



# A novel approach for breast cancer detection using optimized ensemble learning framework and XAI

Raafat M. Munshi<sup>a</sup>, Lucia Cascone<sup>b</sup>, Nazik Alturki<sup>c</sup>, Oumaima Saidani<sup>c</sup>, Amal Alshardan<sup>c</sup>, Muhammad Umer<sup>d,\*</sup>

<sup>a</sup> Department of Medical Laboratory Technology (MLT), Faculty of Applied Medical Sciences, King Abdulaziz University, Rabigh, Saudi Arabia

<sup>b</sup> Department of Computer Science, University of Salerno, Fisciano, Italy

<sup>c</sup> Department of Information Systems, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia

<sup>d</sup> Department of Computer Science & Information Technology, The Islamia University of Bahawalpur, Bahawalpur 63100, Pakistan

## ARTICLE INFO

### Keywords:

Breast cancer detection  
Image processing  
Healthcare  
Transfer learning  
Ensemble learning  
Deep convoluted features

## ABSTRACT

Breast cancer (BC) is a common and highly lethal ailment. It stands as the second leading contributor to cancer-related deaths in women worldwide. The timely identification of this condition is of utmost importance in mitigating mortality rates. This research paper presents a novel framework for the precise identification of BC, utilising a combination of image and numerical data features with explainable Artificial Intelligence (XAI). The utilisation of the U-NET transfer learning model is employed for image-based prediction. Additionally, an ensemble model is constructed by integrating characteristics from a customised convolutional neural network (CNN) model with an ensemble comprising random forest (RF) and support vector machine (SVM). The experiments aim to evaluate the influence of original features compared to convoluted features. A comparative analysis is carried out to assess the efficacy of various classifiers in accurately detecting BC, utilising the Wisconsin dataset. The model under consideration exhibits promising capabilities in enhancing BC diagnosis, with a remarkable accuracy rate of 99.99%. The present study contributes to the advancement of BC diagnosis by introducing a novel strategy based on machine learning and discussing the interpretation of the variables using XAI. The primary objective of this approach is to get a notable level of precision, hence facilitating the early and reliable identification of BC. Ultimately, the implementation of this approach is expected to enhance patient outcomes.

## 1. Introduction

Cancer is a worldwide flagellum, leading to millions of deaths per year [1]. As per GLOBOCAN 2020 estimates of cancer incidence and mortality produced by the International Agency for Research on Cancer [2], approximately 2,261,419 new cases of BC were diagnosed worldwide in 2020; there were 24.5% of cancer cases and 15.5% of cancer deaths worldwide among women [3]. Saudi Health Council data indicates that BC accounted for 15.9% of all reported cancers among Saudi citizens and 28.7% among women of all ages. According to the World Health Organization (WHO), breast cancer affects approximately 2.3 million women each year, resulting in 685,000 deaths, accounting for 13.6% of all female cancer-related fatalities [4]. Prompt detection is

critical in reducing the death rate associated with this condition. According to Globocan 2018 data [5], breast cancer is found in one out of every four cases of cancer in women, making it the sixth most prevalent cause of death globally. Breast cancer often develops in the breast tissue, especially the inner linings of milk ducts or lobules. Changes or mutations in Deoxyribonucleic acid (DNA) or Ribonucleic acid (RNA) cause cancer cell formation. Air pollutants, bacteria, nuclear radiation, fungi, physical cell-level damage, viruses, parasites, high temperatures, water contaminants, electromagnetic radiation, dietary elements, free radicals, aging of DNA and RNA, and genetic changes are all factors that can lead to breast cancer. Most breast cancers are no special type (NST), also called invasive ductal carcinomas (IDCs), followed by invasive lobular carcinomas (ILCs) [6], inflammatory breast cancer (IBC) [7], and

\* Corresponding author.

E-mail addresses: [rmonshi@kau.edu.sa](mailto:rmonshi@kau.edu.sa) (R.M. Munshi), [lcascone@unisa.it](mailto:lcascone@unisa.it) (L. Cascone), [namalturki@pnu.edu.sa](mailto:namalturki@pnu.edu.sa) (N. Alturki), [ocsaidani@pnu.edu.sa](mailto:ocsaidani@pnu.edu.sa) (O. Saidani), [amalshardan@pnu.edu.sa](mailto:amalshardan@pnu.edu.sa) (A. Alshardan), [umersabir1996@gmail.com](mailto:umersabir1996@gmail.com) (M. Umer).

<https://doi.org/10.1016/j.imavis.2024.104910>

Received 24 September 2023; Received in revised form 23 November 2023; Accepted 11 January 2024

Available online 20 January 2024

0262-8856/© 2024 Elsevier B.V. All rights reserved.

lobular breast cancer (LBC) [8].

Breast cancer, a dangerous disease with high death rates, accounts for 2.5% of all fatalities, impacting one out of every thirty-nine women [9]. Early identification and treatment of breast cancer are critical since the disease can spread to other parts of the body if left untreated. Early detection and treatment can increase the chance of survival by up to 80%. This emphasizes the critical need of early identification and treatment of breast cancer. Various procedures and methodologies, such as screening examinations, self-assessments, and frequent visits with healthcare specialists, help to detect breast cancer early [10]. Mammography is a popular and efficient approach for detecting breast cancer early. Several studies have confirmed its efficacy in detecting early-stage breast cancer. Another often-used approach for identifying breast cancer is a biopsy. Tissue is removed from the afflicted breast region and examined under a microscope to detect and categorize the tumor during a biopsy [11]. Biopsy is an often-utilized approach for identifying BC. During a biopsy, tissue is removed from the afflicted region of the breast and examined under a microscope in order to diagnose and categorise the tumour. Although present approaches help to reduce breast cancer death rates, there is still room for improvement, particularly in terms of more simplified and automated diagnostic processes. Data mining is a technique used to extract important and noteworthy insights from large data collections. It is recognized as a critical tool for the early detection of several ailments such as renal disease, heart disease [12], diabetes [13], and cancer. Using data mining techniques makes it easier to recognize patterns and trends in data, which aids in the early diagnosis and treatment of certain disorders. Its importance grows when it comes to recognizing illnesses like cancer, when early identification improves survival chances dramatically. Traditional cancer detection consists on three techniques: physical examination, pathology testing, and radiological imaging. These old approaches, however, are time-consuming and prone to false negatives. Machine learning approaches, on the other hand, are gaining popularity due to their superior results. They provide dependability, precision, and speed, revolutionizing disease detection beyond traditional methods and giving more consistent findings across a wide range of ailments.

Over the past decade, recent advances in machine learning have opened up new areas for the medical fields. Machine learning algorithms require the capacity to comprehend predictions, which is referred to as explainable artificial intelligence (XAI). Justification of specific outcomes yields valuable, understandable insights, facilitating informed decision-making and knowledge generation. Researcher applied XAI in detecting breast cancer [14,15]. In a recent study authors applied machine learning with XAI for deting breast cancer in Indonesian patients [16]. Due to the aforementioned benefits, this study proposes a machine learning-based approach leveraging XAI for detecting breast cancer with the goal of achieving high accuracy. This study makes the following contributions in this regard.

- This study presents a complete framework for accurate breast cancer diagnosis that makes use of both image analysis and numerical data characteristics. For image-based prognosis, the predictive model employs U-NET transfer learning.
- An ensemble strategy is proposed in this study to anticipate breast cancer among patients by merging features from a personalized convolutional neural network (CNN) model. The proposed ensemble model combines random forest (RF) and support vector machine (SVM) techniques, with the final prediction determined by a voting process.
- The experiments involve utilizing both the inherent features from the dataset and those derived from the CNN model. The study aims to assess the influence of the original features in comparison to the performance of models utilizing derived features.
- Moreover, the efficacy of the proposed model is assessed by contrasting its performance with state-of-the-art methodologies concerning accuracy, precision, recall, and the F1 score.

The subsequent sections of this study are outlined as follows: [Section 2](#) encompasses recent literature on breast cancer diagnosis and detection. [Section 3](#) elaborates on the dataset, the proposed methodology, and the machine learning classifiers employed. [Section 4](#) comprises results and a comparative analysis. Discussions are presented in [Section 5](#). Lastly, [Section 6](#) encapsulates the conclusion and outlines avenues for future research.

## 2. Related work

Many approaches for BC detection and diagnosis have emerged last years. Recent research has primarily focused on identifying malignancies through CAD and decision support systems. While the majority of investigations depend on individual models to attain dependable results, a handful explore the utilization of ensemble approaches. This section presents an overview of these approaches.

To achieve precise and accurate breast cancer diagnosis, Yadav and Jadhav [17] proposed a machine learning-driven system employing thermal infrared imaging. The researchers evaluated various baseline models alongside transfer learning models such as VGG16 and InceptionV3. Their experiments encompassed both augmented and non-augmented data scenarios. The findings reveal that the transfer learning model InceptionV3 outperforms other models, achieving an accuracy rate of 93.1% without augmentation and 98.5% with augmentation. In a separate study [18], researchers leveraged genetic programming to optimize features for automated breast cancer diagnosis. They experimented with nine machine learning classifiers. The outcomes highlight genetic programming's effectiveness in amalgamating preprocessing and feature selection, leading to the identification of the best model. The AB classifier attained the highest accuracy score of 98.23%. In another research work, researchers proposed an approach for BC detection using extreme learning machine based on feature fusion with CNN deep features. [19]

The authors devised an automated breast cancer detection system employing various models like LR, KNN, SVM, and diverse CNN variations [20]. During experimentation, the authors scrutinized zones of hostile ductal carcinoma tissue within entire slide images. Results indicated that the CNN variant attained the highest accuracy at 87%, surpassing the accuracy of machine learning models by 9%. This suggests that the proposed deep learning system significantly enhances breast cancer detection accuracy. The researchers introduced a voting classifier based on ensemble learning for breast cancer detection [21]. Their findings demonstrated that the ensemble learning model outperformed individual machine learning models. A study proposed a breast tumor type detection approach and achieved 98.1% accuracy [22]. Another study utilized diverse-density mammograms in breast cancer detection, achieving an overall accuracy score of 88.1% [23].

Diverse approaches and techniques, including Breast Magnetic Resonance Imaging (MRI), Mammography, breast ultrasound, biopsy, breast self-exam, and clinical breast exam, help in the timely detection of cancer. Breast mammography and ultrasound are typical screening methods. [24]. MRI is often used in conjunction with mammography or as a supplemental method for women with dense breasts or high-risk BC cancer. Ultrasound has played an important role in BC detection for many years thanks to its portability, availability, and their ability to assess a variety of important aspects from multiple planes with high resolution. These approaches focused on combining machine learning methods with various feature extraction techniques to enhance accuracy in breast mass classification. Results showed high accuracy rates, with DLA-EABA achieving 97.2% accuracy. Aljuaid et al. proposed a transfer learning-based approach in [25] for breast cancer classification. The authors conducted the experiment in two ways: binary classification and multi-class classification.

Mangukiya et al. [26] conducted a study focused on efficient and accurate early breast cancer diagnosis employing multiple techniques. The authors employed a range of machine learning algorithms on a

dataset containing features with diverse units and magnitudes, requiring standardization through standard scaling. Their findings highlighted that, with standard scaling, the XGBoost machine learning algorithm achieved an accuracy score of 98.24%. Similarly, [27] introduced a deep ensemble learning model for breast cancer detection using whole slide images. They utilized various deep-learning models and proposed an ensemble model. The outcomes demonstrated the superiority of the ensemble model over other models, achieving an accuracy score of 86.21%. XAI has been used in many cancer detection tasks like ovarian cancer analysis [28], cancer biomarkers evaluation [29], and cervical cancer cell migration [30]. This research work is a continuation of our previously published paper [31]. Muhammad Umer the co-author of both works helped us in implementing feature based analysis and prediction of breast cancer in this research work and further extension of breast cancer to image based analysis too.

Numerous investigations have delved into employing machine learning models to detect breast cancer, seeking to enhance classification accuracy and minimize pathological errors in automated diagnosis. Table 1 provides an overview of select literature pertaining to breast cancer detection through machine learning models.

### 3. Materials and methods

This section covers the dataset utilized for breast cancer detection, outlines the proposed approach, and details the methodology steps. Additionally, it provides a concise overview of the machine learning classifiers incorporated in this research.

#### 3.1. Dataset for experiments

This research applies supervised machine learning models to detect breast cancer, aiming to assess their performance. The study proceeds through several stages, starting with the acquisition of the dataset [32]. Specifically, the “Breast Cancer Wisconsin Dataset” is sourced from the UCI machine learning repository, accessible to the public. The dataset

contains 32 features including ‘Texture SE’, ‘Texture Mean’, ‘Concave Points Mean’, ‘Concave Points SE’, ‘ID’, ‘Area Worst’, ‘Smoothness Mean’, ‘Symmetry Worst’, ‘Compactness SE’, ‘Radius Mean’, ‘Texture Worst’, ‘Concave Points Worst’, ‘Perimeter SE’, ‘Fractal Dimension SE’, ‘Area Mean’, ‘Perimeter Worst’, ‘Fractal Dimension Mean’, ‘Compactness Worst’, ‘Compactness Mean’, ‘Radius Worst’, ‘Perimeter Mean’, ‘Concavity SE’, ‘Smoothness SE’, ‘Fractal Dimension Worst’, ‘Concavity Mean’, ‘Smoothness Worst’, ‘Symmetry Mean’, ‘Symmetry SE’, ‘Area SE’, ‘Radius SE’, ‘Concavity Worst’, ‘Diagnosis’ (target class). Within the dataset, two distinct target classes exist: benign and malignant. Sample distribution indicates a composition of 45% in the malignant class and 55% in the benign class. The dataset’s 32 features encompass various classifications like numeric, nominal, and binary, among others. Notably, the target class represents a categorical attribute, while the rest of the attributes are numeric.

#### 3.2. Machine learning models for breast cancer prediction

Machine learning classification involves supervised learning, wherein the system assimilates knowledge from a particular dataset to categorize new observations. The dataset can comprise binary or multi-class classifications. This section addresses machine learning classifiers employed in breast cancer detection. The implementation of these models is executed using the scikit-learn library within the Python environment.

##### 3.2.1. Random Forest

Random Forest (RF) stands as a prevalent ensemble learning technique in machine learning, extensively applied for classification and regression tasks [33,34]. It operates by amalgamating multiple decision trees, wherein each tree’s outputs contribute to the ultimate prediction. The core principle involves training numerous decision trees, each on a distinct subset of the data, and subsequently amalgamating their predictions to derive the final output.

##### 3.2.2. Decision Tree

Decision Trees (DT) stand among the most prevalent methods for classification and prediction [35]. Represented as a tree-like structure akin to a flowchart, a DT showcases logical steps. Within this structure, internal nodes portray attribute tests, branches signify attribute test outcomes, and leaf nodes indicate class labels. Decision trees offer significant advantages in data classification, executing this task swiftly and with minimal computational demands.

##### 3.2.3. K-nearest neighbor

The k-NN algorithm, a non-parametric method in machine learning applicable to regression and classification tasks, operates through lazy learning or instance-based learning [36]. It identifies the k closest training instances to a new data point and determines the majority class among these k neighbors to classify the new data point. Grounded on the similarity concept between input and training data, the algorithm stores all available cases and employs a similarity measure, often a distance function, to classify new cases. Renowned for its simplicity, the k-NN algorithm proves straightforward and easily implementable.

##### 3.2.4. Logistic Regression

Logistic Regression (LR) serves as a statistical model employed in supervised learning specifically for binary classification problems. Its prevalent use occurs when the outcome variable is binary, such as predicting the presence or absence of a disease in a patient, or determining if an email is spam or not. LR aims to gauge the probability of a binary outcome considering specific inputs, utilizing this estimation for predictions. Employing the logistic function, also known as the sigmoid function, LR models the probability of a binary outcome, utilizing the resulting output to make predictions [33,37].

**Table 1**  
Summary of the discussed research works.

Ref.	Models	Dataset	Achieved accuracy
[17]	Baseline models and transfer learning models (VGG16 and Inception V3)	PROENG dataset	93.1% without augmentation; 98.5% with augmentation with Inception V3
[18]	k-NN, SVM, GB, GNB, DT, RF, LR, ADA, and LDA	Wisconsin Breast Cancer dataset	98.23% with AB
[20]	LR, KNN, SVM, CNN variants	Kaggle 162 H&E	87% CNN model 3; 78.56% SVM
[21]	RF, KNN, DT, SVM, LR, GBM, proposed (LR + SGD)	Breast Cancer Wisconsin Dataset	100% with (LR + SGD)
[24]	Deep Learning based model (DLA-EABA)	<a href="https://wiki.cancerimagingarchive.net/BreakHis">https://wiki.cancerimagingarchive.net/BreakHis</a>	97.2% using DLA-EABA
[25]	ResNet, Inception-V3Net, and ShuffleNet		99.7% for binary classification with ResNet; 97.81% for multi-class using ResNet
[26]	RF, k-NN, DT, SVM, NB, XGBoost, ADA	Wisconsin breast cancer Dataset	98.24% using XGboost
[27]	CNN, DNN, LSTM, GRU, BiLSTM, CNN-GRU	Histopathologic Cancer Detection	86.21% CNN-GRU
[22]	DT, SVM, RF, LR, k-NN, NB and rotation forest	The University of Wisconsin Hospital dataset	98.1% using logistic regression
[23]	EfficientNet-B5, DenseNet-169	Hallym University Sacred Heart Hospital dataset	88.1% DenseNet-169

### 3.2.5. Support Vector Machine

The Support Vector Machine (SVM), a recognized supervised learning algorithm [38], is employed in machine learning for classification and regression purposes. SVM's core principle revolves around identifying the optimal boundary, referred to as a hyperplane, that segregates data points into distinct classes. This boundary is constructed to maximize the margin, representing the distance between the boundary and the nearest data points from each class, often referred to as support vectors. SVM demonstrates applicability to both linear and non-linear classification tasks.

### 3.2.6. Gradient boosting machine

GBM stands as a machine learning algorithm employed for classification and regression tasks, falling under the boosting ensemble learning family [39]. GBM amalgamates forecasts from multiple weak models, like decision trees, to generate a robust model. The essence of gradient boosting involves iteratively training weak models, such as decision trees, and sequentially integrating them into the ensemble. Each new tree is trained to rectify the errors made by the preceding trees, focusing on the training instances that were previously misclassified. The predictions from all trees are amalgamated to derive the ultimate prediction.

### 3.2.7. Extra tree classifier

ETC operates as an ensemble learning technique employing randomized trees to produce a conclusive classification outcome by amalgamating independent trees within a decision tree forest [40]. Its foundational idea aligns with RF, yet ETC diverges in constructing decision trees within the forest. In ETC, decision-making involves randomly selecting a subset of the top K features and determining the optimal solution using the Gini index. This methodology guarantees the creation of uncorrelated trees within ETC. Gini feature importance assumes a pivotal role in the process of feature selection.

### 3.2.8. Gaussian Naive Bayes

GNB is a popular machine learning technique that is used for classification tasks based on Bayes' Theorem. The likelihood of a hypothesis (class label) given particular evidence (feature values) equals the likelihood of the evidence given the hypothesis multiplied by the prior probability of the hypothesis, according to this theorem [41]. The word 'naive' in GNB refers to the algorithm's strong premise, known as class conditional independence, which states that while the class name is known, all characteristics are independent. Although this assumption seldom matches real-world conditions, the method frequently gives reliable results in actual applications.

### 3.2.9. Stochastic gradient descent

SGD serves as an optimization algorithm employed to reduce a function, especially during model training in machine learning applications like linear regression, logistic regression, and neural networks [42]. It represents a derivative of the gradient descent (GD) algorithm and earns its stochastic nomenclature due to utilizing a random data subset, termed a mini-batch, for estimating the gradient at every iteration. The primary objective of the SGD algorithm revolves around adjusting the model parameters in a direction contrary to the gradient of the loss function concerning the parameters, employing a constant step size designated as the learning rate.

## 3.3. Deep learning models for breast cancer prediction

Deep learning is an increasingly explored subject of artificial intelligence that has shown tremendous potential in data modeling. The incorporation of automated techniques by medical specialists has shown to be quite beneficial and helpful in illness diagnosis. Deep learning is a popular approach for dealing with large amounts of data, eliminating the need for manual feature extraction and finding significant use in

medical data processing.

### 3.3.1. Multilayer perceptron neural network

For relatively small training sets, along with considerations for ease of implementation, rapidity, and swift outcomes, the Multi-Layer Perceptron emerges as the optimal selection [38]. The inherent architecture of MLP encompasses three layers: input, output, and hidden layers. The hidden layer acts as an intermediary linking the input and output layers during neuron processing. The fundamental operation within MLP revolves around the multiplication of input neurons by weights  $w_{ij}$ , where the output  $y_j$  is the resultant sum. Mathematically, its computation is expressed as follows:

$$y_j = f\left(\sum w_{ij} * O_i\right),$$

Within this equation, the weight assignments  $w$  correspond to the gradient descent algorithm, while the hidden layers are symbolized by  $O$ .

## 3.4. RNN

The Recurrent Neural Network (RNN) is the most popular choice for investigating sequential neural networks [43]. When processing, the input sequence from one neuron is passed along to the next in a weighted sequence that mimics the word order in a phrase. The structure of RNN sequences is designed to produce the sequence and predict the next word in the loop.

### 3.4.1. Convolutional neural network

The CNN is a potent neural network design that is skilled at understanding complex correlations between various data properties. In its capacity as a deep learning model, the CNN carefully evaluates incoming pictures, assigns weight to unique objects and characteristics within these images, and makes distinctions between them. The CNN architecture, which consists of input, output, node, and hidden layers, is vital. In order to provide better results, this study uses a customized CNN architecture [44]. The eight layers in the suggested architecture—two dense, two max-pooling, and two convolution layers—play a crucial role. In terms of medical categorization, CNN has exceptional efficacy and precision. The CNN model shows its usefulness by using the Sigmoid as the error function and being supported by a backpropagation method. CNN is useful for classifying a wide range of illnesses, including heart disease, lung disorders, and brain tumors. At the moment, it is essential for handling large-scale medical datasets. There are two pooling possibilities in CNN's pooling layer: average pooling works well for extracting flat features, while maximal pooling works well for sharp feature extraction.

## 3.5. Long short term memory

A refined RNN variant called LSTM shows improved performance when managing long-term sequences. CitePascone et al. (2023). With LSTM, you can overcome the vanishing gradient issue that RNNs have, and you can also see better performance and pattern memorization. With its three key gates—the forget, output, and input gates—LSTM represents a noteworthy development. The following equations provide examples of word sequences.

$$i_t = \sigma(x_t U^i + h_{t-1} W^i + b_i) \quad (1)$$

$$o_t = \sigma(x_t U^o + h_{t-1} W^o + b_o) \quad (2)$$

$$f_t = \sigma(x_t U^f + h_{t-1} W^f + b_f) \quad (3)$$

where  $x_t$  is the input sequence,  $h_{t-1}$  is the preceding hidden state at current step  $t$ ,  $i_t$  is the input gate,  $o_t$  is the output gate and  $f_t$  is the forget



gate.

### 3.5.1. Architecture of convolutional neural network for feature extraction

In this research focusing on breast cancer detection, the CNN deep learning model is employed as a technique for feature extraction [44]. CNN stands as a widely utilized deep learning framework, primarily applied in classification tasks. Leveraging its feature extraction capabilities, convoluted features are harnessed for breast cancer detection. The conventional CNN model comprises four layers: the pooling layer, embedding layer, convolutional layer, and flattening layer. In this breast cancer detection framework, the initial layer utilized in CNN is the embedding layer, boasting an embedding size of 20,000 and an output dimension of 300. Subsequently, the second layer manifests as the convolutional layer, equipped with 5000 filters, a  $2 \times 2$  kernel size, and employing a rectified linear unit (ReLU) as its activation function. Following this, the third layer constitutes the max pooling layer, employing a  $2 \times 2$  size for extracting significant feature maps from the convolutional layer's output. Ultimately, a flatten layer is employed to convert the output into a 1D array conducive for the learning models.

Consider a tuple set  $(fs_i, tc_i)$  originating from the breast cancer dataset, where  $fs$  denotes the feature set,  $tc$  signifies the target class column, and  $I$  represents the index of the tuple. To facilitate the transformation of the training set into the necessary input, the embedding layer is utilized.

$$EL = \text{embedding\_layer}(Vs, Os, I) \quad (4)$$

$$EOs = EL(fs) \quad (5)$$

Here, the embedding layer is denoted by  $EL$ , while the output of the embedding layer is indicated by  $EOs$ . The typical layer receives this output as its input. Three independent factors control  $EL$ :  $Vs$ , which is the vocabulary size;  $I$ , which is the input length; and  $Os$ , which is the output dimension. The  $EL$  size in this breast cancer detection research is 20,000, meaning that it can accept inputs in the range of 0 to 20,000. The output dimension,  $Os$ , is defined as 300, while the input length is set at 32. All of the incoming data is processed by  $EL$ , which also provides the output that CNN needs for its further processing steps.  $EOs = (\text{None}, 32, 300)$  is the expression for the output dimension of  $EL$ .

$$1D - \text{Convs} = \text{CNN}(F, Ks, AF) \leftarrow EOs \quad (6)$$

The output of the convolutional layer derives from the output of  $EL$ . The CNN structure integrates 500 filters denoted as  $F = 500$ ,

accompanied by a kernel size of  $2 \times 2$ . The ReLU activation function serves to nullify negative values to zero while preserving all other values unchanged.

$$f(x) = \max(0, E)s \quad (7)$$

The architecture of the used CNN together with the prediction model are shown in Fig. 1.

### 3.6. Methodology proposed for numerical data features

Ensemble models are becoming more popular and have demonstrated increased effectiveness and accuracy in classification tasks. Combining different classifiers can make it possible to outperform individual models. An ensemble learning approach is used in this study to improve breast cancer diagnosis. The suggested method uses a voting classifier that uses the soft voting criterion to merge RF and SVM. The class with the greatest number of votes determines the final forecast. The described ensemble model operates as follows, as seen in Algorithm 1:

$$\hat{p} = \text{argmax} \sum_i^n RF_i, \sum_i^n SVM_i \quad (8)$$

**Algorithm 1.** Ensemble Voting Classifier for Breast Cancer Prediction.

---

**Require:** Training data  $X$ , labels  $y$

- 1: Split data into training and testing sets
- 2: Train Random Forest classifier  $RF$  on  $X, y$
- 3: Train Support Vector Machine classifier  $SVM$  on  $X, y$
- 4: Initialize empty array  $predictions$
- 5: **for**  $i$  in  $1$  to  $\text{len}(X_{\text{test}})$  **do**
- 6:  $pred_{RF} \leftarrow RF.predict(X_{\text{test}}[i])$
- 7:  $pred_{SVM} \leftarrow SVM.predict(X_{\text{test}}[i])$
- 8:  $ensemble\_vote \leftarrow \text{majority\_vote}(pred_{RF}, pred_{SVM})$
- 9: Append  $ensemble\_vote$  to  $predictions$
- 10: **end for**
- 11: **return**  $predictions$

---

The predictive probabilities for individual test samples are generated by  $\sum_i^n RF_i$  and  $\sum_i^n SVM_i$ . These probabilities, as depicted in Fig. 2, undergo the soft voting criterion, generating probabilities for each test case through RF and SVM. The voting process is visualized in Fig. 3.

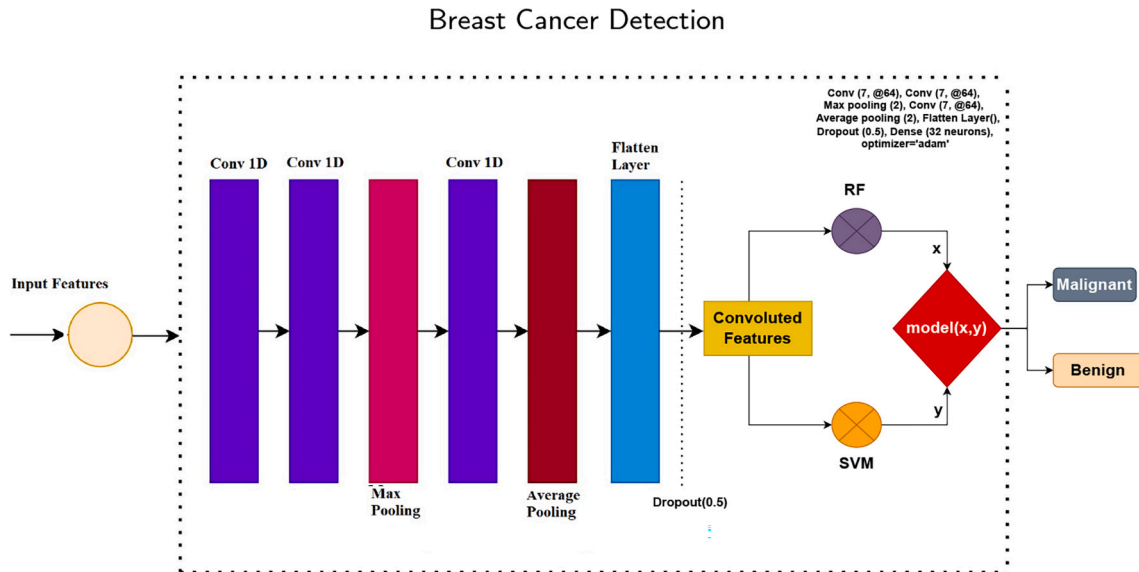


Fig. 1. Architecture diagram of the CNN with voting classifier (RF + SVM) model.

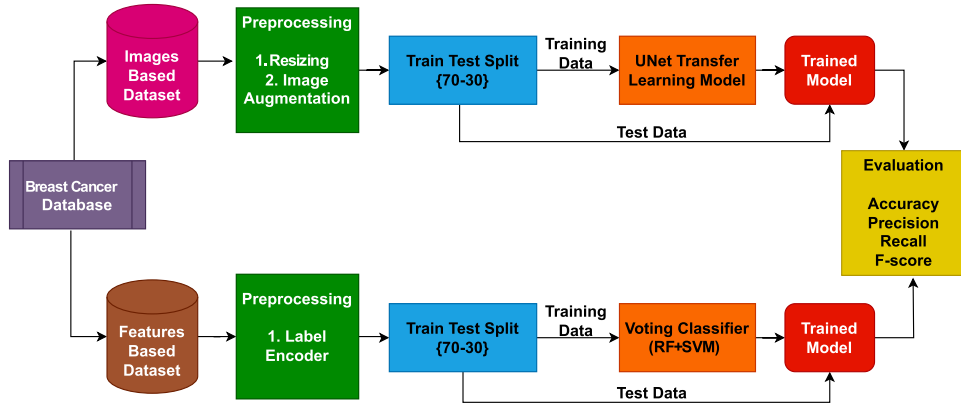


Fig. 2. Workflow diagram of the proposed voting classifier (RF + SVM) model.

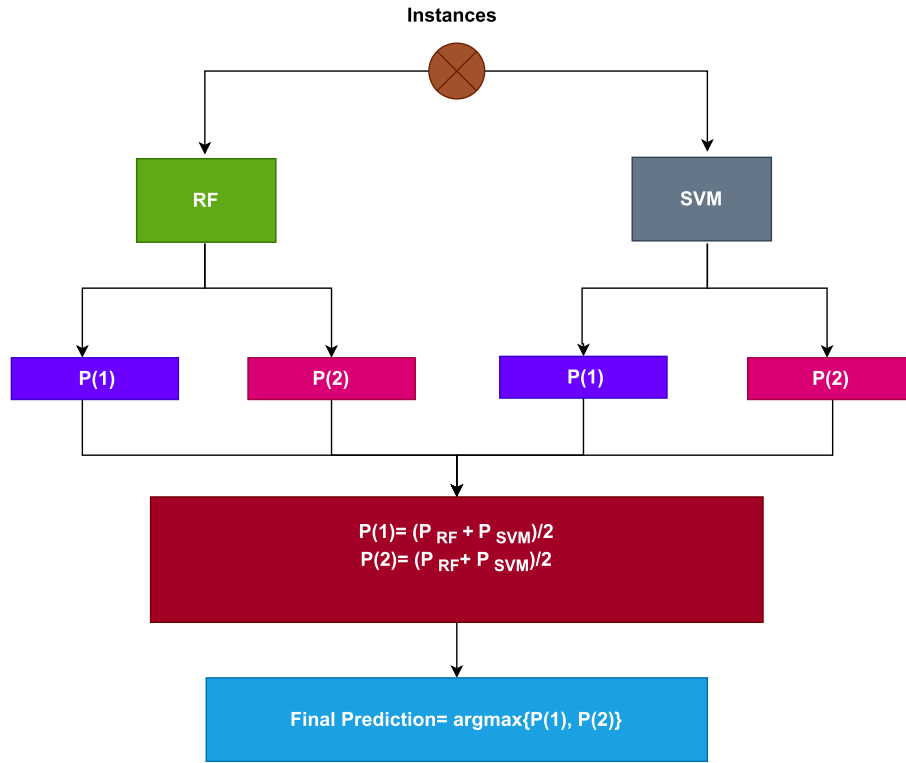


Fig. 3. Architecture of the proposed voting classifier (RF + SVM) model.

$$VC(RF + SVM) = \text{argmax}(g(x)) \quad (9)$$

To assess the performance of the proposed VC(RF + SVM) model, it undergoes testing on the 'Breast Cancer Wisconsin Dataset' in a two-stage evaluation. Initially, breast cancer detection is executed using all 32 features available in the dataset. In the subsequent experimental stage, the dataset is preprocessed to employ convolutional features for machine learning models. The data is partitioned into two segments, with 70% allocated for training purposes and the remaining 30% reserved for testing. This strategy, recognized as the training-testing split, serves as a common methodology in machine learning, facilitating the evaluation of the model's accuracy on new and unseen data.

### 3.7. Transfer learning models

Transfer learning stands as a widely embraced method in machine learning and computer vision, utilizing pre-trained models from

expansive datasets to effectively address new tasks. When applied to brain tumor classification via image data, two frequently employed transfer learning models are U-Net and MobileNet.

#### 3.7.1. U-Net

U-Net, brought forth by Olaf Ronnenberg and associates in 2015, stands as a significant model in the realm of image segmentation. Originally tailored for biomedical image analysis, U-Net swiftly garnered acclaim due to its impressive accuracy and efficacy [45]. Its notable strength lies in achieving remarkable results, even when confronted with a scarcity of training data, a common challenge in medical image segmentation. The model's architecture embodies two primary pathways, resembling an auto-encoder. The initial pathway, termed the contracting or compressive path, functions as the encoder, utilizing a standard deep CNN network. Conversely, the decoder or expanding path (also known as the up-sampling or synthesizing path in some literature) integrates deconvolutional and convolutional layers. The contracting

path reduces the resolution of input images, while the expanding path skillfully restores the original image quality and spatial structure via optimized techniques, such as concatenated skip connections.

As the network progresses, it assimilates spatial classification information, delivering dense predictions at higher resolutions along the ascending path. It also enhances the output image's resolution, leading to final processing through a convolutional layer, yielding a segmented image mirroring the dimensions of the input image. This process maintains the constant shape of the input image throughout segmentation, ensuring precise representation. While classification remains essential in medical image analysis, it lacks pixel-level contextual representation, providing a single label for the entire image. U-Net, in conjunction with subsequent refinement techniques, was explicitly engineered to proficiently handle multi-dimensional tensors, accommodating inputs with three or four dimensions. The network produces an output mirroring the input's shape. Since its inception, U-Net has laid a robust foundation for extensive exploration in medical image segmentation. Numerous advancements have surfaced, either through adjustments to the original U-Net structure or its integration with alternative architectures.

### 3.7.2. Mobilenet

MobileNet is a streamlined CNN structure, aims to facilitate efficient processing on resource-constrained devices, such as mobile phones [46]. It capitalizes on depthwise-separable convolutions, significantly reducing parameter count and computational complexity while upholding model effectiveness. Due to its efficiency, MobileNet is widely applied in diverse image classification tasks. Leveraging transfer learning with a pre-trained MobileNet model enables the extraction of valuable features from extensive image datasets, thereby enhancing the accuracy of brain tumor classification. The MobileNet architecture is meticulously designed to excel in both efficiency and effectiveness, especially in scenarios with limited computational resources or tasks requiring minimal features, such as palmprint recognition. Its salient aspect lies in its depthwise structure, allowing high performance while mitigating computational demands. The consideration of pointwise complexity involving a  $1 \times 1$  convolution is pivotal, and ReLU activation is employed to maintain pointwise connections while generating deep abstraction layers.

Furthermore, a resolution multiplier variable labeled as  $w$  is introduced, aimed at diminishing the dimensionality of both the input image and the internal representation within each layer. This variable facilitates adjustments in network dimensions while ensuring a consistent value for " $w$ " across all layers, thereby bolstering the overall efficiency of the model. The feature vector map, marked as  $F_m$  in size, corresponds with a filter size of  $F_s$ , while the input and output variables are symbolized by  $p$  and  $q$  respectively. The equation below offers a measure of the computational workload for the primary abstraction layers within the architecture, where the variable  $ce$  represents the computational effort.

$$c_e = F_s \cdot F_s \cdot w \cdot \alpha F_m + w \cdot \rho \cdot \alpha F_m \cdot \alpha F_m \quad (10)$$

In the MobileNet framework, the multiplier variable  $w$  assumes values contextual to the study. When focusing on brain tumor classification experiments, this multiplier typically ranges between 1 and  $n$ . Meanwhile, the resolution multiplier  $r$  remains constant at 1. The equation following calculates the computational workload, represented as  $cost_e$ .

$$cost_e = F_s \cdot F_s \cdot w \cdot \rho \cdot F_m \cdot F_m \quad (11)$$

### 3.8. Evaluation metrics

The evaluation of the machine learning models in this study relies on metrics like accuracy, precision, recall, and F1 score, computed from the values in the confusion matrix. These matrices range between 0 and 1,

indicating the minimum and maximum values, respectively.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (12)$$

$$Precision = \frac{TP}{TP + FP} \quad (13)$$

$$Recall = \frac{TP}{TP + FN} \quad (14)$$

$$F1score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (15)$$

## 4. Results

A series of experiments were undertaken to identify breast cancer, employing machine learning models with both the original features and the convoluted features. The outcomes were analyzed, and further experiments were conducted using an ensemble comprising the top four individual machine learning models, applied to both sets of features.

### 4.1. Experiment set up

Experiments are conducted on a system with an Intel Core i7 processor and a graphics processing unit (GPU) from NVIDIA using the Python 3.8 programming environment. Table 2 contains the specifications of the hardware and software used in the experiment. The proposed model took around 57 min to get trained and making predictions.

### 4.2. Performance comparison of individual machine learning models on original and convoluted features

The present study utilizes a total of nine machine learning models, each fine-tuned with optimized hyperparameters to yield improved results. These parameters are determined through empirical methods to ensure a high level of accuracy. Notably, RF demonstrates superior performance when employed in conjunction with the original features, achieving an accuracy score of 91%, while the ETC model attains a score of 89%. Among the tested models, the k-NN algorithm exhibits the lowest performance, with an accuracy score of 81%. Refer to Table 3 for a detailed presentation of the accuracy scores obtained by all classifiers when applied using the original features.

Table 4 outlines the classification accuracy of machine learning models when applied with convoluted features. The results of the experiments indicate that the RF and ETC models outperform other models, achieving accuracy scores of 93.75% and 93.74%, respectively. Similarly, it is evident that both SVM and LR classifiers boast superior accuracy compared to the other classifiers.

### 4.3. Performance of ensemble models using original features

ChatGPT Initially, individual models are employed to analyze both the original features and convoluted features, with the outcomes presented in Table 3 and Table 4. Among the nine machine learning models considered, it is observed that the RF, ETC, LR, and SVM models consistently yield the most favorable outcomes across both feature sets. In this phase of the study, ensembles comprising these machine learning

**Table 2**

Experimental setup for the proposed system.

Libraries	Scikitlearn, TensorFlow
Language	Python 3.8
RAM	8GB
OS	64-bit window 10
CPU	Core i7, 7th Gen with 2.8 GHz processor
GPU	Nvidia, 1060, 8GB

**Table 3**

Accuracy of models with original features.

Model	Accuracy with original features
RF	91.78%
ETC	89.47%
LR	88.59%
SVM	88.47%
GNB	84.89%
KNN	81.77%
GBM	85.86%
DT	86.88%
SGD	84.47%

**Table 4**

Classifiers accuracy with convoluted features.

Model	Accuracy with convoluted features
RF	93.75%
ETC	93.74%
LR	91.85%
SVM	92.34%
GNB	89.47%
KNN	86.53%
GBM	87.84%
DT	90.37%
SGD	90.69%

models are evaluated using the original features. The findings from the ensemble learning models indicate that the RF + SVM ensemble model exhibits superior performance in terms of accuracy, achieving a rate of 95%. This accuracy rate is almost 2% greater than that of the other ensemble learning models. The subsequent model, SVM + ETC, attains a classification accuracy of 92%. The RF + SVM model demonstrates a precision of 95%, recall of 98%, and an F1 score of 96% in the context of breast cancer diagnosis. Table 5 displays the outcomes of the ensemble learning models on the original feature set.

#### 4.4. Performance of ensemble model on convoluted features

The performance of the ensemble models is also evaluated using the features extracted by the customized CNN model. The results of these experiments are presented in Table 6. The findings indicate that the RF + SVM model outperforms alternative models, achieving an accuracy of 99% as well as precision, recall, and F1 scores of 99% each. The ETC + LR model has demonstrated the least favorable outcomes, achieving an accuracy rate of 94%. The utilization of features from the CNN model in conjunction with ensemble learning models yields superior outcomes compared to the utilization of the original features alone.

#### 4.5. Results of K-fold cross-validation

K-fold cross-validation is implemented to evaluate the effectiveness of the proposed model. The objective of cross-validation is to validate the results of the presented model and assess its robustness. This technique is employed to evaluate the model's performance across various subsets of the data. In this study, a 5-fold cross-validation approach is utilized, and the corresponding results are detailed in Table 7. The cross-

**Table 5**

Ensemble model results using Original Features set.

Model	Accuracy	Precision	Recall	F1 score
RF + SVM [31]	95.89%	95.91%	98.54%	96.99%
RF + ETC [31]	93.34%	93.45%	95.11%	94.37%
RF + LR [31]	89.55%	90.65%	88.25%	89.17%
ETC + SVM [31]	94.14%	93.78%	95.64%	94.24%
ETC + LR [31]	90.34%	91.45%	91.67%	91.55%
SVM + LR [31]	91.73%	92.64%	96.98%	95.74%

**Table 6**

Ensemble model results using convoluted Features set.

Model	Accuracy	Precision	Recall	F1 score
RF + SVM [31]	99.99%	99.99%	99.99%	99.99%
RF + ETC [31]	97.21%	97.65%	98.47%	97.54%
RF + LR [31]	95.62%	96.81%	97.14%	96.67%
ETC + SVM [31]	97.77%	97.45%	97.45%	97.45%
ETC + LR [31]	94.39%	95.27%	97.69%	96.44%
SVM + LR [31]	96.25%	97.34%	97.74%	97.54%

**Table 7**

Results for k-fold cross-validation of the proposed ensemble model.

Fold Number	Accuracy	Precision	Recall	F1 score
Fold-1	99.23%	99.96%	99.94%	99.95%
Fold-2	99.34%	99.96%	99.95%	99.96%
Fold-3	99.45%	99.97%	99.96%	99.96%
Fold-4	99.11%	99.94%	100.0%	99.99%
Fold-5	99.24%	99.99%	99.98%	99.99%
Average	99.27%	99.96%	99.96%	99.97%

validation results illustrate that the ensemble model proposed in this study achieves an average accuracy score of 0.996. Additionally, the average precision, recall, and F1 scores for the model are 0.998, 0.998, and 0.997, respectively.

## 5. Discussion

### 5.1. Deep learning model performance with original features

The application of deep learning models involves utilizing both original and convoluted features, and the results from these models are outlined in Table 8. Among the three deep learning models, it is evident that CNN exhibits superior performance on both sets of features. In this phase of the experiments, the proposed model's validity is assessed by comparing it to state-of-the-art deep learning models. However, the ensemble model outperforms the deep learning models in terms of accuracy. The accuracy of MLP decreases when using CNN features, while LSTM accuracy improves. This improvement can be attributed to LSTM's capacity to extract more meaningful features, facilitating sequence generation. CNN's accuracy remains nearly the same because it receives the same convoluted features and an additional layer for predictions.

#### 5.1.1. Results comparison of transfer learning models

This section will undertake an analysis of the outcomes derived from two transfer learning models, namely MobileNet and UNET, in relation to the utilization of image data. Table 9 demonstrates the complete classification report of the two models and clearly states that the UNET model exhibits superior accuracy, precision, recall, and F1 score compared to the MobileNet model.

#### 5.1.2. Comparative evaluation of image and feature-based datasets

In order to assess the efficacy of the U-Net transfer learning model, a comparative analysis was undertaken to analyze the performance of both the machine learning model and transfer learning models using numerical and image data. The findings indicate a significant enhancement in the efficacy of the transfer learning models. The outputs

**Table 8**

Accuracy of deep learning models with original and convoluted features.

Model	Accuracy	
	Original features	Convoluted features
MLP	87.69%	84.41%
CNN	90.22%	90.70%
LSTM	85.95%	88.34%



**Table 9**

Results comparison of transfer learning models.

Model	Accuracy	Precision	Recall	F1 score
MobileNet	98.59%	98%	98%	98%
U-Net	99.97%	99%	99%	99%

obtained by the machine learning models in both scenarios are presented in Table 10, providing a comprehensive overview that enables a full study of their performance.

### 5.2. Performance comparison with existing studies

To showcase the effectiveness of the proposed model in comparison to prior state-of-the-art models, a comprehensive analysis is conducted by contrasting the outcomes with those obtained from existing models. To fulfill the objectives of this study, a meticulous selection process was employed to identify the nine most relevant research articles. For instance, [47] utilized PCA features with the SVM machine learning model for breast cancer detection, achieving an accuracy score of 96.99%. In another study [48], autoencoder was employed, yielding the highest accuracy score of 98.40%. The use of Quadratic SVM by [49] resulted in an accuracy score of 98.11%. For the same task, [50] applied XgBoost and achieved an accuracy score of 97.11%. Similarly, [51,52] utilized Chi-square features and the LR machine learning model with accuracy scores of 98.21% and 98.10%, respectively. The performance comparison between the planned and existing studies is detailed in Table 11. The results underscore the superior performance of the proposed model.

### 5.3. Significance of the proposed model

In order to validate the performance of the proposed model, we tested it on two further independent datasets. The first dataset [56] is 'Breast Cancer Survival', which contains 330 patient records with the feature Patient\_ID, Age, Gender, and expression levels of four proteins (Protein1, Protein2, Protein3, Protein4). The dataset also includes the Breast cancer stage of the patient (Tumor\_Stage), Histology (type of cancer), ER, PR, and HER2 status, Surgery\_type, Date of Surgery, Date of Last Visit, and Patient Status (Alive/Dead). The second dataset [57] contains 10 Quantitative features to show the presence or absence of breast cancer in a patient. The features are Age (years), BMI (kg/m2), Glucose (mg/dL), Insulin ( $\mu$ U/mL), HOMA, Leptin (ng/mL), Adiponectin ( $\mu$ g/mL), Resistin (ng/mL), MCP-1(pg/dL), and Labels (absence or presence). Fig. 4 presents the accuracy comparison of the classifiers used in experiments. The proposed model obtained 97.34% accuracy on the first dataset and 96.67% accuracy on the dataset which greatly shows the stability of the proposed model on all kinds of datasets. Important terms are explained in Table 13.

### 5.4. Shapley additive explanations

Deciphering the connections between inputs and outputs in learning models is often intricate due to the perceived complexity of these models as mysterious, black-box algorithms. This lack of transparency poses a significant challenge in grasping the significance of features, especially

**Table 10**

Comparison analysis of image and feature-based dataset techniques.

Features Based Data			Image Data	
Best ML Performing Models' Results			Best TL Models' Results	
Models	Original Features	CNN Features	Models	Accuracy
VC (RF + SVM)	95.8%	99.9%	U-Net	99.99%
VC (ETC + SVM)	94.1%	97.7%	MobileNet	98.5%

**Table 11**

Performance comparison with state-of-the-art studies.

Ref.	Technique	Accuracy
[53]	K-means clustering	92.01%
[47]	PCA features with SVM	96.99%
[49]	Quadratic SVM	98.11%
[48]	Auto-encoder	98.40%
[54]	GF-TSK	94.11%
[50]	XgBoost	97.11%
[55]	Five most significant features with LightGBM	95.03%
[51]	Chi-square features	98.21%
[52]	LR with all features	98.10%
<b>Proposed</b>	Deep convoluted features with voting classifier (RF + SVM)	99.99%

in supervised learning scenarios with labeled data. Fortunately, the SHAP technique, introduced as a recent breakthrough, provides a quantitative remedy to this challenge, allowing us to enhance model interpretability. First proposed by Lee and Lundberg in 2017 and expanded upon by Lundberg et al. in 2018 [58,59], this innovation has notably advanced our understanding of the components within the model.

SHAP employs a linear additive feature attribution method rooted in concepts from cooperative game theory to clarify complex models. This method assigns an importance value to each feature, taking into account its impact on the model's predictions based on its presence or absence during SHAP estimation. Such an explanatory approach simplifies the functioning of intricate models, rendering them more comprehensible. The principles underlying the linear additive feature attribution technique, drawing from cooperative game theory, are expounded upon in works by Lee and Lundberg (2017) and further developed by Lundberg et al. (2020) [58,59].

$$f(a) = g(a') = \phi_0 + \sum_{j=1}^j \phi_j a'_j \quad (16)$$

In this scenario, we denote the original learning model as (a), while the more straightforward explanatory model is referred to as  $g(a')$ . Here,  $a'_j$ , where  $j$  represents a simplified seismic attribute number, represents these attributes. To calculate SHAP values, labeled as "j," we consider all potential input orderings. In the estimation process,  $a'_j$  indicates the presence or absence of a specific seismic attribute, using an input vector. Finally,  $\phi_0$  represents the model prediction when none of the attributes are taken into account during the estimation. The comprehensive feature importance, calculated using SHAPly, is presented in descending order in Table 12.

The SHAP analysis underscores the significance of features in the prediction of breast cancer. While SHAP feature importance surpasses traditional methods, relying solely on it provides somewhat constrained additional insights. SHAPly XAI provides a more comprehensive and information-laden portrayal of SHAP values, unveiling not just the relative importance of features but also their intricate connections with the predicted outcome. SHAP explanations furnish valuable insights into how features contribute to a specific instance. The cumulative sum of feature contributions, in conjunction with the bias term, equals the model's initial prediction, representing the forecast before applying the inverse link function.

### 5.5. Limitations

The present study contributes significantly by proposing a machine learning-based approach leveraging XAI for breast cancer detection, aiming for high accuracy. However, certain limitations are evident:

- The study's effectiveness might be influenced by the size of the dataset used for training and testing the models. A larger and more

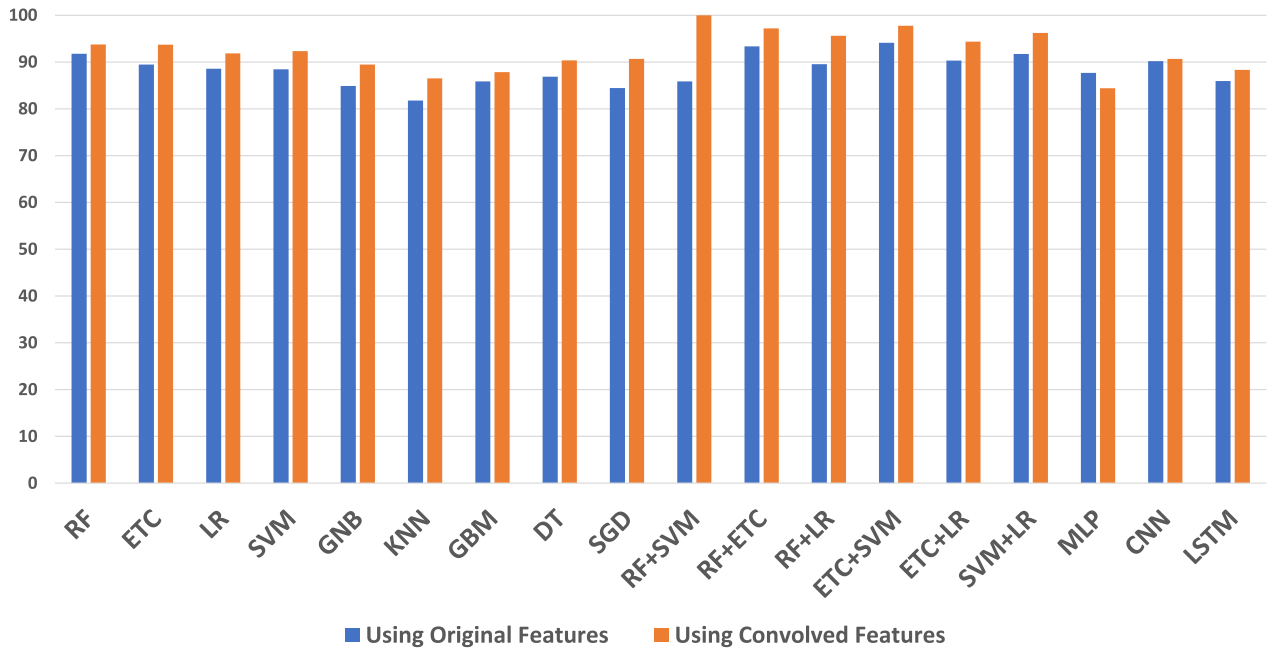


Fig. 4. Accuracy comparison of the classifiers.

**Table 12**  
Feature importance calculated using SHAPly.

Weight	Feature
0.1756 ± 0.0468	Texture SE
0.0271 ± 0.0482	Concave Points SE
0.0464 ± 0.0474	Concave Points Mean
0.0027 ± 0.0077	Smoothness Mean
0.0046 ± 0.0141	Area Worst
0.0037 ± 0.0244	Symmetry Worst
0.0910 ± 0.0849	Texture Mean
0.0048 ± 0.0426	Radius Mean
0.0059 ± 0.0074	Compactness SE
0.0011 ± 0.0241	Texture Worst
0.0011 ± 0.0011	Perimeter SE
0.0011 ± 0.0221	Concave Points Worst
-0.0038 ± 0.0067	Area Mean
-0.0048 ± 0.0077	Fractal Dimension SE
-0.0069 ± 0.0245	Perimeter Worst
-0.0049 ± 0.0088	Compactness Worst
-0.0029 ± 0.0054	Fractal Dimension Mean
-0.0031 ± 0.0048	Compactness Mean
-0.0068 ± 0.0046	Perimeter Mean
-0.0048 ± 0.0345	Radius Worst
0.0011 ± 0.0261	Concavity SE
-0.0038 ± 0.0077	Fractal Dimension Worst
0.0011 ± 0.0011	Smoothness SE
-0.0038 ± 0.0057	Concavity Mean
-0.0039 ± 0.0044	Smoothness Worst
-0.0039 ± 0.0275	Perimeter Worst
-0.0049 ± 0.0058	Symmetry Mean
-0.0048 ± 0.0365	Area SE
-0.0051 ± 0.0078	Symmetry SE
-0.0068 ± 0.0056	Concavity Worst
-0.0068 ± 0.0056	Radius SE

diverse dataset could offer a more comprehensive understanding of the model's performance.

- The models' robustness and generalizability could benefit from external validation on datasets from different sources or demographics, ensuring their applicability beyond the current dataset.
- The computational resources required for implementing some of the proposed models.

**Table 13**  
Acronym Table.

Abbreviation	Description
AB	AdaBoost
BiLSTM	Bidirectional LSTM
CAD	computer-aided diagnostics
CNN	convolutional neural network
DNA	Deoxyribonucleic acid
DCIS	Ductal Carcinoma in situ
DNN	deep neural network
DT	Decision Tree
ETC	extra tree classifier
GNB	Gaussian Naive Bayes
GBM	Gradient boosting machine
GRU	gated recurrent unit
IDC	Invasive ductal carcinoma
KNN	K-nearest neighbor
LDA	Latent Dirichlet Allocation
LBC	Lobular breast cancer
LSTM	long short-term memory
LR	Logistic regression
MBC	Mucinous breast cancer
MRI	magnetic resonance imaging
MTBC	Mixed tumor Breast cancer
RF	Randon Forest
RNA	Ribonucleic acid
RNN	Recurrent Neural Network
SGD	stochastic gradient descent
SVM	Support Vector Machine
XAI	explainable artificial intelligence

Despite these limitations, the study's contributions in proposing a comprehensive framework for accurate breast cancer detection using machine learning techniques, including the ensemble model, and comparing their performance against various benchmarks are valuable steps toward advancing diagnostic methodologies.

## 6. Conclusions

This research paper presents a novel framework for the precise identification of BC, utilising a combination of image and numerical data features with explainable Artificial Intelligence (XAI). The

utilisation of the U-NET transfer learning model is employed for image-based prediction. Additionally, an ensemble model is constructed by integrating characteristics from a customised convolutional neural network (CNN) model with an ensemble comprising random forest (RF) and support vector machine (SVM). The experiments aim to evaluate the influence of original features compared to convoluted features. A comparative analysis is carried out to assess the efficacy of various classifiers in accurately detecting BC, utilising the Wisconsin dataset. The model under consideration exhibits promising capabilities in enhancing BC diagnosis, with a remarkable accuracy rate of 99.99%. The present study contributes to the advancement of BC diagnosis by introducing a novel strategy based on machine learning and discussing the interpretation of the variables using XAI. The primary objective of this approach is to get a notable level of precision, hence facilitating the early and reliable identification of BC. Ultimately, the implementation of this approach is expected to enhance patient outcomes. This study aims to establish a comprehensive framework for effectively identifying BC using explainable Artificial Intelligence (XAI) coupled with image and numerical data features, thereby mitigating the associated mortality risk linked to this prevalent reason of death among women. We employed U-NET transfer learning model for image-based prediction. In addition, We proposed an ensemble model, leveraging its reported superior performance in the existing literature. Furthermore, our training process employs features derived from the CNN model. The accuracy of the suggested model in distinguishing patients from normal ones is an impressive 0.999. Moreover, the use of CNN-based features in conjunction with models consistently leads to more optimal outcomes. We demonstrate the efficiency and durability of the provided model by doing K-fold cross-validation and comparing its performance to established state-of-the-art models. In future studies, we plan to apply our model to a variety of datasets spanning different domains, including BC pictures, and analyse the corresponding numerical values produced from the microscopic characteristics in these images.

#### CRedit authorship contribution statement

**Raafat M. Munshi:** Data curation, Resources, Writing – original draft. **Lucia Cascone:** Project administration, Writing – review & editing. **Nazik Alturki:** Conceptualization, Methodology, Funding, Project administration, Writing – review & editing. **Oumaima Saidani:** Conceptualization, Methodology, Writing – original draft. **Amal Alshardan:** Conceptualization, Methodology, Software, Visualization. **Muhammad Umer:** Methodology, Software, Writing – original draft.

#### Declaration of competing interest

The authors declare no conflict of interest.

#### Data availability

The datasets generated during and/or analyzed during the current study can be downloaded from the following link. <https://github.com/MUmerSabir/MDPIDiagnostic>

#### Acknowledgments

This research work was funded by Institutional Fund Projects under grant no. (IFPIP: 1328-415-1443) and by Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2024R333), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia. The authors gratefully acknowledge the technical and financial support provided by the Ministry of Education and King Abdulaziz University, DSR, Jeddah, Saudi Arabia.

#### References

- [1] WHO. World Health Organization, Cancer: Key Facts. 2022. <https://www.who.int/news-room/fact-sheets/detail/cancer>, July 2022. Online; accessed 10 January 2023.
- [2] Yi-Sheng Sun, Zhao Zhao, Zhang-Nv Yang, Xu Fang, Lu Hang-Jing, Zhi-Yong Zhu, Wen Shi, Jianmin Jiang, Ping-Ping Yao, Han-Ping Zhu, Risk factors and preventions of breast cancer, *Int. J. Biol. Sci.* 13 (11) (2017) 1387.
- [3] WHO. World Health Organization, Breast Cancer. <https://www.who.int/news-room/fact-sheets/detail/breastcancer>, July 2022. Online; accessed 10 January 2023.
- [4] WHO. World Health Organization, Breast Cancer. <https://www.who.int/news-room/fact-sheets/detail/breast-cancer>, March 2021. Online; accessed 10 January 2023.
- [5] Freddie Bray, Jacques Ferlay, Isabelle Soerjomataram, Rebecca L. Siegel, Lindsey A. Torre, Ahmedin Jemal, et al., Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries, *CA Cancer J. Clin.* 68 (6) (2018) 394–424.
- [6] Amrita Ray Chaudhury, Ranjani Iyer, Kaveri K. Iychettira, A. Sreedevi, Diagnosis of invasive ductal carcinoma using image processing techniques, in: *2011 International Conference on Image Information Processing, IEEE*, 2011, pp. 1–6.
- [7] Fredika M. Robertson, Melissa Bondy, Wei Yang, Hideko Yamauchi, Shannon Wiggins, Samira Kamrudin, Savitri Krishnamurthy, Huong Le-Petross, Luc Bidaut, Audrey N. Player, et al., Inflammatory breast cancer: the disease, the biology, the treatment, *CA Cancer J. Clin.* 60 (6) (2010) 351–375.
- [8] Serena Masciari, Nina Larsson, Janine Senz, Niki Boyd, Pardeep Kaurah, Michaela J. Kandel, Lyndsay N. Harris, Hugo C. Pinheiro, Armelle Troussard, Penelope Miron, et al., Germline e-cadherin mutations in familial lobular breast cancer, *J. Med. Genet.* 44 (11) (2007) 726–731.
- [9] American Cancer Society. American Cancer Society, Breast Cancer. <https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/breast-cancer-facts-and-figures/breast-cancer-facts-and-figures-2015-2016.pdf>, July 2022. Online; accessed 10 January 2023.
- [10] A. Pasumpon, Pandian., Identification and classification of cancer cells using capsule network with pathological images, *J. Artif. Intell.* 1 (01) (2019) 37–44.
- [11] Andrei Chekkoury, Parmeshwar Khurd, Jie Ni, Claus Bahlmann, Ali Kamen, Amar Patel, Leo Grady, Maneesh Singh, Martin Groher, Nassir Navab, et al., Automated malignancy detection in breast histopathological images, in: *Medical Imaging 2012: Computer-Aided Diagnosis* 8315, SPIE, 2012, pp. 332–344.
- [12] Furqan Rustam, Abid Ishaq, Kashif Munir, Mubarak Almutairi, Naila Aslam, Imran Ashraf, Incorporating cnn features for optimizing performance of ensemble classifier for cardiovascular disease prediction, *Diagnostics* 12 (6) (2022) 1474.
- [13] Vaibhav Rupapara, Furqan Rustam, Abid Ishaq, Ernesto Lee, Imran Ashraf, Chi-square and pca based feature selection for diabetes detection with ensemble classifier, *Intell. Automat. Soft Comput.* 36 (2) (2023).
- [14] Muhammad Idrees, Ayesha Sohail, Explainable machine learning of the breast cancer staging for designing smart biomarker sensors, *Sensors Int.* 3 (2022) 100202.
- [15] M. Rodríguez-Sampaio, Mariano Rincón, Sonia Valladares-Rodríguez, M. Bachiller-Mayoral, Explainable artificial intelligence to detect breast cancer: A qualitative case-based visual interpretability approach, in: *International Work-Conference on the Interplay between Natural and Artificial Computation*, Springer, 2022, pp. 557–566.
- [16] Fabián Silva-Aravena, Hugo Núñez Delafuente, Jimmy H. Gutiérrez-Bahamondes, Jenny Morales, A hybrid algorithm of ml and xai to prevent breast cancer: a strategy to support decision making, *Cancers* 15 (9) (2023) 2443.
- [17] Samir S. Yadav, Shivajirao M. Jadhav, Thermal infrared imaging based breast cancer diagnosis using machine learning techniques, *Multimed. Tools Appl.* (2022) 1–19.
- [18] Habib Dhahri, Eslam Al Maghayreh, Awais Mahmood, Wail Elkilani, Mohammed Faisal Nagi, et al., Automated breast cancer diagnosis based on machine learning algorithms, *J. Healthcare Eng.* 2019 (2019).
- [19] Z. Wang, M. Li, H. Wang, H. Jiang, Y. Yao, H. Zhang, J. Xin, Breast cancer detection using extreme learning machine based on feature fusion with CNN deep features, *IEEE Access* 7 (2019) 105146–105158.
- [20] Saad Awadh Alanazi, M.M. Kamruzzaman, Md Nazirul Islam Sarker, Madallah Alruwaili, Yousef Alhwaiti, Nasser Alshammari, Muhammad Hameed Siddiqi, Boosting breast cancer detection using convolutional neural network, *J. Healthcare Eng.* 2021 (2021).
- [21] Muhammad Umer, Mahum Naveed, Fadwa Alrowais, Abid Ishaq, Abdullah Al Hejaili, Shtwai Alsubai, Ala' Abdulmajid Eshmawi, Abdullah Mohamed, and Imran Ashraf., Breast cancer detection using convoluted features and ensemble machine learning algorithm, *Cancers* 14 (23) (2022) 6015.
- [22] Muhammet Fatih Ak, A comparative analysis of breast cancer detection and diagnosis using data visualization and machine learning applications, *Healthcare* 8 (2) (2020) 111.
- [23] Yong Joon Suh, Jaewon Jung, Bum-Joo Cho, Automated breast cancer detection in digital mammograms of various densities via deep learning, *J. Personal. Med.* 10 (4) (2020) 211.
- [24] Jing Zheng, Denan Lin, Zhongjun Gao, Shuang Wang, Mingjie He, Jipeng Fan, Deep learning assisted efficient adaboost algorithm for breast cancer detection and early diagnosis, *IEEE Access* 8 (2020) 96946–96954.
- [25] Hanan Aljuaid, Nazik Alturki, Najah Alsubaie, Lucia Cavallaro, Antonio Liotta, Computer-aided diagnosis for breast cancer classification using deep neural networks and transfer learning, *Comput. Methods Prog. Biomed.* 223 (2022) 106951.

- [26] Manav Mangukiyi, Anuj Vaghani, Meet Savani, Breast cancer detection with machine learning, *Int. J. Res. Appl. Sci. Eng. Technol.* 10 (2) (2022) 141–145.
- [27] Xiaomei Wang, Ijaz Ahmad, Danish Javeed, Syeda Armana Zaidi, Fahad M. Alotaibi, Mohamed E. Ghoneim, Yousef Ibrahim Daradkeh, Junaid Asghar, Elsayed Tag Eldin, Intelligent hybrid deep learning model for breast cancer detection, *Electronics* 11 (17) (2022) 2767.
- [28] Liang Tang, Jianming Li, Ju Meihua Bao, Yiwei Chen Xiang, Yan Wang, Genetic association between her2 and esr2 polymorphisms and ovarian cancer: a meta-analysis, *Onco Targets Ther* (2018) 1055–1066.
- [29] Xiaoliang Xie, Xulin Wang, Yuebin Liang, Jingya Yang, Wu Yan, Li Li, Xin Sun, Pingping Bing, Binsheng He, Geng Tian, et al., Evaluating cancer-related biomarkers based on pathological images: a systematic review, *Front. Oncol.* 11 (2021) 763527.
- [30] Sihe Chen, Yongheng Chen, Yu Lanting, Hu. Xiangmei, Overexpression of socs4 inhibits proliferation and migration of cervical cancer cells by regulating jak1/stat3 signaling pathway, *Eur. J. Gynaecol. Oncol.* 42 (3) (2021) 554–560.
- [31] H. Karamti, R. Alharthi, M. Umer, H. Shaiba, A. Ishaq, N. Abuzinadah, S. Alsubai, I. Ashraf, Breast cancer detection employing stacked ensemble model with convolutional features, *Cancer Biomarkers* (2023) 1–16 (Preprint).
- [32] UCI Repository. Uci machine learning repository. URL <https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+%28Diagnostic%29>.
- [33] Leo Breiman, Bagging predictors, *Mach. Learn.* 24 (2) (1996) 123–140.
- [34] Gérard Biau, Erwan Scornet, A random forest guided tour, *Test* 25 (2) (2016) 197–227.
- [35] Mubariz Manzoor, Muhammad Umer, Saima Sadiq, Abid Ishaq, Saleem Ullah, Hamza Ahmad Madni, Carmen Bisogni, Rfcnn: traffic accident severity prediction based on decision level fusion of machine and deep learning model, *IEEE Access* 9 (2021) 128359–128371.
- [36] Afaq Juna, Muhammad Umer, Saima Sadiq, Hanen Karamti, Ala' Abdulmajid Eshmawi, Abdullah Mohamed, Imran Ashraf, Water quality prediction using knn imputer and multilayer perceptron, *Water* 14 (17) (2022) 2592.
- [37] Elham Besharati, Marjan Naderan, Ehsan Namjoo, Lr-hids: logistic regression host-based intrusion detection system for cloud environments, *J. Ambient. Intell. Humaniz. Comput.* 10 (9) (2019) 3669–3692.
- [38] Samina Sarwat, Naeem Ullah, Saima Sadiq, Robina Saleem, Muhammad Umer, Ala' Abdulmajid Eshmawi, Abdullah Mohamed, Imran Ashraf, Predicting students' academic performance with conditional generative adversarial network and deep svm, *Sensors* 22 (13) (2022) 4834.
- [39] Imran Ashraf, Manideep Narra, Muhammad Umer, Rizwan Majeed, Saima Sadiq, Fawad Javaid, Nouman Rasool, A deep learning-based smart framework for cyber-physical and satellite system security threats detection, *Electronics* 11 (4) (2022) 667.
- [40] Muhammad Umer, Saima Sadiq, Michele Nappi, Muhammad Usman Sana, Imran Ashraf, et al., Etcnn: extra tree and convolutional neural network-based ensemble model for covid-19 tweets sentiment classification, *Pattern Recogn. Lett.* 164 (2022) 224–231.
- [41] Rizwan Majeed, Nurul Azma Abdullah, Muhammad Faheem Mushtaq, Muhammad Umer, Michele Nappi, Intelligent cyber-security system for iot-aided drones using voting classifier, *Electronics* 10 (23) (2021) 2926.
- [42] Muhammad Umer, Saima Sadiq, Malik Muhammad Saad Missen, Zahid Hameed, Zahid Aslam, Muhammad Abubakar Siddique, Michele Nappi, Scientific papers citation analysis using textual features and smote resampling techniques, *Pattern Recogn. Lett.* 150 (2021) 250–257.
- [43] Lucia Cascone, Saima Sadiq, Saleem Ullah, Seyedali Mirjalili, Hafeez Ur Rehman Siddiqui, Muhammad Umer, Predicting household electric power consumption using multi-step time series with convolutional lstm, *Big Data Res.* 31 (2023) 100360.
- [44] Ahmad Hameed, Muhammad Umer, Umair Hafeez, Hassan Mustafa, Ahmed Sohaib, Muhammad Abubakar Siddique, Hamza Ahmad Madni, Skin lesion classification in dermoscopic images using stacked convolutional neural network, *J. Ambient. Intell. Humaniz. Comput.* (2021) 1–15.
- [45] Yuewu Hou, Zhaoying Liu, Ting Zhang, Yujian Li, C-unet: complement unet for remote sensing road extraction, *Sensors* 21 (6) (2021) 2153.
- [46] Parvathaneni Naga Srinivasu, Jalluri Gnana SivaSai, Muhammad Fazal Ijaz, Akash Kumar Bhoi, Wonjoon Kim, James Jin Kang, Classification of skin disease using deep learning neural networks with mobilenet v2 and lstm, *Sensors* 21 (8) (2021) 2852.
- [47] D. Lavanya, Dr K. Usha, Rani., Analysis of feature selection with classification: breast cancer datasets, *Indian J. Comput. Sci. Eng. (IJCSSE)* 2 (5) (2011) 756–763.
- [48] Sonam Jawahar Singh, Ramanathan Rajaraman, Tanmay Tulsidas Verlekar, Breast cancer prediction using auto-encoders, in: *International Conference on Data Management, Analytics & Innovation*, Springer, 2023, pp. 121–132.
- [49] O. Ibrahim Obaid, Mazin Abed Mohammed, M.K.A. Ghani, A. Mostafa, Fahad Taha, et al., Evaluating the performance of machine learning techniques in the classification of Wisconsin breast cancer, *Int. J. Eng. Technol.* 7 (4.36) (2018) 160–166.
- [50] Prithwish Ghosh. Breast cancer wisconsin (diagnostic) prediction.
- [51] Ravi Kumar Sachdeva, Priyanka Bathla, A machine learning-based framework for diagnosis of breast cancer, *International Journal of Software Innovation (IJSI)* 10 (1) (2022) 1–11.
- [52] Muhammet Fatih Ak, A comparative analysis of breast cancer detection and diagnosis using data visualization and machine learning applications, *Healthcare* 8 (2) (2020), <https://doi.org/10.3390/healthcare8020111>. ISSN 2227-9032. URL, <https://www.mdpi.com/2227-9032/8/2/111>.
- [53] Ashutosh Kumar Dubey, Umesh Gupta, Sonal Jain, Analysis of k-means clustering approach on the breast cancer Wisconsin dataset, *Int. J. Comput. Assist. Radiol. Surg.* 11 (11) (2016) 2033–2047.
- [54] Allison Murphy, Breast cancer Wisconsin (diagnostic) data analysis using gfs-tsk, in: *North American Fuzzy Information Processing Society Annual Conference*, Springer, 2021, pp. 302–308.
- [55] Sami Akbulut, Ipek Balikci Cicek, Cemil Colak, Classification of breast cancer on the strength of potential risk factors with boosting models: a public health informatics application, *Med. Bull. Haseki/Haseki Tip Bulteni* 60 (3) (2022).
- [56] Kreesheh Rajani. Breast cancer survival dataset. URL <https://www.kaggle.com/datasets/kreeshehrajani/breast-cancer-survival-dataset/code>.
- [57] Ankit Khushal Barai. Uci machine learning repository. URL <https://www.kaggle.com/datasets/ankitbarai507/breast-cancer-dataset/code>.
- [58] Muhammad Aurangzeb Ahmad, Carly Eckert, Ankur Teredesai, Interpretable machine learning in healthcare, in: *Proceedings of the 2018 ACM International Conference on Bioinformatics, Computational Biology, and health INFORMATICS*, 2018, pp. 559–560.
- [59] Scott M. Lundberg, Su-In Lee, A unified approach to interpreting model predictions, *Adv. Neural Inf. Proces. Syst.* 30 (2017).