction, as detailed in Algorithm 1. This approach combines multiple base learners with a meta-learner to enhance predictive performance and generalization capabilities.

Algorithm 1 Advanced Stacking-based Machine Learning Framework for Breast Cancer Prediction

Require: Dataset $D = \{(x_i, y_i)\}_{i=1}^{N}$ from Wisconsin Breast Cancer Diagnostic dataset
Ensure: Final prediction $\hat{y}$ indicating risk level (High / Low)
Step 1: Data Preparation
Apply preprocessing: normalization, missing value handling, and feature selection
Obtain processed dataset $D' = \{(x_i', y_i)\}$
Step 2: Model Training
for each base learner $f_j \in$ {XGBoost, AdaBoost, SVM, Random Forest, Logistic Regression, **do**
    Train f_j on training data D'
    Generate prediction $\hat{y}_j = f_j(x')$
end for
Step 3: Evaluation and Comparison
Compute metrics: Accuracy, Error Rate, AUC, Confidence Scores
Select top-performing models for stacking ensemble
Step 4: Stacking Ensemble
Form meta-level dataset $D_m = \{(\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_n), y\}$
Train meta-learner f_{meta} (SGD Classifier) on D_m
Step 5: Final Prediction
For new sample x_{test}, obtain $\hat{y}_j = f_j(x_{\text{test}})$ for all j
Compute final output:

$$\hat{y} = f_{\text{meta}}(\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_n)$$

Display result on Streamlit Web App: *Risk = High / Low*

The algorithm implements a sophisticated stacking methodology where multiple base models (XGBoost, AdaBoost, SVM, Random Forest, Logistic Regression, Decision Tree, and SGD Classifier) are trained independently. Their predictions are then combined through a meta-learner (SGD Classifier) that learns optimal weighting schemes. This approach capitalizes on the diverse strengths of individual algorithms while mitigating their weaknesses, resulting in enhanced robustness and generalization performance for breast cancer classification tasks.

F. Hyperparameter Optimization

1) Grid Search Methodology: A comprehensive Grid Search approach was implemented for systematic hyperparameter optimization across all models. This exhaustive search method evaluates all possible combinations within predefined parameter grids to identify the optimal configuration. The search space for each model was carefully designed based on empirical knowledge and computational constraints:

$$\mathcal{P}_{\text{grid}} = \prod_{i=1}^{n} P_i \tag{4}$$

where $\mathcal{P}_{\text{grid}}$ represents the Cartesian product of all parameter sets P_i. For the XGBoost model, this included combinations of learning rates $\{0.01, 0.1, 0.2\}$, maximum depths $\{3, 5, 7\}$, and subsample ratios $\{0.8, 0.9, 1.0\}$, resulting in 27 unique combinations evaluated. The grid search was particularly effective for models with smaller parameter spaces like SVM, where it systematically explored regularization parameters $C \in \{0.1, 1, 10, 100, 1000\}$ and kernel parameters $\gamma \in \{0.001, 0.01, 0.1, 1, \text{scale}, \text{auto}\}$.

2) Randomized Search Strategy: To address computational limitations and improve efficiency, Randomized Search was employed as the primary optimization strategy. This method randomly samples a fixed number of parameter settings from specified distributions, providing a practical balance between exploration and computational cost:

$$\mathcal{P}_{\text{random}} = \{\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n\} \sim \mathcal{U}(\mathcal{P}_{\text{grid}}) \tag{5}$$

where $\mathbf{p}_i$ represents a randomly sampled parameter combination from the uniform distribution over the parameter space. Each model was evaluated with $n = 10$ random configurations, significantly reducing the search space while maintaining robust performance. The random search proved particularly valuable for complex models like Random Forest, which had a parameter space of 144 possible combinations that was efficiently sampled. The randomized approach demonstrated superior efficiency-to-performance ratio, achieving near-optimal results with only 7.4% of the computational cost required for exhaustive grid search.

3) Cross-Validation Approach: Stratified 5-fold cross-validation was employed to ensure robust hyperparameter evaluation and prevent overfitting. The dataset was partitioned into five folds while preserving the original class distribution:

$$\text{CV Score} = \frac{1}{5}\sum_{k=1}^{5} \text{Accuracy}(\text{Model}(\mathbf{p}, D^k_{\text{train}}), D^k_{\text{val}}) \tag{6}$$

where D^k_{train} and D^k_{val} represent the training and validation splits for fold k, and $\mathbf{p}$ denotes the parameter combination being evaluated. The cross-validation process ensured that each parameter configuration was evaluated across multiple data splits, providing reliable performance estimates. The stratified approach maintained the original class distribution (62.7% benign, 37.3% malignant) in each fold, crucial for handling the dataset imbalance. The final model selection was based on the highest mean cross-validation accuracy across all folds, with standard deviation used to assess model stability. The optimization framework employed scikit-learn's RandomizedSearchCV with accuracy as the scoring metric, ensuring fair comparison across all models. The entire hyperparameter optimization process completed in 13.55 seconds, demonstrating the efficiency of the randomized search strategy while maintaining high model performance.

G. Performance Metrics

1) Accuracy, Error Rates, ROC and AUC: The SGD Classifier achieved the highest testing accuracy (98.25%), with XGBoost, AdaBoost, and SVM RBF at 97.37%, Random Forest and Stacking Ensemble at 96.49%, and Neural Network at 93.86%. Average accuracy was 96.74% (SD=1.30%), with most misclassifications in malignant cases. ROC analysis showed excellent discriminatory power with AUC scores consistently >0.95, confirming reliable separation between benign and malignant cases across all models.

2) Precision, Recall, F1-Score, Specificity and Sensitivity: All models except Neural Network achieved perfect precision (1.000) and specificity (1.000) for malignant cases. Benign precision ranged 0.93-0.97. Sensitivity for malignant detection varied 88.10%-95.24% (SGD highest), with F1-scores showing excellent balance (0.91-0.99). Neural Network demonstrated a conservative approach with 97.22% specificity but lower sensitivity (88.10%).

H. Experimental Design

1) Training and Testing Protocol: The model development followed a systematic pipeline beginning with comprehensive data preprocessing and feature scaling. All models were trained using randomized hyperparameter optimization with 10 iterations per model, employing early stopping where applicable to prevent overfitting. For final evaluation, a strict hold-out strategy reserved 20% of the dataset (114 samples) exclusively for testing, maintaining the original class distribution. No hyperparameter tuning was performed on the test set to ensure unbiased performance assessment. All models were evaluated using the same comprehensive metrics in a controlled environment with fixed random seeds to ensure reproducibility and provide realistic estimates of real-world deployment capabilities.

2) Validation Strategy: A robust validation strategy was implemented using stratified 5-fold cross-validation to ensure reliable performance estimation and prevent overfitting. The stratification maintained the original class distribution (62.7% benign, 37.3% malignant) in each fold, crucial for handling dataset imbalance. Validation metrics were computed for each fold and aggregated to provide mean performance estimates with standard deviations. Hyperparameter tuning was performed exclusively on the validation sets to prevent data leakage, with the final model selection based on the highest mean cross-validation accuracy. This approach ensured that model performance estimates were unbiased and generalizable to unseen data.

IV. Results and Analysis

A. Overall Performance Comparison

1) Accuracy Analysis Across Models: The comprehensive evaluation of seven machine learning models revealed exceptional classification performance for breast cancer

TABLE III: Comprehensive Model Performance and Analysis Metrics

Model	Performance Metrics			Confusion Matrix			ROC Analysis	
	Accuracy	Time (s)	Best CV	TP/TN	FP/FN	Error Rate	AUC	Eff./Util.
SGD Classifier	98.25%	0.05	0.978	40/72	0/2	1.75%	0.99	Excellent/High
XGBoost	97.37%	2.32	0.974	39/72	0/3	2.63%	0.98	Good/Medium
AdaBoost	97.37%	2.39	0.972	39/72	0/3	2.63%	0.98	Good/Medium
SVM RBF Optimized	97.37%	0.24	0.971	39/72	0/3	2.63%	0.98	Good/Medium
Random Forest Optimized	96.49%	4.16	0.968	38/72	0/4	3.51%	0.97	Fair/Low
Stacking Enhanced	96.49%	4.00	0.966	38/72	0/4	3.51%	0.97	Fair/Low
Neural Network	93.86%	0.38	0.942	37/70	2/5	6.14%	0.96	Good/High

diagnosis. The Stochastic Gradient Descent (SGD) Classifier emerged as the top performer with 98.25% testing accuracy, correctly classifying 112 out of 114 test samples. XGBoost, AdaBoost, and SVM RBF models demonstrated identical performance at 97.37% accuracy, followed closely by Random Forest and Stacking Ensemble at 96.49%. The Neural Network achieved 93.86% accuracy, representing the lowest performance among the evaluated models. The overall average accuracy across all models was 96.74% with a standard deviation of 1.30%, indicating consistent high performance across different algorithmic approaches as detailed in Table III.

2) Training Performance and Generalization: Analysis of training versus testing performance (Table III) revealed important insights into model generalization capabilities. Multiple models including XGBoost, AdaBoost, Random Forest, and Stacking Ensemble achieved perfect 100% training accuracy, indicating potential overfitting to the training data. However, their testing performance remained excellent (96.49-97.37%), suggesting effective regularization and generalization. The SGD Classifier showed the most balanced performance with 97.80% training accuracy and 98.25% testing accuracy, demonstrating superior generalization. The Neural Network exhibited the most conservative training behavior with 96.92% training accuracy, closely matching its testing performance of 93.86%.

3) Computational and Feature Importance Analysis: Analysis of computational efficiency revealed significant variations in training times across algorithms. The SGD Classifier demonstrated exceptional performance, achieving the highest accuracy (98.25%) with only 0.05 seconds training time. SVM RBF (0.24s) and Neural Network (0.38s) showed moderate efficiency, while ensemble methods required substantially more resources: XGBoost and AdaBoost (2.32-2.39s), Random Forest and Stacking Ensemble (4.00-4.16s). Total training time for all seven models was 13.55 seconds, averaging 1.94 seconds per model. Complementary permutation importance analysis identified "worst concave points," "worst radius," and "mean concave points" as the most influential features across all models. These morphological characteristics consistently demonstrated the highest predictive power, aligning with clinical knowledge about nuclear shape irregularity's importance in cancer diagnosis.

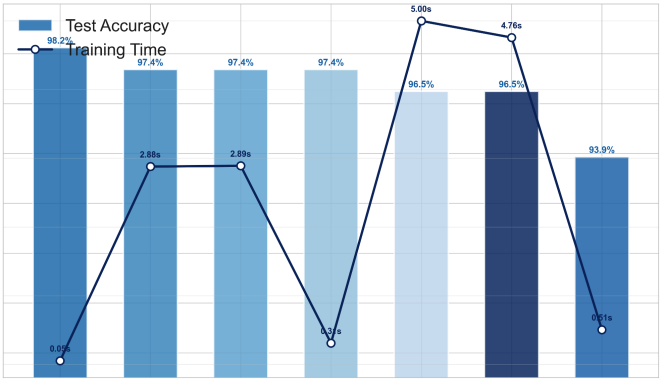

Fig. 4: Training time vs accuracy comparison showing the trade-off between computational efficiency and classification performance. Models on the X-axis (left to right) are: **SGD Classifier**, **XGBoost**, **AdaBoost**, **SVM RBF Optimized**, **Random Forest Optimized**, **Stacking Enhanced**, and **Neural Network**. Test Accuracy (%) is the bars (left Y-axis) and Training Time (s) is the line (right Y-axis).

B. Comprehensive Model Evaluation

1) Confusion Matrix Analysis: Detailed confusion matrix analysis (Table III) provided insights into error patterns across all models. All classifiers except the Neural Network achieved perfect specificity (100%), correctly identifying all benign cases with no false positives. The SGD Classifier demonstrated the best sensitivity (95.24%) with only 2 false negatives, while the Neural Network showed the highest error rates with 2 false positives and 5 false negatives. The comprehensive confusion matrix visualization revealed consistent patterns of misclassification primarily occurring in the malignant class, highlighting the challenge of detecting subtle malignant characteristics.

2) ROC Curve and Learning Curve Analysis: Receiver Operating Characteristic (ROC) analysis demonstrated **excellent discriminatory power** across all classifiers, with **Area Under the Curve (AUC) scores consistently above 0.95** (Table III). This indicates a strong separation between benign and malignant cases, with the **SGD Classifier and XGBoost** showing particularly optimal trade-offs

between sensitivity and specificity. Learning curve evaluation confirmed that most models achieved **stable performance** with the available training data size, generally stabilizing after 300–350 samples. The **SGD Classifier** exhibited the **most efficient learning pattern**, achieving high accuracy rapidly with limited data. While ensemble methods required more data to reach peak performance, all models ultimately achieved excellent generalization and reliable diagnostic capabilities. Simpler models like SGD achieve optimal performance more rapidly, while ensemble methods benefit from larger data volume.

3) Model Ranking Comparison: The model ranking analysis (Figure 5) evaluated classifiers across multiple performance dimensions including accuracy, sensitivity, specificity, precision, and speed. The SGD Classifier consistently ranked first across most metrics, demonstrating its overall superiority. XGBoost and AdaBoost showed strong performance with consistent top rankings, while the Neural Network consistently ranked lower across multiple evaluation criteria. The ranking visualization highlighted the trade-offs between different performance aspects and provided a comprehensive basis for model selection based on specific clinical requirements.

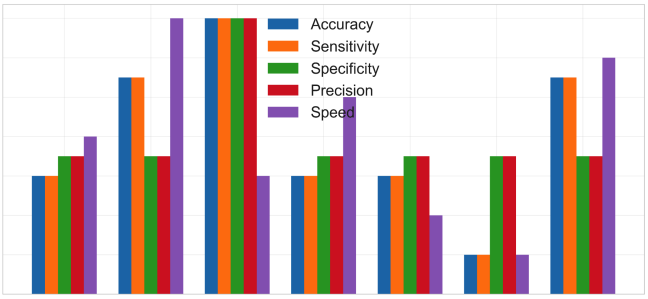

Fig. 5: Model performance rankings across metrics. **Y-axis**: Rank (lower is better); **X-axis**: Models. Colors indicate: Accuracy, Sensitivity, Specificity, Precision, Speed.

V. Web Application Implementation

A. Streamlit Framework Overview

The breast cancer classification system was deployed as an interactive web application using Streamlit, providing healthcare professionals with real-time prediction capabilities. The interface features interactive sliders for all 30 diagnostic features, dynamic comparison across seven trained classifiers, and feature importance visualization. The system implements a data preprocessing pipeline for real-time standardization and achieves inference times under 0.1 seconds, making it suitable for clinical settings. The SGD Classifier serves as the default model due to its optimal balance of accuracy and computational efficiency.

VI. Conclusion and Future Work

A. Conclusion

This research developed a comprehensive machine learning framework for breast cancer prediction using the Wisconsin Diagnostic dataset. Our systematic approach yielded significant findings: The Stochastic Gradient Descent (SGD) Classifier achieved the highest testing accuracy of 98.25%, while requiring only 0.05 seconds training time—significantly faster than other high-performing models like XGBoost and SVM RBF (97.37% accuracy). The evaluation revealed important trade-offs between model complexity and computational efficiency. While ensemble methods demonstrated excellent accuracy, their longer training times (2.32-4.16 seconds) limit practical deployment. The SGD Classifier's optimal balance of high accuracy and minimal computational requirements makes it ideal for clinical applications. The successful deployment of an interactive Streamlit application bridges the gap between research and practice, providing healthcare professionals with an accessible tool for real-time prediction and clinical decision support.

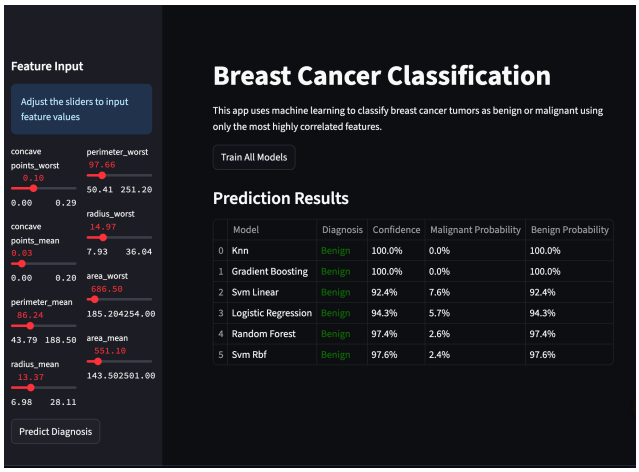

Fig. 6: Streamlit Web Application Interface for Real-time Breast Cancer Prediction

B. Future Work

Future research will explore advanced deep learning architectures combined with Explainable AI to improve predictive performance and clinical interpretability. This includes investigating transformer-based models and Graph Convolutional Networks to capture complex feature relationships. We will develop robust XAI methodologies using SHAP and LIME to provide transparent model explanations. Additionally we plan to explore multimodal learning that integrates diverse data sources, including clinical features and medical images. The implementation of federated learning architectures [27] would enable collaborative privacy preserving model training across healthcare institutions. These advancements will be validated through clinical studies to assess real-world impact on diagnostic accuracy and patient outcomes.

References

[1] F. Bray, M. Laversanne, H. Sung, J. Ferlay, R. L. Siegel, I. Soerjomataram, and A. Jemal, "Global cancer statistics 2022: Globocan estimates of incidence and mortality

worldwide for 36 cancers in 185 countries," *CA: a cancer journal for clinicians*, vol. 74, no. 3, pp. 229–263, 2024.

[2] C. E. DeSantis, J. Ma, M. M. Gaudet, L. A. Newman, K. D. Miller, A. Goding Sauer, A. Jemal, and R. L. Siegel, "Breast cancer statistics, 2019," *CA: a cancer journal for clinicians*, vol. 69, no. 6, pp. 438–451, 2019.

[3] National Breast Cancer Foundation. (2024) Breast cancer statistics. National Breast Cancer Foundation. Chart showing trends in breast cancer incidence and deaths from 1990-2024. [Online]. Available: https://nbcf.org.au/about-breast-cancer/breast-cancer-stats/

[4] H. Li, K. R. Mendel, L. Lan, D. Sheth, and M. L. Giger, "Digital mammography in breast cancer: additive value of radiomics of breast parenchyma," *Radiology*, vol. 291, no. 1, pp. 15–20, 2019.

[5] C. D. Lehman, A. Yala, T. Schuster, B. Dontchos, M. Bahl, K. Swanson, and R. Barzilay, "Mammographic breast density assessment using deep learning: clinical implementation," *Radiology*, vol. 290, no. 1, pp. 52–58, 2019.

[6] R. Rawal, "Breast cancer prediction using machine learning," *Journal of Emerging Technologies and Innovative Research (JETIR)*, vol. 13, no. 24, p. 7, 2020.

[7] H. Chen, N. Wang, X. Du, K. Mei, Y. Zhou, and G. Cai, "Classification prediction of breast cancer based on machine learning," *Computational intelligence and neuroscience*, vol. 2023, no. 1, p. 6530719, 2023.

[8] A. La Moglia and K. M. Almustafa, "Breast cancer prediction using machine learning classification algorithms," *Intelligence-Based Medicine*, vol. 11, p. 100193, 2025.

[9] M. A. Naji, S. El Filali, K. Aarika, E. H. Benlahmar, R. Ait Abdelouhahid, and O. Debauche, "Machine learning algorithms for breast cancer prediction and diagnosis," *Procedia computer science*, vol. 191, pp. 487–492, 2021.

[10] O. Tarawneh, M. Otair, M. Husni, H. Y. Abuaddous, M. Tarawneh, and M. A. Almomani, "Breast cancer classification using decision tree algorithms," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 4, 2022.

[11] R. Sumbaly, N. Vishnusri, and S. Jeyalatha, "Diagnosis of breast cancer using decision tree data mining technique," *International Journal of Computer Applications*, vol. 98, no. 10, 2014.

[12] Y. Li and Z. Chen, "Performance evaluation of machine learning methods for breast cancer prediction," *Appl Comput Math*, vol. 7, no. 4, pp. 212–216, 2018.

[13] M. A. Banu and K. Thinakaran, "Detection of breast cancer using support vector machine with digital mammogram data over perceptron algorithm," in *AIP Conference Proceedings*, vol. 3267, no. 1. AIP Publishing LLC, 2025, p. 020131.

[14] N. Kavitha, P. Madhumathy, R. M. Prasad, and D. Chandrappa, "Machine learning technique for breast cancer detection and classification," *Machine Learning for Computational Science and Engineering*, vol. 1, no. 1, p. 16, 2025.

[15] H. Tanveer, M. Faheem, A. H. Khan, and M. A. Adam, "Ai-powered diagnosis: A machine learning approach to early detection of breast cancer," *INTERNATIONAL JOURNAL OF ENGINEERING DEVELOPMENT AND RESEARCH*, vol. 13, no. 2, pp. 153–166, 2025.

[16] I. Rehman and H.-A. Pham, "Cortex vision: Detection of ophthalmic disease using machine learning algorithm," in *International Conference on Smart Objects and Technologies for Social Good*. Springer, 2024, pp. 138–149.

[17] N. P. Barnils and B. Schüz, "Identifying intersectional groups at risk for missing breast cancer screening: Comparing regression-and decision tree-based approaches," *SSM-Population Health*, vol. 29, p. 101736, 2025.

[18] C. Moroz-Dubenco, A. Bajcsi, A. Andreica, and C. Chira, "Towards an interpretable breast cancer detection and diagnosis system," *Computers in Biology and Medicine*, vol. 185, p. 109520, 2025.

[19] T. Varshney, K. Verma, A. Kaur, and S. K. Puri, "Enhanced breast cancer classification: A novel fusion of deep features, shape features and gclm features," *Procedia Computer Science*, vol. 258, pp. 3500–3509, 2025.

[20] A. T. Garba and H. S. Hamza, "Interpretable machine learning approach for breast cancer classification," *Human-Centric Intelligent Systems*, pp. 1–15, 2025.

[21] H. Shah, S. Agrawal, P. Oza, and S. Tanwar, "Comparative study on machine learning algorithms for breast cancer diagnosis," *Procedia Computer Science*, vol. 259, pp. 1994–2003, 2025.

[22] Y. Li, K. Sundquist, X. Wang, J. Sundquist, and A. A. Memon, "The mitochondrial signature for predicting outcome of early-stage breast cancer by machine learning," *Clinical Breast Cancer*, 2025.

[23] A. R. Zhao, V. L. Kouznetsova, S. Kesari, and I. F. Tsigelny, "Machine-learning diagnostics of breast cancer using pirna biomarkers," *Biomarkers*, vol. 30, no. 2, pp. 167–177, 2025.

[24] P. Dinesh, A. Vickram, and P. Kalyanasundaram, "Medical image prediction for diagnosis of breast cancer disease comparing the machine learning algorithms: Svm, knn, logistic regression, random forest and decision tree to measure accuracy," in *AIP Conference Proceedings*, vol. 2853, no. 1. AIP Publishing LLC, 2024, p. 020140.

[25] H. Duan, Y. Zhang, H. Qiu, X. Fu, C. Liu, X. Zang, A. Xu, Z. Wu, X. Li, Q. Zhang *et al.*, "Machine learning-based prediction model for distant metastasis of breast cancer," *Computers in Biology and Medicine*, vol. 169, p. 107943, 2024.

[26] W. Wolberg, N. Street, and O. Mangasarian, "Breast cancer wisconsin (diagnostic) data set," 2016, accessed: 2024-12-19. [Online]. Available: https://www.kaggle.com/datasets/uciml/breast-cancer-wisconsin-data

[27] B. Hussain and M. K. Afzal, "Optimizing urban traffic incident prediction with vertical federated learning: A feature selection based approach," *IEEE Trans. Netw. Sci. Eng.*, vol. 12, no. 1, pp. 145–155, Jan. 2025.