IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 2 August 2023, accepted 14 August 2023, date of publication 24 August 2023, date of current version 9 January 2025.

*Digital Object Identifier 10.1109/ACCESS.2023.3308446*

RESEARCH ARTICLE

# An Explainable Artificial Intelligence Model for the Classification of Breast Cancer

**TAREK KHATER[1], ABIR HUSSAIN[1,2], (Senior Member, IEEE), RIYAD BENDARDAF[3,4], IMAN M. TALAAT[3,5], HISSAM TAWFIK[1], (Senior Member, IEEE), SAM ANSARI[1], AND SOLIMAN MAHMOUD[1,6], (Senior Member, IEEE)**

[1]Electrical Engineering Department, University of Sharjah, Sharjah, United Arab Emirates
[2]School of Computer Science and Mathematics, Faculty of Engineering, Liverpool John Moores University, L3 3AF Liverpool, U.K.
[3]Clinical Sciences Department, University of Sharjah, Sharjah, United Arab Emirates
[4]Department of Oncology, University Hospital Sharjah, Sharjah, United Arab Emirates
[5]Pathology Department, Faculty of Medicine, Alexandria University, Alexandria 5424041, Egypt
[6]University of Khorfakan, Sharjah, United Arab Emirates

Corresponding author: Abir Hussain (abir.hussain@sharjah.ac.ae)

**ABSTRACT** Breast cancer is the most common cancer among women and globally affects both genders. The disease arises due to abnormal growth of tissue formed of malignant cells. Early detection of breast cancer is crucial for enhancing the survival rate. Therefore, artificial intelligence has revolutionized healthcare and can serve as a promising tool for early diagnosis. The present study aims to develop a machine-learning model to classify breast cancer and to provide explanations for the model results. This could improve the understanding of the diagnosis and treatment of breast cancer by identifying the most important features of breast cancer tumors and the way they affect the classification task. The best-performing machine-learning model has achieved an accuracy of 97.7% using k-nearest neighbors and a precision of 98.2% based on the Wisconsin breast cancer dataset and an accuracy of 98.6% using the artificial neural network with 94.4% precision based on the Wisconsin diagnostic breast cancer dataset. Hence, this asserts the importance and effectiveness of the proposed approach. The present research explains the model behavior using model-agnostic methods, demonstrating that the bare nuclei feature in the Wisconsin breast cancer dataset and the area's worst feature Wisconsin diagnostic breast cancer dataset are the most important factors in determining breast cancer malignancy. The work provides extensive insights into the particular characteristics of the diagnosis of breast cancer and suggests possible directions for expected investigation in the future into the fundamental biological mechanisms that underlie the disease's onset. The findings underline the potential of machine learning to enhance breast cancer diagnosis and therapy planning while emphasizing the importance of interpretability and transparency in artificial intelligence-based healthcare systems.

**INDEX TERMS** Artificial intelligence, breast cancer, explainable machine learning, model-agnostic, permutation importance, partial dependence plot, SHAP, supervised learning.

## I. INTRODUCTION

According to the World Health Organization (WHO), breast cancer poses the most significant risk to women [1], since it is the leading cause of cancer mortality in women worldwide. Reducing this high mortality rate requires early detection of the disease. With 2.3 million new cases annually or 11.7% of all new cancer cases in 2020 [2], breast cancer is the most frequently diagnosed malignancy. Additionally, 7.8 million women alive today have received a breast cancer diagnosis in the last five years [3].

In the United Arab Emirates (UAE), breast cancer is the most common cancer, especially in women younger than 50 years. From around 4000 cases diagnosed with cancer in the UAE, 21% have been diagnosed with breast cancer [1]. Breast cancer results from the aberrant tissue growth formed

The associate editor coordinating the review of this manuscript and approving it for publication was Mehul S. Raval.

by cancerous cells [4]. Regular breast screening can aid in early detection and permit treatment, particularly for those with a high or moderate breast cancer risk [4].

Hence, the earlier the detection of breast cancer, the higher the possibility of treatment, and the better the chances of survival. Early research has significantly helped in the treatment of breast cancer because scientists were aware of the risks posed by emphasizing cancer from the onset. The mortality rate has demonstrated a consistent and lowering trend over the past few decades thanks to research efforts and early identification techniques. According to figures from Cancer Research in the UK, the five-year survival rate for breast cancer can be as low as 15% if discovered at a later stage but is virtually 100% if discovered at an early stage. Manual diagnosis of breast cancer from the images takes a lot of time, which makes it difficult for the clinician to categorize the illness. Therefore, it is imperative to automate the diagnosis of cancer using multiple diagnostic methods.

Mammograms are currently the most used test, although they still include false positive (high-risk) results that show abnormal cells and can result in pointless biopsies and procedures. Surgery to remove lesions may occasionally discover that they are benign. This implies that the patient will undergo needless expensive and unpleasant surgery. Because of the increasing availability of structured and unstructured data and the rapid development of analytical methods, artificial intelligence (AI) is revolutionizing the healthcare industry. With the increasing importance and applications of AI in healthcare, there are growing concerns about a lack of transparency and explainability, as well as potential bias in model predictions. AI can be used to enhance the detection and diagnosis of breast cancer as well as limit overtreatment. However, combining AI with machine learning (ML) approaches makes predictions possible and facilitates precise decision-making. For instance, determining if the patient needs surgery based on the findings of the biopsy for the detection of breast cancer. Therefore, in order to make such decisions, Explainable Artificial Intelligence (XAI) is employed.

XAI is a collection of techniques and methods that can be used to explain the outcomes of the development of ML models in a way that is understandable to humans [5]. XAI encompasses two main approaches or methods: the intrinsic approach [6] in which the internal parameters of the model are used to generate explanations, and the model-agnostic approach, which is employed when the model is regarded as a black box, and the internal parameters can not be accessed. ML algorithms employed in this study are classified as black box models. Consequently, model-agnostic methods are utilized to interpret the inner workings of these models and provide a clearer understanding. Model-agnostic methods possess the capability to generate explanations independently of the internal workings of ML models that are considered ''opaque'' [6]. A key advantage of these methods is their versatility, as they can be applied to any ML model. In this paper, three model-agnostic methods have been used to provide explanations for the ML model outcome, including permutation importance, Partial Dependence Plot (PDP), and Shapley Additive Operations (SHAP).

First, the permutation importance approach [6] aims to address the most important characteristics of the model. Hence, after permuting the feature to assess its relevance, the increase in the model prediction error is calculated. A feature is deemed ''essential'' if rearranging its values leads to an elevation in model error as the feature was utilized by the model for forecasting purposes [7]. The PDP [6] is utilized to investigate how these features affect the predictions. It is a plot that demonstrates the functional connection between one or more inputs and the desired outcome. PDP shows how the most important elements could influence the evolved prediction. The relationship between the target and a feature can be shown on the PDP whether it is linear, monotonic, or more complex [7]. Finally, the SHAP method depends on Shapley values that provide explanations on specific instances instead of just global explanations [7]. SHAP has emerged and launched as a Python toolkit for ML [8]. that delivers a roster of Shapley values for a particular data point, associating with each feature. This conceptually assumes that predictions can be clarified by regarding each feature as a ''player'' in a game, where the prediction represents the reward [8].

In light of the above, the contribution of this study can be summarized as follows.

- We introduce an innovative and explainable machine learning-based model for breast cancer diagnosis, contributing to advancements in this field.
- The model accurately classifies breast cancers as benign or malignant, identifying influential factors including bare nuclei and worst area. These insights enhance understanding and provide a basis for further investigations into breast cancer mechanisms as we tried to link two different breast cancer datasets. To the best knowledge of the authors, the link between the two features has not been identified in the literature previously.
- By employing model-agnostic methods, the research addresses the need for interpretability and transparency in healthcare AI systems. The model explanations aid in understanding the decision-making process, improving breast cancer diagnosis and therapy planning practices.

The increasing need for interpretable and transparent ML models in the healthcare industry stems from the desire to enhance trust and acceptance among healthcare professionals and patients. By providing explanations and insights into their decision-making process, these models can foster greater confidence and understanding. This heightened demand arises from the recognition that transparency is a crucial aspect of deploying machine learning models in healthcare settings. Consequently, the development of interpretable models has become a vital area of research, aiming to address this need and promote the widespread adoption of ML solutions in healthcare.

The major contribution of the paper is the development of an explainable machine learning-based model for breast cancer diagnosis. The model can accurately classify cancers

as benign or malignant with high accuracy and provides insights into the most important factors that contribute to malignancy. This information can enhance the accuracy of diagnosis and facilitate the development of more effective treatment strategies.

The remainder of this paper is organized as follows: Section II provides a comprehensive review of the existing literature on breast cancer diagnosis, covering two datasets, including the Wisconsin breast cancer (WBC) dataset and Wisconsin diagnostic breast cancer (WDBC) dataset. In section III, we introduce the datasets utilized in this study, offering a concise description of each one. Section IV delves into the methodology employed for this research. Subsequently, in section V, we present the simulation results and engage in a thorough discussion of their implications. Finally, section VI summarizes the key conclusions derived from this study and outlines potential avenues for future research.

## II. LITERATURE REVIEW

This section presents a comprehensive overview of the existing literature and research that is relevant to the research question, which focuses on the classification of breast cancer using explainable ML techniques. The section is divided into two parts: a literature review on the use of ML for breast cancer classification using clinical datasets such as the WBC and WDBC, and a discussion of relevant related works on ML-based detection of the disease using genetic breast cancer datasets. This examination aims to provide a comprehensive understanding of the research question.

### A. CLINICAL BREAST CANCER DATASETS

There are a lot of works that predict or classify breast cancer based on clinical datasets, such as the WBC and WDBC datasets. For instance, Alshayeji et al. [9] aimed to classify breast tumors based on the WBC and WDBC datasets using an artificial neural network (ANN). The ANN model contains one hidden layer without employing feature selection or optimization techniques. In the WBC dataset, the shallow ANN model performed well, with an average precision of 99.85%. The average accuracy for breast cancer detection with WDBC was 99.47%. However, the use of 100 neurons in one hidden layer may result in longer training convergence time, and the model may not effectively capture the complex features in the dataset. Khandaker et al. [10] developed an explainable ML model to predict breast cancer. Initially, they employed gradient-boosting algorithms to train their model, resulting in a 99% accuracy rate for light gradient boosting (LGBM) as their best-performing model. Additionally, the authors applied the SHAP method to provide interpretations for their model and to investigate the impact and contribution of each feature in the dataset. Afolayan et al. [11] used Particle Swarm Optimization (PSO) to optimize the performance of the Decision Tree (DT) algorithm on the WBC dataset. The results showed that the system achieved an accuracy of 92.26%, helping to minimize the incidence of breast cancer by providing early detection and diagnosis.

Birchha and Nigam [12] used the averaged perceptron ML algorithm to classify breast cancer based on the WBC dataset. Their ML model achieved 98.4% accuracy with zero false negative classifications which means the recall equals one. Hence, the averaged perceptron ML classifier can provide accurate breast cancer classification with zero positive or negative classifications. One of the limitations of this study is that the testing set is small which is 12% leading to unreliability results. Singh et al. [13] proposed a new feature selection method based on the Eagle Strategy (ESO) Optimization, Gravitational Search Optimization (GSO) algorithm, and their hybrid algorithm. The method was used to classify breast cancer into two groups using the WDBC dataset. The results showed that the proposed hybrid algorithm achieved an accuracy of 98.9578%, sensitivity of 0.9705, specificity of 1.000, precision of 1.000, F1-score of 0.9696, and Area Under the Curve (AUC) of 0.9980. The authors indicated that the proposed method could be used to develop a clinical prediction system for breast cancer. A big data-based two-class breast cancer (BC) classification model was developed by Saad et al. [14] using Deep Reinforcement Learning (DRL). The model's stages include data collection, preprocessing, feature selection, classification, and explanations. Gorilla Troops Optimization (GTO) algorithm was used for feature selection, Deep Q learning (DQL) for classification, and LIME for explanation. The model underwent evaluation on three datasets from the UCI repository: WBC, WDBC, and WPBC (Wisconsin Prognostic Breast Cancer). The authors claimed that their proposed model outperformed the traditional methods, achieving 98.90% accuracy for the WBC dataset, 99.02% for WDBC, and 98.88% for the WPBC dataset, respectively. To extract features and use ANN to classify the images, Tahmooresi et al. [15] and Salma et al. [16] chose two distinct datasets from WBC and KDD, and they both employed the Factorization Machine ANN (FM-ANN). The authors contrasted the outcomes with those of other methods, namely Radial Basis Function Network (RBF), Feedforward Neural Network (FNN), and Modular Neural Network (MNN). Due to the higher number of features, KDD achieved a superior accuracy of 99.96% after training and testing. When comparing the outcomes, FM-ANN was shown to be more precise. Additionally, it is worthwhile to assess the computational efficiency and scalability of FM-ANN in relation to other ML techniques.

In another study, Khuriwal et al. [17] used deep learning to help in breast cancer diagnosis based on the WBC dataset, achieving 99.67% accuracy but around 93% precision, which is not as good as the other algorithms. However, they used certain pre-processing algorithms such as label encoder, normalizer, and StandardScaler for scaled datasets before training the model. Ahmed et al. [18] evaluated the efficacy of several variables of the original WDBC for predicting breast cancer diagnosis using various ML classification algorithms to properly forecast the target class and enhance it. These algorithms include naive Bayes (NB), multilayer perceptron (MLP), random forest (RF), and J48. Performance

parameters, including accuracy, precision, kappa statistic, f1-score, Matthews correlation coefficient (MCC), recall, receiver operating characteristic (ROC), and precision-recall curve (PRC), were utilized to compare the results. Among the employed algorithms, the NB classifier produced the best results based on the values of the performance indicators with 97.2779% accuracy. NB assumes independence between the features, and the features in such a dataset have a kind of correlation. Hence, other ML algorithms such as support vector machine (SVM), RF, and XG-boost capture the complex relationships between the features.

Predictions regarding the types of breast tumors were made using data from the WBC dataset on breast cancer tumors by Ak [19]. Various ML methods such as k-nearest neighbors (KNN), logistic regression (LR), DT, RF, SVM, NB, and rotation forests were employed along with data visualization techniques. The LR model with all features yielded the highest classification accuracy of 98.1%, and the proposed method demonstrated improved accuracy performance. Rahman et al. [20] conducted a comparative analysis of different ML methods, including SVM, DT, NB, and KNN. They performed research on the WBC dataset using adaptive boosting (AdaBoost), extreme gradient boosting (XGBoost), and RF. The primary objective was to assess the accuracy, precision, specificity, and sensitivity of data classification achieved by each algorithm, considering their effectiveness and efficiency. Based on the experimental findings, XGBoost exhibited the highest accuracy of 98.24% and the lowest error rate.

Magdy et al. [21] introduced an optimized framework for identifying breast cancer types and predicting breast cancer recurrence using seven ML algorithms: LR, XGboost, NB, RF, KNN, DT, and multilayer perception (MLP) of neural network. Grid search was employed to optimize the ML algorithms.The framework's performance was evaluated on the following Wisconsin datasets: the WBC dataset, the WDBC dataset, and the Wisconsin prognosis breast cancer (WPBC) dataset to determine the best-performing classifier. The results showed an accuracy of 98.3% for the WBC dataset, 99.2% for the WDBC dataset, and 78.6% for the WPBC dataset in predicting cancer recurrence. In another study, Islam et al. [22] compare five supervised ML techniques, including KNN, SVM, ANNs, RF, and LR. They assessed the effectiveness of the different ML algorithms in terms of accuracy, precision, sensitivity, specificity, negative predictive value, false positive rate, F1 score, false negative rate, and Matthews Correlation Coefficient. Additionally, the PRC, AUC, and ROC were evaluated for various strategies. The findings show that SVM received accuracy, precision, and F1 scores of 97.14%, 95.65%, and 0.9777, respectively, while ANNs obtained the highest scores of 98.57%, 97.82%, and 0.9890, respectively. Mridha [23] applied many ML algorithms such as gradient booster, SVM, NB, LR, RF, KNN, and ANN. Each of these algorithms' accuracy, cross-validation, sensitivity, and specificity gains were calculated and compared. They concluded from the trials that KNN has the least accuracy (91.22%), whereas RF has the best accuracy (98.83%). The accuracy of predictions has been increased using deep learning algorithms ANN. Overall accuracy in the ANN example was 99.73%, correspondingly.

Ara et al. [24] objective is to examine the dataset and assess how well different ML algorithms perform at predicting breast cancer. To categorize tumors into benign and malignant types, SVM, LR, KNN, NB, DT, and RF classifiers have been used. To choose the best algorithm, the accuracy of each is calculated and compared. Based on the investigation, SVM and RF outperform other classifiers with an accuracy of 96.5%. Durai et al. [25] selected data mining for disease detection, specifically breast cancer. The authors compared a linear regressive classifier (LRC) with BFI, Iterative Dichotomiser 3 (ID3), J48, and SVM. The results indicate that LRC achieved the highest accuracy of 99.25%. Six alternative SVM algorithms were worked on by Azar et al. [26]. In order to evaluate the performance in terms of accuracy, sensitivity, specificity, and ROC, they compared standard SVM (ST-SVM) with linear programming SVM (LPSVM), Lagrangian SVM (LSVM), smooth SVM (SSVM), proximal SVM (PSVM), and finite Newton SVM (NSVM). LPSVM demonstrated the best performance with an accuracy of 97.1429%, sensitivity of 98.2456%, specificity of 95.082%, and ROC of 99.38%. Therefore, LPSVM exhibits the highest performance. Deng et al. [27] utilized a novel technique called the weighted hierarchical adaptive voting ensemble (WHAVE). They contrasted WHAVE's precision with seven other techniques that had the best precisions in earlier studies. The maximum performance value of 99.8% was achieved by WHAVE. Egwom et al. [28] developed an ML algorithm to categorize breast cancer. SVMs were used for classification, and linear discriminant analysis (LDA) was used for feature extraction to accomplish this. When LDA was employed, and the median was utilized to compute missing values, they used two datasets, WBC and WPBC. On the WBC dataset, they achieved an accuracy of 99.2%, recall of 98.0% and precision of 98.0%, and accuracy of 79.5%, recall of 76.0%, and precision of 59.0% on the WPBC dataset.

Manikandan et al. [29] proposed a practical approach based on ML for classifying the SEER breast cancer dataset. The researchers employed a two-step feature selection method, which combined variance threshold and principal component analysis, to identify relevant features from the SEER breast cancer dataset. supervised and ensemble learning techniques such as Ada, XG, gradient, NB, and DT were utilized to classify the dataset. The performance of various ML algorithms was assessed using both the train-test split and k-fold cross-validation methods. The DT algorithm achieved an accuracy of 98% in both the train-test split and cross-validation, outperforming other supervised and ensemble learning algorithms in this study on the SEER dataset.

Hou et al. [30] conducted a study to evaluate and compare the predictive performance of four ML algorithms for

detecting breast cancer among Chinese women. The study utilized a dataset comprising 7127 breast cancer cases and 7127 matched healthy controls for model training and testing. Model performance metrics such as AUC, sensitivity, specificity, and accuracy were calculated using repeated five-fold cross-validation. Among the three advanced ML algorithms (XGBoost, RF, and deep neural network), all three outperformed LR in terms of accuracy, sensitivity, and area under the ROC curves (ROC AUC). XGBoost exhibited the highest performance with an AUC of 0.742, followed by the RF and the deep neural network with an AUC of 0.728, 0.742, respectively.

Wang et al. [31] employed Microwave Tomography Imaging (MTI). In this study, the two methodologies Gaussian Mixture Modeling (GMM) and KNN were contrasted. According to their findings, KNN has a sensitivity of 87%, compared to 67% for GMM. Accuracy was at 85% for KNN and 75% for GMM, respectively. Because mammography scans are less expensive, Massari et al. [32] proposed an ontological ML model to predict breast cancer based on the DT algorithm. The approach involves deriving rules from the DT algorithm that differentiates malignant and benign breast cancer patients. These rules are subsequently applied to the ontological reasoner using the Semantic Web Rule Language. They demonstrated that the ontological model attained a prediction accuracy of 97.10%. Tanzeel et al. [33] employed diverse ML methods to predict and detect breast cancer symptoms early. The utilized algorithms were DT, KNN, Multilayer Perceptron (MLP) classifiers, SVM, and RF. Their aim was to differentiate between benign and malignant cancer cells. Their results revealed that the MLP model exhibited the highest accuracy of 86% compared to the other techniques examined.

Rabiei et al. [34] tried to predict breast cancer using different ML approaches applying demographic, laboratory, and mammographic data. In this analytical investigation, the database from Motamed Cancer Institute (ACECR), Tehran, Iran, had 5,178 independent records, 25% of which belonged to breast cancer patients, and each record contained 24 attributes. This study made use of RF, MLP, gradient boosting trees (GBT), and genetic algorithms (GA). Models were initially trained using laboratory and demographic data (20 features). When compared to other approaches, RF performed better (80% accuracy, 95% sensitivity, 80% specificity, and an AUC of 0.56). Gradient boosting (AUC=0.59) outperformed the neural network in terms of performance. Mugahed et al. [35] used mammograms to identify breast mass using deep learning, particularly You-Only-Look-Once (YOLO) approach. Initially, they used a full-resolution convolutional network for mammogram segmentation. Subsequently, a CNN model was trained on the INbreast dataset to detect and classify the masses as benign or malignant. The findings reveal that overall accuracy is 98.96% and F1-score of 99.24%m using 4 fold-cross validation. Furthermore, the utilization of FrCN demonstrated an overall accuracy of 92.97% and F1-score of 92.69%. The performance of the CNN model was evaluated, resulting in an accuracy of 95.64%, AUC of 94.78%, and F1-score of 96.84%. Massafra et al. [36] developed an XAI framework to understand breast cancer invasive disease events (IDEs) such as second cancers, contralateral, and recurrence, The study was performed on 486 breast cancer patients enrolled at IRCCS Istituto Tumori ''Giovanni Paolo II'' in Bari, Italy. They designed an ML model to predict the IDEs using SVM, RF, NB, and XG-Boost. The best-performing model was XG-Boost with AUC values equal to 93.7% and 91.7% for the 5-year and 10-year IDE predictions, respectively. The authors determined the main influencers behind the IDE by analyzing the Shapley values within two widely employed timeframes in clinical settings: 5 years and 10 years from the initial tumor diagnosis. Maouche et al. [37] designed an XAI model for predicting breast cancer metastasis using clinicopathological data. They trained their model using the CatBoost classifier achieving precision of 76.5%, recall of 79.5%, and f1-score of 77%. The LIME method assessed patient and treatment effects on breast cancer metastasis, uncovering varying impacts. High impact factors include no adjuvant chemotherapy, whereas moderate impact encompasses medullary histological type. Low-impact factors include oral contraception usage. Silva-Aravena et al. [38] proposed a decision support strategy for health teams based on ML tools and XAI. Their findings showed that XG-Boost was the best-performing algorithm with an accuracy of 81%. In order to identify the relevant variables and their level of significance in the prediction and quantify the impact of these features on the clinical condition of the patients, the researchers used the SHAP. They claimed that the results would allow health teams to offer early and personalized alerts for each patient.

It should be noted that our work is distinguished from all the aforementioned works, which employ ML models to predict or classify breast cancer. Our primary focus lies in offering comprehensive explanations and interpretations for the outcomes generated by the ML model.

### B. GENETIC BREAST CANCER DATASETS

Various researchers have utilized genetic datasets for the prediction and classification of breast cancer. The previous studies employed an ML approach to predict the risk of breast cancer by identifying the combination of interacting genetic variants known as single nucleotide polymorphisms (SNPs) and demographic risk factors. The research focused on two distinct groups: group 1, which consisted of factors associated with familial history, and group 2, which pertained to estrogen metabolism. The objective was to determine the interactions between genetic and demographic risk factors that would yield the highest accuracy in predicting breast cancer risk. By incorporating both interacting genetic features and group one features, the proposed approach achieved a mean average precision (mAP) of 77.78 on the Kuopio Breast Cancer Project (KBCP) dataset. This performance surpassed the mAPs obtained when using only group one

features (74.19) or interacting SNPs (73.65). When considering solely group two features, the system achieved an mAP of 72.57. However, integrating interacting genetic features with group two features resulted in an improved mAP of 78.00. Furthermore, the study generated gene interaction maps based on genes associated with SNPs that interacted with demographic risk factors. These maps revealed biologically significant entities relevant to breast cancer, such as networks associated with angiogenesis, apoptosis, and estrogen. Interestingly, the findings also indicated that individual demographic risk factors possess greater predictive value for breast cancer risk compared to genetic variations. Lee et al. [39] conducted a study to identify specific germline single nucleotide polymorphisms (SNPs) that can effectively predict the occurrence of radiation-associated contralateral breast cancer (RCBC). The aim was not only to predict RCBC risk but also to gain new insights into the underlying carcinogenic process. To achieve this, the researchers employed a preconditioned RF regression method for forecasting the probability of developing RCBC. The model was evaluated using hold-out validation data, and it yielded an AUC of 0.62 ($p = 0.04$). This AUC value indicates the model's ability to discriminate between individuals who are at higher or lower risk of RCBC. The application of ML and bioinformatics techniques to genome-wide genotyping data demonstrated significant potential in uncovering plausible biological correlates associated with the risk of RCBC.

With the advancements in multi-omic data analysis, Rajpal et al. [40] attempted to uncover the molecular heterogeneity of breast cancer using Copy Number Variation (CNV) data, known for its stability as a genetic variation. However, existing algorithms often produce biomarkers that are too complex for clinical interpretation. To address this, the authors introduced XAI-CNVMarker, an explainable AI-based framework for discovering a small set of interpretable CNV biomarkers. Deep learning is employed for breast cancer classification, and different explainable AI methods are used to identify 44 CNV biomarkers. Through gene set analysis, the paper identifies subtype-specific enriched pathways, druggable genes, and prognostic outcome-related biomarkers. The framework efficacy is validated on METABRIC, showcasing the potential of explainable AI in discovering clinically relevant biomarkers. The study achieves a classification accuracy of 0.712 with a 95% confidence interval using 5-fold cross-validation.

In another study, Kumar and Das [41] aimed to identify diagnostic biomarkers for breast cancer using XAI on XG-Boost models trained on a binary classification dataset. It analyzed expression data of Peripheral blood mononuclear cell from 252 breast cancer patients and 194 healthy women. By incorporating SHAP values into the XG-Boost model, the authors discovered ten important genes associated with breast cancer development, which can be potential biomarkers. The findings indicated that SVIP, BEND3, MDGA2, LEF1-AS1, PRM1, TEX14, MZB1, TMIGD2, KIT, and FKBP7 genes significantly influence model prediction. They claimed that these genes have the potential to serve as early, non-invasive diagnostic and prognostic biomarkers for breast cancer patients.

Table 1 indexes various literature reviews for the detection of breast cancer using ML and various breast cancer datasets. As can be noticed, various researchers have reported high values for accuracy, although the AUC values as illustrated by [30], [34], and [39] may indicate poor fitting during the training process by the utilized ML algorithms. Based on the previous literature review, we can deduce that the authors focused on proposing ML models to detect breast cancer but without providing extensive interpretations for their models. Certain papers such as Mohi et al. [10] and Almutairi et al. [14] provide explanations for their ML models but they confine themselves to a single XAI technique for feature ranking, without delving into the specific impact of these features on ML classification. In contrast, this paper presents an explainable ML model specifically designed for the detection of breast cancer, addressing the gap in the existing research.

## III. DATASET DESCRIPTION

In this section, a detailed description of both the WBC and the WDBC datasets is presented. These datasets have been widely employed in various studies and research endeavors pertaining to classification and prediction tasks within the domain at hand. By thoroughly examining the features of these datasets, we hope to foster a deeper understanding and appreciation for their significance in facilitating accurate and reliable classification and prediction analyses.

### A. WISCONSIN BREAST CANCER DATASET

The WBC dataset is a well-known dataset in the field of ML and data analysis of health applications, which is widely used for classification and regression tasks. The dataset contains information about breast cancer tumors, including characteristics of the tumor, such as its size, shape, texture, and other features. The dataset was first introduced in 1992 by Dr. William H. Wolberg of the University of Wisconsin Hospitals. The WBC dataset contains 699 instances, or samples, of breast cancer tumor data, each of which has 10 features associated with it [42].

The first nine features describe various characteristics of the tumor, such as its radius, texture, smoothness, and symmetry, while the last feature is a binary label indicating whether the tumor is malignant or benign. Every feature is assessed using a scale ranging from 1 to 10, where a score of 1 indicates a closer proximity to benign characteristics, and a score of 10 indicates a closer proximity to malignant characteristics. For example, the clump Thickness feature measures the thickness of cell clusters in the breast tissue sample and is rated on a scale from 1 to 10, with 1 being the thinnest and 10 being the thickest [42].

The marginal adhesion feature evaluates how well the cells in the breast tissue sample adhere to one another [43], with 1 being the least sticky and 10 being the most. Single

**TABLE 1. List of the selected articles and the ML techniques used.**

| Ref. | ML algorithms | Evaluation metrics | Dataset |
|---|---|---|---|
| [9] | shallow ANN | 99.85% precision and 99.47% accuracy | WBC WDBC |
| [10] | LGBM | 99% accuracy | WDBC |
| [11] | Decision Tree | 92.26% accuracy | WBC |
| [12] | Averaged perceptron ML | 98.4% accuracy | WBC |
| [13] | Eagle Strategy (ESO) Optimization, Gravitational Search Optimization (GSO) | 98.9578% accuracy, 0.9705 sensitivity, 1.000 specificity 1.000 precision, 0.9696 F1-score | WDBC |
| [14] | Deep Reinforcement Learning | 98.90% accuracy for WBC, 99.02% for WDBC, and 98.88% for WPBC | WBC, WDBC, WPBC |
| [15] | RBF and MNN | 99.96% accuracy | WBC KDD |
| [17] | Deep neural network | 99.67% accuracy and 93% precision | WBC |
| [18] | NB, MLP, RF,and J48 | 97.2779% accuracy for NB | WDBC |
| [19] | DT, RFs, LR, KNN, SVM, NB, and rotation forests. | 98.1% accuracy for LR | WBC |
| [20] | SVM, DT, NB, KNN, and XG-boost | 98.24% accuracy for XG-boost | WBC |
| [21] | LR, XGboost, MLP, NB, RF, KNN, and DT | 98.3% for WBC dataset, 99.2% for WDBC dataset, and 78.6% of accuracy in the WPBC. | WBC WDBC WPBC |
| [22] | SVM,KNN, RF, ANNs, and LR | ANNs obtained the highest scores of 98.57% accuracy, 97.82% precision, and 0.9890 F1-score | WBC |
| [23] | Gradient booster, SVM, NB, LR, RF, KNN, and ANN | 99.73% for ANN | WDBC |
| [24] | SVM, NB, LR, RF, KNN, and DT | 96.5% accuracy for SVM and RF. | WBC |
| [25] | LRC, ID3, J48, and SVM. | 99.25% accuracy | WBC |
| [26] | ST-SVM with LPSVM, LSVM, SSVM, PSVM, and NSVM. | Accuracy of 97.1429%, a sensitivity of 98.2456%, a specificity of 95.082%, and ROC of 99.38% for LPSVM | WBC |
| [27] | WHAVE | 99.8% accuracy | WBC |
| [28] | LDA-SVM | WBC accuracy of 99.2%, recall of 98.0% and precision of 98.0% WPBC dataset accuracy of 79.5%, recall of 76.0% and precision of 59.0% | WBC WPBC |
| [30] | XG-Boost, RF, and Deep Neural Network | AUC 0.742 for XG-boost | Chinese women breast cancer dataset |
| [34] | RF, MLP, GBT, and GA | RF performed better (80% accuracy, 95% sensitivity, 80% specificity, and an AUC of 0.56). | Motamed Cancer Institute (ACECR) |
| [35] | Convolutional neural network (CNN) | 95.64% accuracy 96.84% F1-score. | INbreast dataset |
| [39] | RF regression | (AUC) of 0.62 ($p = 0.04$) | Germline (SNPs) |

Epithelial Cell Size is a measure for the size of each individual cell in the breast tissue sample and is measured and graded using a range of 1 to 10, with 1 denoting the smallest cell and 10 the largest. When a cell's nucleus is unenclosed by a cytoplasmic membrane, this condition is referred to as having "bare nuclei". This feature is zero when the membrane is absent, while a value of 10 indicates the highest degree of clumpiness. "Bland Chromatin" is a characteristic that assesses how chromatin, which is the component of chromosomes, appears in the breast tissue sample's cells. The mitoses feature, which counts the number of dividing cells seen in the breast tissue sample, is evaluated on a scale of 1 to 10, with 1 denoting a low number of mitoses and 10 denoting a high number. Table 2 summarizes the various parameters and details for the WCB dataset.

**TABLE 2. WBC description.**

| Parameters | Values |
|---|---|
| Number of instances | 699 instances |
| Number of features | 10 features |
| Number of inputs | 9 inputs (each input has from a scale 1: 10) |
| Number of outputs | two outputs (Malignant and Benign) |

**TABLE 3. WDBC description.**

| Parameters | Values |
|---|---|
| Number of instances | 569 instances |
| Number of features | 30 features |
| Number of inputs | 28 inputs (the inputs vary from 0 to 2500) |
| Number of outputs | two outputs (Malignant and Benign) |

This dataset has been widely used for research in the field of ML, particularly for binary classification tasks, as it provides a rich set of features for each sample, making it an excellent dataset for testing the performance of different classification algorithms. Many researchers have also used this dataset to develop and test feature selection and feature extraction techniques. Overall, the WBC dataset is a valuable resource for researchers and practitioners in the field of ML, as it provides a well-defined and well-documented dataset that can be used to test and compare various classification algorithms and techniques.

### B. WISCONSIN DIAGNOSTIC BREAST CANCER

WDBC is a public dataset that contains the medical records of breast cancer patients. The dataset was collected by Dr. William H. Wolberg of the University of Wisconsin Hospitals in the early 1990s, and it is widely used for research and development of ML algorithms. The WDBC dataset includes 569 observations, each of which contains 30 attributes. The first attribute is an ID number, which is unique to each patient. The second attribute is the diagnosis of breast cancer, which can be either malignant or benign. The other 28 attributes describe different characteristics of the tumor, such as its size, shape, and texture.

Table 3 outlines the various parameters and details for the WDCB dataset. The WDBC dataset is a valuable resource for

researchers and healthcare professionals who are interested in developing models that can accurately diagnose breast cancer. One of the most important aspects of this dataset is that it is highly accurate. The dataset has also been preprocessed to remove any redundant or irrelevant features, making it an ideal starting point for researchers who are interested in developing ML models for breast cancer diagnosis. One notable example is the work of Dr. David J. Hand of Imperial College London, who used the WDBC dataset to compare the performance of different ML algorithms for diagnosing breast cancer.

Dr. Hand found that a simple DT algorithm was the most effective for diagnosing breast cancer, achieving an accuracy of over 95%. The WDBC dataset has also been used to develop more advanced ML models, such as neural networks and SVM. These models can achieve even higher levels of accuracy, but they are also more complex and require more computational resources to train. In summary, the WDBC dataset is a highly accurate and well-documented dataset that is widely used for research and development of ML models for breast cancer diagnosis.

## IV. METHODOLOGY

This section outlines the approach and techniques employed in this study to develop an innovative and explainable machine learning-based model for breast cancer diagnosis. In order to successfully accomplish the objective of this research, the following series of steps are diligently executed:

1) The first step of this study involves gathering relevant datasets that comprise information about women diagnosed with breast cancer, such as the WBC and WDBC datasets. These datasets usually consist of different attributes or features associated with breast cancer, including tumor size, shape, texture, and location. Through the collection of these datasets,
2) After the data has been collected, the subsequent step involves the cleaning and preprocessing of the data. This encompasses various tasks aimed at ensuring the quality and suitability of the data for the ML model. Duplicate data points are eliminated to prevent any biases that may arise from redundant information. Outliers, which refer to extreme values that can adversely impact the model's performance, are identified and appropriately handled, either through removal or by applying statistical techniques to mitigate their influence. Additionally, the data is normalized or scaled to ensure that the features are on a consistent scale. This normalization process helps prevent any particular feature from dominating the model's training process due to its larger magnitude.
3) Following data cleaning and preprocessing, important features of the model are identified. These features play a crucial role in breast cancer diagnosis and can be determined through statistical analysis, domain knowledge, or using feature importance techniques provided by machine learning algorithms. By selecting the most relevant features, the model can focus on the most informative aspects of the data and improve its overall performance.
4) Once the relevant features are determined, an appropriate machine-learning algorithm is selected for the breast cancer diagnosis problem. Commonly used algorithms in medical diagnosis include SVMs, ANN, and RF. The selection of the algorithm depends on various factors such as the size of the dataset, the complexity of the problem, and the interpretability required.
5) The selected algorithm is then trained on the preprocessed dataset using a portion of the data for training (typically 75%) and another portion for testing (usually 25%) to assess the model's performance. This training process involves adjusting the algorithm's parameters and optimizing its performance on the given data.
6) After training, the model's performance is evaluated using the testing dataset. This evaluation measures the model's accuracy, sensitivity, specificity, and other relevant metrics to assess its diagnostic capabilities. By analyzing these metrics, researchers can gain insights into how well the model performs in correctly identifying breast cancer cases.
7) Finally, the model's output and predictions are analyzed and interpreted using XAI techniques. XAI helps researchers understand the factors and features that contribute to the model's predictions. It provides insights into the potential diagnosis of breast cancer by highlighting the most influential factors and providing a transparent explanation for the model's decision-making process.

The block diagram in Fig. 1 illustrates the entire process of building the machine learning model for breast cancer diagnosis, encompassing data collection, cleaning and preprocessing, feature selection and engineering, algorithm selection and training, evaluation, and interpretation using XAI techniques. This comprehensive approach aims to develop an accurate and interpretable model that can aid in the diagnosis and understanding of breast cancer.

## V. RESULTS AND DISCUSSION

In this section, the results of using ML to classify breast cancer based on two datasets (WBC & WDBC) are presented and discussed.

### A. WBC DATASET RESULTS

#### 1) DESCRIPTIVE ANALYSIS

Firstly, descriptive analysis was performed to represent and describe the data. Fig. 2 presents the correlation matrix of the dataset. It is concluded from the correlation matrix that the uniformity of cell size and uniformity of cell shape are highly correlated. In addition to that, these features also have a high correlation with the output class as well as the bare nuclei feature. To understand the distribution of the data,

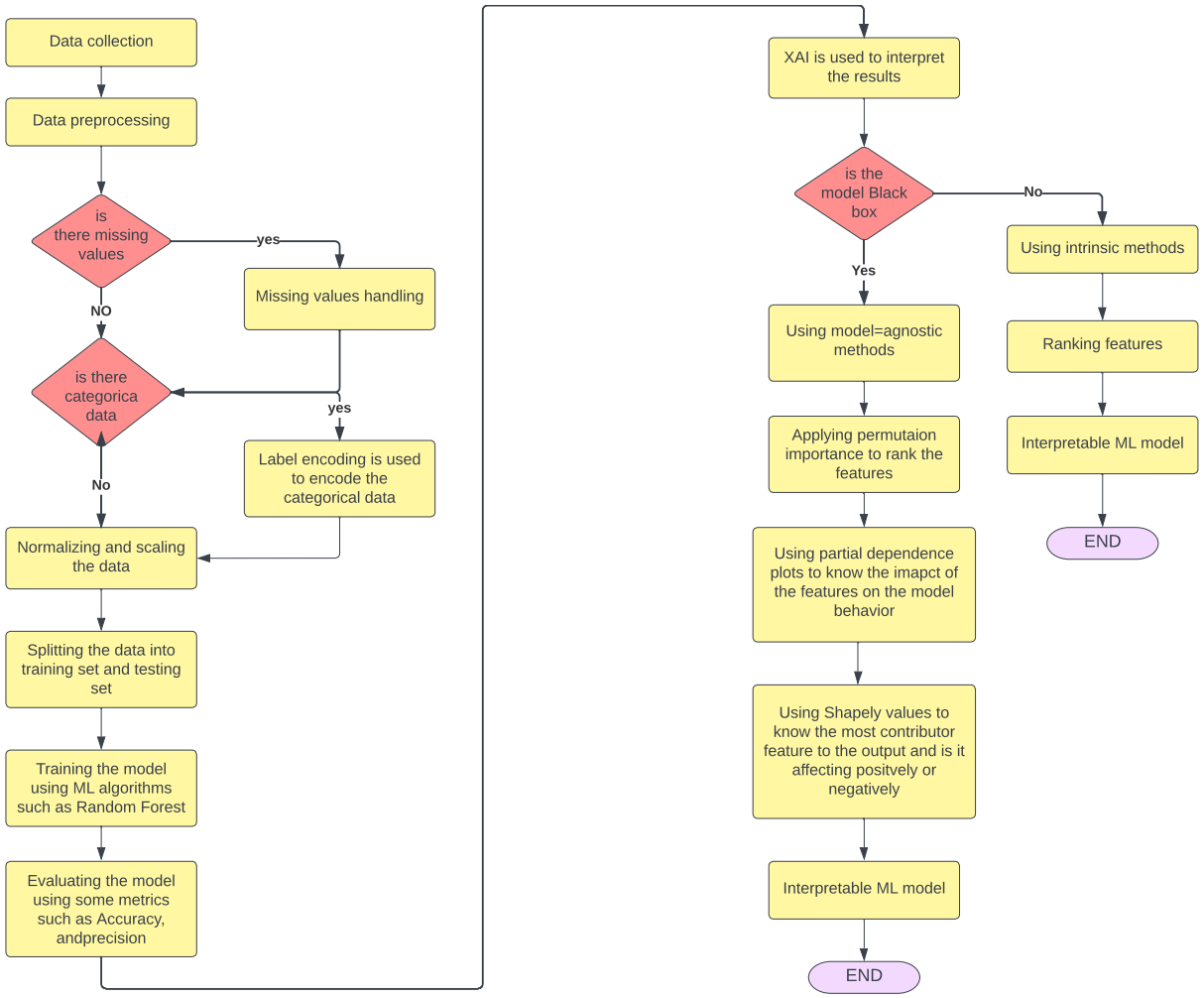


**FIGURE 1.** The complete process of building our ML model.

we used box plots to compare the distribution of multiple features across different categories. Box plots provide a visual summary of the distribution of a dataset, including the spread, center, and any outliers. Fig. 3 illustrates the box plots for all the features across the output class, which is malignant and benign. Fig. 4 depicts the histogram of the Malignant and Benign classes.

### 2) ML MODEL RESULTS

The study involved the use of different ML algorithms such as SVM, KNN, RF, and XG-boost. We trained our ML model using these algorithms, and XG-boost achieved the highest accuracy. Fig. 5 shows the accuracy and precision of each algorithm. The precision, which is defined as the number of correct instances retrieved divided by all retrieved instances [44], is considered very important in the classification of any disease. Hence, we focused on the precision of the ML model which ensures accurate identification of positive disease cases and minimizes false positives for enhanced diagnostic reliability.

Fig. 6 shows the precision percentage of each algorithm for each class. It is clear that KNN is performing better in classifying the malignant class than XG-boost. Due to the high correlation between the uniformity of cell size and uniformity of cell shape, we can use only one of them as a trial to improve the model performance. Hence, the uniformity of cell size was removed. The results indicated that there is an improvement in the performance in the case of KNN, RF, and SVM; meanwhile, XG-boost yielded the same result, as shown in Fig. 7. The model achieved 97.7% using KNN. As KNN is our best-performing model, its computational complexity relies on the dataset size, feature count, and the chosen value of k which are 699, 8, and 5, respectively. The time complexity for a single query point in the KNN algorithm is

$$O(k * log(n)), \tag{1}$$

where $n$ represents the testing examples and $k$ denotes the number of neighbors. Table 4 shows the time complexity for each algorithm to train and test the ML model empowered by

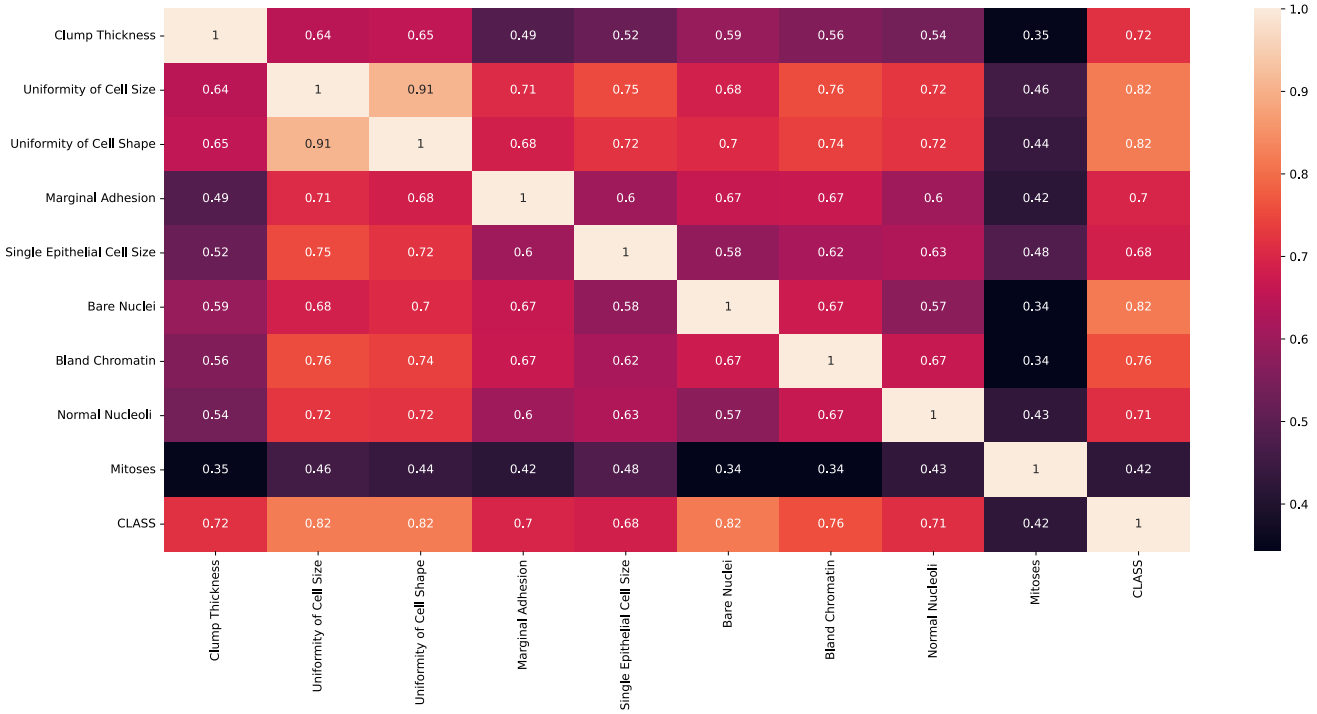


**FIGURE 2.** The correlation matrix for WBC dataset.

**TABLE 4.** Time for the training and testing process for each algorithm in WBC dataset.

| Algorithm | Elapsed time | Time per instance |
|---|---|---|
| XG-boost | 52 ms | 47 us |
| SVM | 7 ms | 7 us |
| RF | 40 ms | 36 us |
| KNN | 8 ms | 8 us |

Intel(R) processor, 16 GB RAM, and Core(TM) i7-10750H CPU @ 2.60GHz. It should be observed that KNN provides a speedup of up to 80.00% and 84.00%, respectively, when compared to RF and XG-boost.

### 3) XAI RESULTS

After the ML model's performance has been assessed, it is critical to explain and analyze the findings in order to comprehend the model's performance. This entails determining which features are crucial for the model's predictions, understanding the connections between the features and the goal variable, and identifying any relevant patterns or trends in the data. This study utilized three model-agnostic techniques including permutation importance, PDP, and SHAP.

#### *a: PERMUTATION IMPORTANCE RESULTS*

The permutation importance method is used to rank the features; hence, the most important features are identified. After permuting the feature, we ascertain its relevance by calculating the rise in the model's prediction error. A feature is deemed "essential" if altering its values causes a rise in model error. A feature is considered "unimportant" if altering its values causes the same model error because the feature was disregarded for the forecast. Fig. 8 shows the results from the permutation of features, indicating that Bare nuclei and clump thickness are the most important features. The y-axis of the permutation importance plot represents the feature importance scores. These scores are calculated as the decrease in a model's score when a particular feature is randomly permuted. Further interpretations are needed to prove the permutation's result. Hence, PDPs and Shap values are performed in the following sections.

#### *b: PARTIAL DEPENDENCE PLOTS RESULTS*

PDP is a global ML interpretation method. This approach considers all instances and provides an assessment of the overall association between a feature and the predicted outcome. Fig. 9 illustrates the PDP for the Bare Nuclei feature. A value of 1 represents the absence of Bare Nuclei, while a value of 10 indicates the highest degree of clumpiness. It is evident that the model's prediction for the malignant class increases as the range of bare nuclei increases, particularly from 8 to 10, where the blue shaded area represents the average behavior of the model. Interactive plots can offer more helpful explanations. For example, clump thickness and marginal adhesion are used to evaluate the overall architecture of a tissue sample. We can explore the relationship

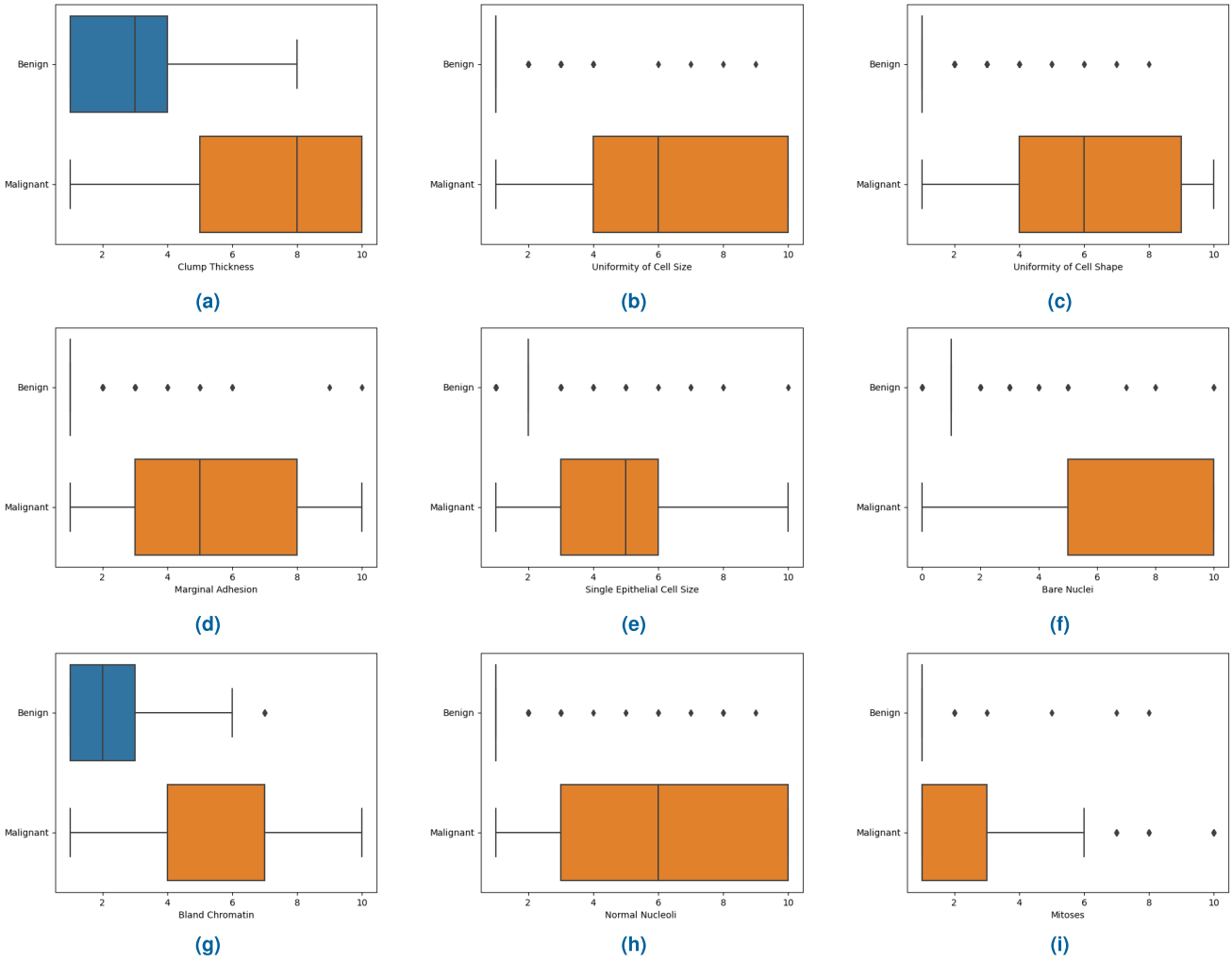


**FIGURE 3.** Box plots of (a) clump thickness feature, (b) uniformity of cell size feature, (c) uniformity of cell shape feature, (d) marginal adhesion feature (e) marginal adhesion feature, (e) single epithelial cell size feature, (f) bare Nuclei feature, (g) Bland Chromatin feature, (h) normal Nucleoli feature, (i) mitosis feature.

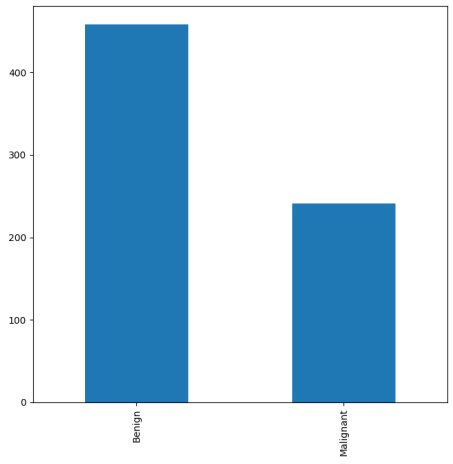


**FIGURE 4.** The histogram of the target class.

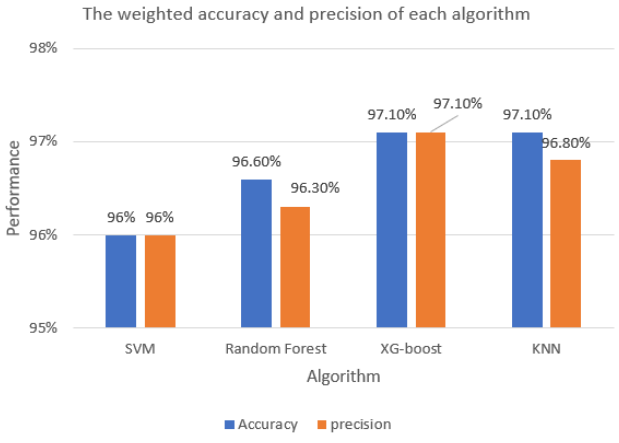


**FIGURE 5.** The weighted accuracy and precision of each algorithm.

between changes in the degree of adhesion between individual cells at the margins of these clusters and changes in the thickness of cellular clusters. This can help us understand the potential connection between breast cancer development and alterations in tissue architecture. Fig. 10 displays the

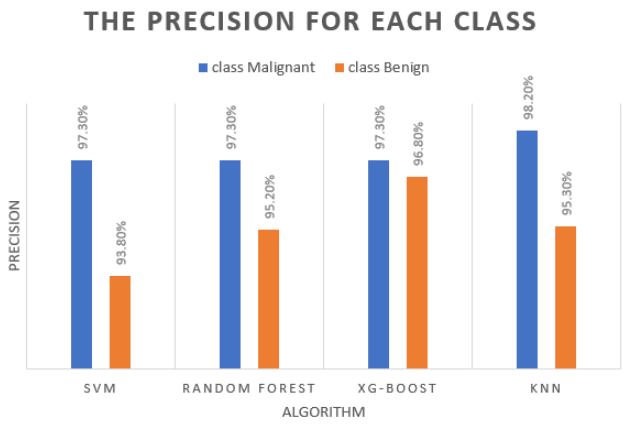


**FIGURE 6.** The precision percentage for Malignant and precision class.

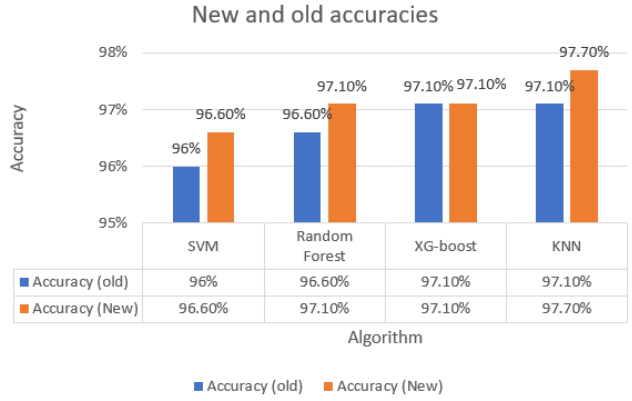


**FIGURE 7.** The weighted accuracy before and after removing Uniformity of cell Shape feature.

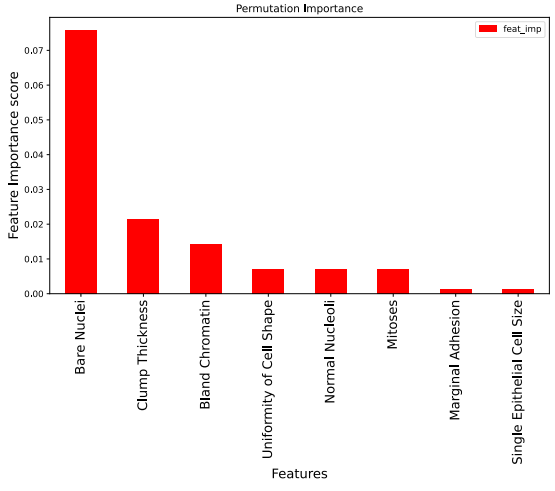


**FIGURE 8.** The permutation importance plot for the WBC dataset.

interaction plot for clump thickness and marginal adhesion features. The plot shows that when the thickness varies from one to three and the marginal adhesion is 10 (sticky cells), the model's probability of predicting the malignant class increases.

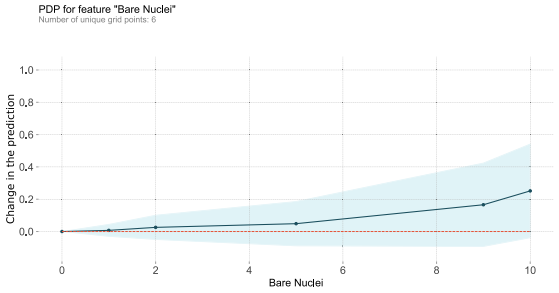


**FIGURE 9.** PDP plot for Bare Nuclei.

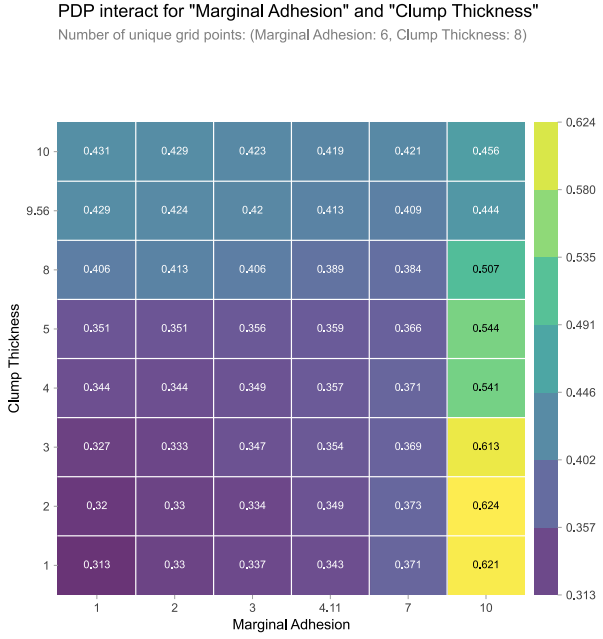


**FIGURE 10.** Interactive PDP plot for clump thickness and marginal adhesion.

### c: SHAP RESULTS

SHAP method [6] depends on Shapley values that provide explanations of specific instances instead of global explanations. We can determine which feature is more important for a given prediction using Shapley values. When we need an answer for a particular prediction and are less concerned with knowing the model's "typical" behavior, SHAP can be useful. SHAP [7] is used to explain the prediction of an instance x by calculating the contribution of each feature to the prediction. A SHAP summary plot was created to assess the contribution of each feature to the classification of breast cancer. Fig. 11 demonstrates that the Bare nuclei feature has the highest contribution, which aligns with the result of the permutation.

## B. WDBC DATASET RESULTS

In this dataset, the features were extracted from digitized images of breast tissue samples. For each cell nucleus in the tissue sample, ten real-valued characteristics are computed. These characteristics consist of the mean, standard deviation,

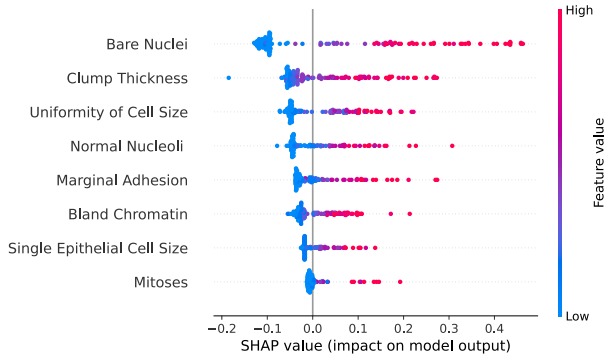


**FIGURE 11.** SHAP summary plot.

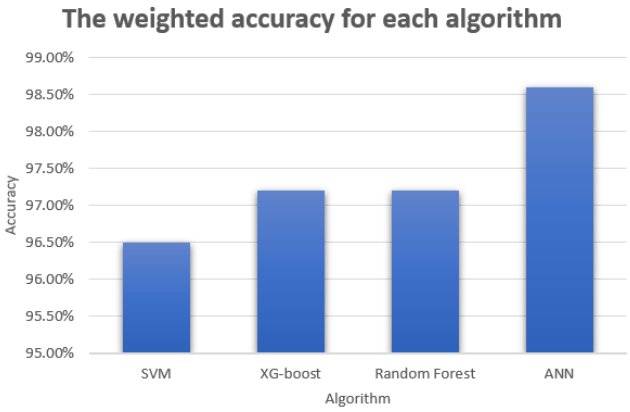


**FIGURE 12.** The weighted accuracy for each algorithm.

and worst (largest) numbers. In this paper, we used only the extreme (worst) values for each feature to train our ML model which are the extreme values for each feature.

There are several benefits to training an ML model using only the ''worst'' characteristics from the WDBC dataset. First, dimensionality reduction is achieved by the selection of 10 out of 30 features. Second, the ML model provides a higher predictive power. compared to the other features in the dataset, the ''worst'' features have a stronger correlation with the prevalence of malignancy. A model may be more accurate in identifying whether a tissue sample is benign or malignant by concentrating only on these characteristics. Finally, improving explainability because fewer features in a model can make it simpler to analyze and comprehend the variables that affect the model's predictions. The best-performing ML model achieved an accuracy of 99% and a precision of 94.4% using ANN. The performance of each model is shown in Fig. 12. The computational complexity for training our best-performing model, a neural network with 3 layers comprising $i, j$, and $k$ nodes (15, 10, 1, respectively), using $t$ training examples equals 426, and $n$ epochs set to 100 can be calculated as follows:

$$O(nt(ij + jk)). \tag{2}$$

**TABLE 5.** Time for the training and testing process for each algorithm in WDBC.

| Algorithm | Elapsed time | Time per instance |
|---|---|---|
| XG-boost | 70 ms | 48 us |
| SVM | 162 ms | 3.6 us |
| RF | 102 ms | 71 us |
| ANN | 4.6 s | 3 ms |

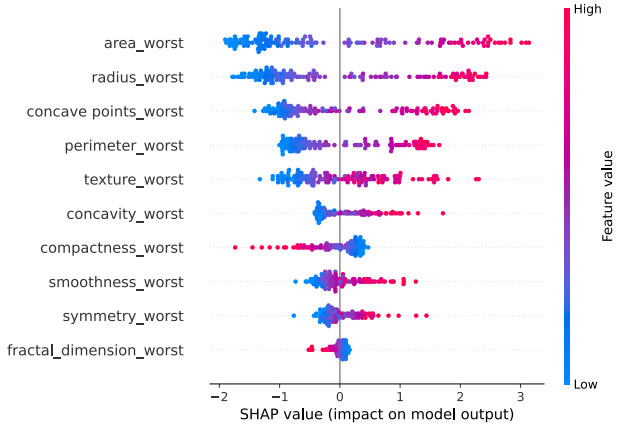


**FIGURE 13.** SHAP summary plot for malignant class WDBC.

Table 5 shows the time complexity for each algorithm to train and test the dataset. Despite that ANN has the best performance, it entails the longest time complexity.

To accomplish the entire process of constructing an interpretable ML model, the Shapely values were calculated for the features that represent only the extreme values in order to understand the effect of these features on the prediction of the breast cancer classes. In terms of interpretability, the ''area worst'' feature, which is The total area occupied by the nucleus, is the most contributing feature in the classification of breast cancer. The SHAP plot in Fig. 13 illustrates that when the area has a large value, it positively affects the classification task. This means the larger the value of the area feature, the higher the model's prediction of malignant breast cancer. Different from the existing literature reviews, we present a systematic framework of an explainable ML model. Initially, the permutation importance method is employed to assess the relative importance of input features, allowing us to identify the most crucial one. Subsequently, a Partial Dependence Plot (PDP) is developed to gain insight into the relationship the feature holds with the output. Then SHAP method is employed to quantify the contribution of each feature, either on a local or global scale, in the classification task.

In the literature review, researchers employ a single XAI method to explore the influential features present within the dataset. However, our extensive research and results indicated that incorporating a diverse range of XAI techniques can gain a more comprehensive understanding of the underlying factors that shape the dataset and its predictive outcomes. Table 6

**TABLE 6. Comparison of the proposed work with some state-of-the-art works.**

| Ref. | Algorithm | ML model Accuracy | XAI technique | Dataset |
|---|---|---|---|---|
| [10] | Gradient Boosting | 99% | SHAP | WDBC |
| [14] | Deep Reinforcement Learning | 98.90% for WBC, 99.02% for WDBC, 98.88% for WPBC. | LIME | WBC, WDBC, and WPBC |
| Proposed work | KNN, SVM, XG-boost, RF, and ANN | 97.9 % for WBC, 98.6 % for WDBC | Permutation importance, Partial dependence plots, SHAP | WBC, WDBC |

shows the benchmark of our proposed model with some recent works utilizing XAI to classify breast cancer. It should be noted that although the authors claimed high accuracy for the ML model provided with interpretations, they limit their approach to only one XAI technique to rank the features without providing any further details about how the features affect the ML classification. In addition to that, our work attempts to link two breast cancer datasets in order to uncover potential correlations, patterns, and insights that may arise from the analysis of these complementary datasets. Through this synergistic approach, we can enhance our understanding of breast cancer and potentially unveil novel findings that may have remained undiscovered by studying each dataset in isolation.

We have emphasized that ''bare nuclei'' in WBC and the ''area worst'' in WDBC are the most contributing feature to the classification of breast cancer malignancy. However, there is an indirect relationship between both features. Specifically, the ''Worst Area'' attribute in the WDBC dataset could be impacted by abnormal cellular proliferation or division, which is observable in the ''Bare Nuclei'' feature. Consequently, this abnormality may lead to enlarged and irregularly shaped cell nuclei. However, XAI techniques have their limitations. Multiple iterations of permutation importance, while providing valuable insights into feature importance, can significantly increase the runtime of the analysis. Additionally, the maximum number of dependent variables that can be plotted simultaneously is limited to two, which can hinder comprehensive visualizations of complex models. PDPs assume that the variables displayed in the plot are not correlated with other variables used in the model, which may not always hold true and can impact the accuracy of interpretations. Furthermore, the use of KernelSHAP, a global SHAP method, can be slow due to the computation of Shapley values for numerous instances, affecting its practicality in large-scale applications. These limitations highlight the need for further research and development in XAI to overcome these challenges and enhance the interpretability and usability of AI models.

## VI. CONCLUSION

Advancements in data science and technology have propelled the interest in developing intelligent systems for early breast cancer detection. This work proposes a framework for breast cancer detection including data collection, data preprocessing, model selection, model evaluation, and finally, model interpretation. Our best-performing model achieves an accuracy of 97.7% and a precision of 98.2%, employing KNN for the WBC dataset. In the WDBC dataset, ANN obtains the best performance by achieving an accuracy of 98.6% and a precision of 94.4%. XAI techniques provide some explanations for the model results, such as permutation importance methods, PDP, and SHAP methods. The permutation method indicates that the Bare Nuclei feature is the most important feature. Besides, PDP aids in our comprehension of the possible link between changes in tissue architecture and the growth of breast cancer by finding the relationship between the thickness of the tissue and the stickiness of the cells. Finally, Shapely values illustrate that Bare nuclei are the most contributing feature to malignant breast cancer detection. The higher values of Bare Nuclei, which means the absence of the cell membrane, the higher the probability of the model predicting the malignant class. We find out that ''bare nuclei'' in WBC and the ''area worst'' in WDBC are the most contributing feature to the classification of breast cancer malignancy. However, maybe there is an indirect relationship between both features. For instance, The ''Worst Area'' feature in the WDBC dataset may be influenced by abnormal cell growth or division, which can be seen in the ''Bare Nuclei'' feature in the dataset. It may also result in bigger and more atypically shaped cell nuclei. However, more investigation and analysis would be required to look into any possible connections between these characteristics or the biological processes they are thought to represent. Future work involves the use of genetic data for the early prediction of breast cancer. Subsequent research directions involve incorporating genetic data for early breast cancer prediction and performing ensemble analyses on the WDBC datasets to uncover additional properties and insights for early detection.

## REFERENCES


[1] (2022). *World Health Organization Breast Cancer*. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/breast-cancer

[2] H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, and F. Bray, ''Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries,'' *CA, A Cancer J. Clinicians*, vol. 71, no. 3, pp. 209–249, May 2021.

[3] M. Arnold, E. Morgan, H. Rumgay, A. Mafra, D. Singh, M. Laversanne, J. Vignat, J. R. Gralow, F. Cardoso, S. Siesling, and I. Soerjomataram, ''Current and future burden of breast cancer: Global statistics for 2020 and 2040,'' *Breast*, vol. 66, pp. 15–23, Dec. 2022.

[4] O. I. Obaid, M. A. Mohammed, M. K. A. Ghani, A. Mostafa, and F. Taha, ''Evaluating the performance of machine learning techniques in the classification of Wisconsin breast cancer,'' *Int. J. Eng. Technol.*, vol. 7, no. 4.36, pp. 160–166, 2018.

[5] A. Thampi, *Interpretable AI: Building Explainable Machine Learning Systems*. New York, NY, USA: Simon and Schuster, 2022.

[6] L. Gianfagna and A. D. Cecco, *Explainable AI With Python*. Cham, Switzerland: Springer, 2021.

[7] C. Molnar. *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*. Accessed: May 2, 2023. [Online]. Available: https://christophm.github.io/interpretable-ml-book

[8] S. Rathi, ''Generating counterfactual and contrastive explanations using SHAP,'' 2019, *arXiv:1906.09293*.

[9] M. H. Alshayeji, H. Ellethy, S. Abed, and R. Gupta, "Computer-aided detection of breast cancer on the Wisconsin dataset: An artificial neural networks approach," *Biomed. Signal Process. Control*, vol. 71, Jan. 2022, Art. no. 103141.
[10] K. M. M. Uddin, N. Biswas, S. T. Rikta, S. K. Dey, and A. Qazi, "XML-LightGBMDroid: A self-driven interactive mobile application utilizing explainable machine learning for breast cancer diagnosis," *Eng. Rep.*, vol. 5, no. 11, 2023, Art. no. e12666.
[11] J. O. Afolayan, M. O. Adebiyi, M. O. Arowolo, C. Chakraborty, and A. A. Adebiyi, "Breast cancer detection using particle swarm optimization and decision tree machine learning technique," in *Intelligent Healthcare: Infrastructure, Algorithms and Management*. Cham, Switzerland: Springer, 2022, pp. 61–83.
[12] V. Birchha and B. Nigam, "Performance analysis of averaged perceptron machine learning classifier for breast cancer detection," *Proc. Comput. Sci.*, vol. 218, pp. 2181–2190, 2023.
[13] L. K. Singh, M. Khanna, and R. Singh, "Artificial intelligence based medical decision support system for early and accurate breast cancer prediction," *Adv. Eng. Softw.*, vol. 175, Jan. 2023, Art. no. 103338.
[14] S. Almutairi, S. Manimurugan, B.-G. Kim, M. M. Aborokbah, and C. Narmatha, "Breast cancer classification using deep Q learning (DQL) and gorilla troops optimization (GTO)," *Appl. Soft Comput.*, vol. 142, Jul. 2023, Art. no. 110292.
[15] M. Tahmooresi, A. Afshar, B. Bashari Rad, K. B. Nowshath, and M. A. Bamiah, "Early detection of breast cancer using machine learning techniques," *J. Telecommun. Electron. Comput. Eng.*, vol. 10, nos. 2–3, pp. 21–27, 2018.
[16] Doreswamy and M. U. Salma, "Fast modular artificial neural network for the classification of breast cancer data," in *Proc. 3rd Int. Symp. Women Comput. Informat.*, Aug. 2015, pp. 66–72.
[17] N. Khuriwal and N. Mishra, "Breast cancer diagnosis using deep learning algorithm," in *Proc. Int. Conf. Adv. Comput., Commun. Control Netw. (ICACCCN)*, Oct. 2018, pp. 98–103.
[18] M. T. Ahmed, M. N. Imtiaz, and A. Karmakar, "Analysis of Wisconsin breast cancer original dataset using data mining and machine learning algorithms for breast cancer prediction," *J. Sci. Technol. Environ. Informat.*, vol. 9, no. 2, pp. 665–672, 2020.
[19] M. F. Ak, "A comparative analysis of breast cancer detection and diagnosis using data visualization and machine learning applications," *Healthcare*, vol. 8, no. 2, p. 111, Apr. 2020.
[20] M. M. Rahman, Z. Ferdousi, P. Saha, and R. A. Mayuri, "A machine learning approach to predict breast cancer using boosting classifiers," *Indian J. Comput. Sci. Eng.*, vol. 14, no. 3, pp. 409–415, Jun. 2023.
[21] M. Zeid, D. El-Bahnasy, and S. Abu-Youssef, "An efficient optimized framework for analyzing the performance of breast cancer using machine learning algorithms," *J. Theor. Appl. Inf. Technol.*, vol. 100, no. 14, pp. 1–14, 2022.
[22] M. M. Islam, M. R. Haque, H. Iqbal, M. M. Hasan, M. Hasan, and M. N. Kabir, "Breast cancer prediction: A comparative study using machine learning techniques," *Social Netw. Comput. Sci.*, vol. 1, no. 5, pp. 1–14, Sep. 2020.
[23] K. Mridha, "Early prediction of breast cancer by using artificial neural network and machine learning techniques," in *Proc. 10th IEEE Int. Conf. Commun. Syst. Netw. Technol. (CSNT)*, Jun. 2021, pp. 582–587.
[24] S. Ara, A. Das, and A. Dey, "Malignant and benign breast cancer classification using machine learning algorithms," in *Proc. Int. Conf. Artif. Intell. (ICAI)*, Apr. 2021, pp. 97–101.
[25] S. G. Durai, S. H. Ganesh, and A. J. Christy, "Novel linear regressive classifier for the diagnosis of breast cancer," in *Proc. World Congr. Comput. Commun. Technol. (WCCCT)*, Feb. 2017, pp. 136–139.
[26] A. T. Azar and S. A. El-Said, "Performance analysis of support vector machines classifiers in breast cancer mammography recognition," *Neural Comput. Appl.*, vol. 24, no. 5, pp. 1163–1177, Apr. 2014.
[27] C. Deng and M. Perkowski, "A novel weighted hierarchical adaptive voting ensemble machine learning method for breast cancer detection," in *Proc. IEEE Int. Symp. Multiple-Valued Log.*, May 2015, pp. 115–120.
[28] O. J. Egwom, M. Hassan, J. J. Tanimu, M. Hamada, and O. M. Ogar, "An LDA–SVM machine learning model for breast cancer classification," *BioMedInformatics*, vol. 2, no. 3, pp. 345–358, Jun. 2022.
[29] P. Manikandan, U. Durga, and C. Ponnuraja, "An integrative machine learning framework for classifying seer breast cancer," *Sci. Rep.*, vol. 13, no. 1, pp. 1–12, 2023.
[30] C. Hou, X. Zhong, P. He, B. Xu, S. Diao, F. Yi, H. Zheng, and J. Li, "Predicting breast cancer in Chinese women using machine learning techniques: Algorithm development," *JMIR Med. Informat.*, vol. 8, no. 6, Jun. 2020, Art. no. e17364.
[31] A. Sahni, S. Singh, and G. Srivastava, "Exploring data science for highlighting breast cancer prediction using Python," in *Proc. Int. Conf. Innov. Comput. Commun. (ICICC)*, 2021, p. 4.
[32] H. E. Massari, N. Gherabi, S. Mhammedi, Z. Sabouri, H. Ghandi, and F. Qanouni, "Effectiveness of applying machine learning techniques and ontologies in breast cancer detection," *Proc. Comput. Sci.*, vol. 218, pp. 2392–2400, 2023.
[33] T. S. Rana, I. Saleem, R. N. Rao, M. Shabbir, and L. W. Chaudhry, "Comparative analysis of breast cancer detection using cutting-edge machine learning algorithms (MLAs)," *Innov. Comput. Rev.*, vol. 3, no. 1, pp. 1–15, Jun. 2023.
[34] R. Rabiei, S. M. Ayyoubzadeh, S. Sohrabei, M. Esmaeili, and A. Atashi, "Prediction of breast cancer using machine learning approaches," *J. Biomed. Phys. Eng.*, vol. 12, no. 3, p. 297, Jul. 2022.
[35] M. A. Al-antari, M. A. Al-masni, M.-T. Choi, S.-M. Han, and T.-S. Kim, "A fully integrated computer-aided diagnosis system for digital X-ray mammograms via deep learning detection, segmentation, and classification," *Int. J. Med. Informat.*, vol. 117, pp. 44–54, Sep. 2018.
[36] R. Massafra, "Analyzing breast cancer invasive disease event classification through explainable artificial intelligence," *Frontiers Med.*, vol. 10, Feb. 2023, Art. no. 1116354.
[37] I. Maouche, L. S. Terrissa, K. Benmohammed, and N. Zerhouni, "An explainable AI approach for breast cancer metastasis prediction based on clinicopathological data," *IEEE Trans. Biomed. Eng.*, early access, Jun. 5, 2023, doi: 10.1109/TBME.2023.3282840.
[38] F. Silva-Aravena, H. Núñez Delafuente, J. H. Gutiérrez-Bahamondes, and J. Morales, "A hybrid algorithm of ML and XAI to prevent breast cancer: A strategy to support decision making," *Cancers*, vol. 15, no. 9, p. 2443, Apr. 2023.
[39] S. Lee, X. Liang, M. Woods, A. S. Reiner, P. Concannon, L. Bernstein, C. F. Lynch, J. D. Boice, J. O. Deasy, J. L. Bernstein, and J. H. Oh, "Machine learning on genome-wide association studies to predict the risk of radiation-associated contralateral breast cancer in the WECARE study," *PLoS ONE*, vol. 15, no. 2, Feb. 2020, Art. no. e0226157.
[40] S. Rajpal, A. Rajpal, M. Agarwal, V. Kumar, A. Abraham, D. Khanna, and N. Kumar, "XAI-CNVMarker: Explainable AI-based copy number variant biomarker discovery for breast cancer subtypes," *Biomed. Signal Process. Control*, vol. 84, Jul. 2023, Art. no. 104979.
[41] S. Kumar and A. Das, "Peripheral blood mononuclear cell derived biomarker detection using eXplainable artificial intelligence (XAI) provides better diagnosis of breast cancer," *Comput. Biol. Chem.*, vol. 104, Jun. 2023, Art. no. 107867.
[42] W. H. Wolberg and O. L. Mangasarian, "Multisurface method of pattern separation for medical diagnosis applied to breast cytology," *Proc. Nat. Acad. Sci. USA*, vol. 87, no. 23, pp. 9193–9196, Dec. 1990.
[43] G. I. Salama, M. Abdelhalim, and M. A.-E. Zeid, "Breast cancer diagnosis on three different datasets using multi-classifiers," *Breast Cancer (WDBC)*, vol. 32, no. 569, p. 2, 2012.
[44] H. Dalianis, *Clinical Text Mining: Secondary Use of Electronic Patient Records*. Berlin, Germany: Springer Nature, 2018.

**TAREK KHATER** was born in TX, USA. He received the B.S. degree in biomedical engineering from Minia University, Egypt, in 2021. He is currently pursuing the M.Sc. degree in biomedical engineering with the University of Sharjah. His M.Sc. dissertation was titled, "Explainable Artificial Intelligence in the Classification of Breast Cancer in Young Women." He is currently a Research and Teaching Assistant with the University of Sharjah. He has a number of publications in the research area of machine learning, explainable artificial intelligence, and their applications in biomedical engineering.

**ABIR HUSSAIN** (Senior Member, IEEE) received the Ph.D. degree from The University of Manchester (UMIST), U.K., in 2000. Her Ph.D. dissertation was titled, ''Polynomial Neural Networks for Image and Signal Processing.'' She is currently a Visiting Professor in machine learning with Liverpool John Moores University, U.K. She has published numerous refereed research papers in conferences and journals in the research areas of neural networks, signal prediction, telecommunication fraud detection, and image compression. She has worked with higher-order and recurrent neural networks and their applications to e-health and medical image compression techniques. She has developed with her research students several recurrent neural network architectures. She is currently a Ph.D. supervisor and an external examiner for research degrees, including M.Phil. and Ph.D. degrees. She is one of the initiators and chairs of the Development in E-Systems Engineering (DeSE) series.

**RIYAD BENDARDAF** received the bachelor's degree in medicine from Arab Medical University, Benghazi, Libya, the master's degree in Finland, and the Ph.D. degree from the Oncology Department and Radiation Therapy, Turku University Hospital, University of Turku, Turku, Finland, with a focus on predictive and prognostic signs of metastatic rectal and colon cancer. He has been a medical oncologist in United Arab Emirates, since 2013. Previous clinical placements include a Consultant of medical oncology with the Medical Oncology Unit, Department of Medicine, Benghazi Medical Centre, Benghazi, and a Medical Oncology Specialist with the Department of Oncology and Radiotherapy, Turku University Hospital, where he initially started as a Oncology Resident. He is a fellow of the Royal College of Physicians Edinburgh, U.K.

**IMAN M. TALAAT** received the M.B.B.Ch. and M.D. degrees, and the M.Sc. and Ph.D. degrees in basic medical sciences in pathology from the Faculty of Medicine, Alexandria University, Egypt. She is currently a Full Professor in pathology with the Clinical Sciences Department, College of Medicine, University of Sharjah, Sharjah, United Arab Emirates, and a Full Professor in pathology with the Pathology Department, Faculty of Medicine, Alexandria University. She is also a consultant anatomic pathologist. She is known for her research work in the field of cancer, focusing mainly on breast, colorectal, prostate, and renal cancers, and other solid tumors. In addition, she has a good experience in immunohistochemistry (IHC), in-situ hybridization (ISH), and fluorescence in situ hybridization (FISH) techniques. She has more than 70 publications in high-impact peer-reviewed journals and she is the PI and co-PI of various internal and external grants. Additionally, she is the Coordinator of the Immuno-Oncology Research Group, Research Institute of Medical and Health Sciences (RIMHS), University of Sharjah.

**HISSAM TAWFIK** (Senior Member, IEEE) received the Ph.D. degree in computer engineering from The University of Manchester, U.K. He is currently a Professor in artificial intelligence. He has a long research track record of refereed publications in reputable international journals and conference proceedings in applied artificial intelligence, biologically inspired systems, big data, and machine learning. He is also a Professor with the College of Engineering, University of Sharjah, United Arab Emirates. Prior to that, he worked for over 20 years for various universities in the U.K. He served as an Editor for the *International Journal of Future Generation Computer Systems* (Elsevier), the *International Journal of Neural Computing and Applications* (Springer), and *International Journal of Reliable Intelligent Environments* (Springer). He is the Chair of the International Conference Series on Developments in eSystems Engineering (DESE) and a Guest Editor of the journal Special Issue on ''Robotics, Sensors and Industry 4.0'' for *Sensors* (MPDI).

**SAM ANSARI** received the B.Sc. degree in telecommunication engineering from Canadian University Dubai, Dubai, United Arab Emirates, and the M.Sc. degree in electrical and computer engineering from Abu Dhabi University, Abu Dhabi, United Arab Emirates. He is currently pursuing the Ph.D. degree with the Department of Electrical and Computer Engineering, University of Sharjah, Sharjah, United Arab Emirates. His research interests include wireless communications, molecular communications, artificial intelligence, machine learning, and renewable energy.

**SOLIMAN MAHMOUD** (Senior Member, IEEE) was born in Cairo, Egypt, in 1971. He received the B.Sc. (Hons.), M.Sc., and Ph.D. degrees from the Electronics and Communications Department, Cairo University, Egypt, in 1994, 1996, and 1999, respectively. He is currently a Professor with the Electrical Engineering Department, University of Sharjah, Sharjah, United Arab Emirates. He is a distinguished academic leader in the field of electrical engineering, will be on leave from the University of Sharjah and will assume the position of Vice Chancellor for Academic Affairs with the University of Khorfakkan, Khorfakkan, United Arab Emirates, in September 2023. He supervised three Ph.D. students, 15 M.Sc. students, and more than 50 senior design projects. He is actively engaged in scholarly research work and has authored or coauthored more than 170 journals and conference publications since joining academia, in 1996. He received ''The German-Egyptian Research Fund'' Grant. The Grant has been used to finance the project ''Design of CMOS Field Programmable Analog Array and Its Applications.'' The project has been carried out in collaboration with the Ulm Microelectronics Institute, Ulm University, Germany. His articles received more than 2100 citations, and his Google Scholar H-index is 25 (His H-index from Scopus is 22). He published six refereed research books. His research interests include mixed analog/digital integrated electronic circuit (IC) design, including mixed mode (voltage/current) analog circuits IC design, mixed (analog/digital) programmable CMOS electronics systems, biomedical circuits, field programmable analog arrays (FPAAs), and multi-standard wireless receiver design. In 2005, he received the Science Prize in Advanced Engineering Technology from the Academy of Scientific Research and Technology, Higher Ministry of Education, Cairo. He received the Distinguished Research Award from the University of Sharjah, from 2011 to 2012 and from 2014 to 2015.