Breast cancer detection employing stacked ensemble model with convolutional features

Hanen Karamti[a], Raed Alharthi[b], Muhammad Umer[c,*], Hadil Shaiba[a], Abid Ishaq[c],
Nihal Abuzinadah[d], Shtwai Alsubai[e] and Imran Ashraf[f,*]

[a]*Department of Computer Sciences, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia*
[b]*Department of Computer Science and Engineering, University of Hafr Al-Batin, Hafar, Saudi Arabia*
[c]*Department of Computer Science and Information Technology, The Islamia University of Bahawalpur, Bahawalpur, Pakistan*
[d]*Faculty of Computer Science and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia*
[e]*Department of Computer Science, College of Computer Engineering and Sciences, Prince Sattam Bin Abdulaziz University, Al-Kharj, Saudi Arabia*
[f]*Department of Information and Communication Engineering, Yeungnam University, Gyeongsan-si, Korea*

Abstract. Breast cancer is a major cause of female deaths, especially in underdeveloped countries. It can be treated if diagnosed early and chances of survival are high if treated appropriately and timely. For timely and accurate automated diagnosis, machine learning approaches tend to show better results than traditional methods, however, accuracy lacks the desired level. This study proposes the use of an ensemble model to provide accurate detection of breast cancer. The proposed model uses the random forest and support vector classifier along with automatic feature extraction using an optimized convolutional neural network (CNN). Extensive experiments are performed using the original, as well as, CNN-based features to analyze the performance of the deployed models. Experimental results involving the use of the Wisconsin dataset reveal that CNN-based features provide better results than the original features. It is observed that the proposed model achieves an accuracy of 99.99% for breast cancer detection. Performance comparison with existing state-of-the-art models is also carried out showing the superior performance of the proposed model.

Keywords: Breast cancer detection, image processing, healthcare, machine learning, ensemble learning, deep convoluted features

1. Introduction

Breast cancer is a prevalent and deadly disease, particularly for women in developing countries [1]. Breast cancer is a common form of cancer in women that is linked to denser breast tissue. It is ranked as the second most common cause of death for women globally [2], impacting 2.1 million individuals annually [3]. The

World Health Organization (WHO) reports that breast cancer affects more than 2.3 million women each year and causes 685000 deaths, comprising 13.6% of all cancer-related deaths in women [4]. Early detection is crucial in reducing the number of deaths from this disease. According to data from Globocan 2018 [5], one in four cancer cases in women is diagnosed as breast cancer, making it the fifth leading cause of death globally. Breast cancer usually originates in the breast tissue, specifically in the inner lining of milk ducts or lobules. The development of cancer cells is caused by mutations or modifications in the Deoxyribonucleic acid (DNA) or Ribonucleic acid (RNA). A variety of factors can contribute to mutations that may lead to breast can-

*Corresponding authors: Muhammad Umer, Department of Computer Science & Information Technology, The Islamia University of Bahawalpur, Bahawalpur, Pakistan. E-mail: umer.sabir@iub.edu.pk. Imran Ashraf, Information and Communication Engineering, Yeungnam University, Republic of Korea. E-mail: imranashraf@ynu.ac.kr.

cer including air pollutants, bacteria, nuclear radiation, fungi, mechanical cell-level injury, viruses, parasites, high temperatures, water contaminants, electromagnetic radiation, dietary factors, free radicals, DNA and RNA aging, and genetic evolution. Several kinds of breast cancer are found like inflammatory breast cancer (IBC) [6], Lobular breast cancer (LBC) [7], Invasive ductal carcinoma (IDC) [8], Mucinous breast cancer (MBC), Mixed tumor Breast cancer (MTBC), Ductal Carcinoma in situ (DCIS).

Breast cancer is a severe disease that carries a high risk of mortality. It accounts for 2.5% of all deaths, with one out of every thirty-nine women suffering from the disease [9]. Detecting and treating breast cancer early is essential because if left untreated, cancer can spread to other parts of the body. Early diagnosis and proper treatment can increase the survival rate by up to 80%. This emphasizes the significance of timely detection and prompt treatment of breast cancer. Several methods and techniques, such as screening tests, self-examinations, and regular visits to healthcare professionals can aid in the early diagnosis of breast cancer [10]. Mammography remains one of the most prevalent and effective techniques for detecting breast cancer in its early stages. Several studies have affirmed the efficacy of mammography in identifying breast cancer at an early stage. Another widely used technique for diagnosing breast cancer is a biopsy. In a biopsy, a tissue sample is collected from the affected area of the breast and examined under a microscope to detect and classify the tumor [11]. The biopsy is also considered a proficient method for breast cancer detection. Examination and analysis of breast cancer cells also help in this regard. Researchers performed nuclei analysis and cell classification to classify the cancerous cells into benign and malignant. While the available methods can help reduce the number of deaths from breast cancer, there is still room for improvement, particularly in terms of more efficient and automated diagnosis.

Data mining is a technique that can be used to extract useful and meaningful information from large amounts of data. It has been recognized as an important tool for the early diagnosis of various diseases such as heart disease [12], diabetes [13], kidney disease, and cancer. With the help of data mining techniques, patterns, and trends can be identified in the data which can help in the early diagnosis and treatment of these diseases. It is especially beneficial for detecting diseases such as cancer, where early detection can greatly increase the chances of survival. Basically, conventional cancer detection methods are comprised of three tests; physical examination, pathological test, and radiological images. All these conventional methods are time-consuming and are prone to false negatives. Aside from the traditional methods, machine learning methods are getting attention due to better results. Machine learning methods are reliable, accurate, and fast. These methods are extensively used in almost every kind of disease detection and produce better and more reliable results. Due to the aforementioned benefits, this study proposes a machine learning-based approach for detecting breast cancer to achieve high accuracy. This study makes the following contributions in this regard.

- A novel ensemble model is designed that uses a convolutional neural network (CNN) to extract features that are used for training. The ensemble model employs random forest (RF) and support vector machine (SVM) using voting to make the final prediction.
- Impact of convolutional features on prediction accuracy is analyzed by performing experiments with the original, as well as, the features extracted from the CNN model. For performance comparison, K-nearest neighbor (KNN), RF, logistic regression (LR), gradient boosting machine (GBM), Gaussian Naive Bayes (GNB), extra tree classifier (ETC), SVM, decision tree (DT) and stochastic gradient descent (SGD) are used.
- Performance of the proposed ensemble model is validated using k-fold cross-validation and comparing its performance with the state-of-the-art approaches. The results show that the proposed model can provide robust and generalizable performance.

The remaining sections of the present study are as follows. Section 2 contains the recent related works on breast cancer diagnosis and detection. The dataset, proposed methodology, and machine learning classifiers are explained in Section 3. Section 4 includes results and performs a comparative analysis. Discussions are presented in Section 5. Finally, Section 6 contains the conclusion and future work.

2. Related work

The early detection of breast cancer is crucial, and computer-aided diagnostics (CAD) plays an essential role in achieving this goal. In this field, various data mining techniques and machine learning algorithms have a significant impact. However, analyzing large and diverse healthcare datasets can be challenging in health

analytics. The latest advancements in CAD and AI offer accurate and precise solutions for medical applications while also handling sensitive medical data. Despite breast cancer being a leading cause of mortality in developed countries, machine learning is widely used in its detection. Recent research has focused on identifying malignancies, especially breast cancer, through CAD and decision support systems. Most studies use single models to obtain reliable results, while a few employ ensemble models. This section examines the latest and innovative breast cancer detection systems that utilize machine learning methods.

For the accurate and precise diagnosis of breast cancer Yadav and Jadhav [14] proposed a machine learning-based system that uses thermal infrared imaging. The authors used several baseline models and transfer learning models like VGG16 and InceptionV3. The authors performed experiments involving data augmentation and without augmentation. Results of the study show that the transfer learning model InceptionV3 outperforms other learning models and achieves an accuracy score of 93.1% without augmentation and 98.5% with augmentation. In another study [15], the authors utilized the genetic programming technique to select the optimal features for automated breast cancer diagnosis. The authors tested nine machine learning classifiers including RF, LR, SVM, DT, AdaBoost (AB), GNB, Latent Dirichlet Allocation (LDA), KNN, and GB. The results demonstrate that genetic programming effectively identifies the best model by merging the preprocessing and models' features. The highest accuracy score of 98.23% is attained using the AB classifier.

Alanazi et al. [16] proposed an automated system for breast cancer detection using deep earning. They also utilized machine learning models including LR, KNN, SVM, and various CNN variants. In experiments, the authors examined the hostile ductal carcinoma tissue zones in the whole slide image. The study's findings reveal that the CNN variant obtained the highest accuracy of 87%, surpassing the machine learning models' accuracy by 9%. It indicates that the proposed deep learning-based system enhances accuracy in breast cancer detection. Umer et al. [17] introduced an ensemble learning-based voting classifier for detecting breast cancer. The study incorporated various machine learning models such as RF, KNN, DT, SVM, LR, and GBM alongside the proposed ensemble learning model. The findings showed that the proposed ensemble learning model achieved better results than machine learning models. For the detection of breast tumor types, the study [18] proposed a machine learning-based system

that achieves an accuracy of 98.1%. Suh et al. [19] used various density mammograms for breast cancer detection. They achieved an overall accuracy score of 88.1%.

In addition to machine learning models, transfer learning models are also developed and utilized for breast cancer classification. From the different imagining techniques such as magnetic resonance imaging (MRI), ultrasound, and mammography, the CNN-based transfer learning model is used in [20]. DLA-EABA is used for the classification of breast masses. The work mainly focuses on the ensemble of the machine learning approaches with the different feature extraction techniques and evaluating the output using segmentation and classification techniques. Results depict that the proposed DLA-EABA achieved an accuracy score of 97.2%. A transfer learning-based approach is proposed by Aljuaid et al. in [21] for breast cancer classification. The authors experimented with two ways; binary classification and multi-class classification. They used the transfer learning models such as ResNet18, ShuffleNet, and InceptionV3. For the binary class classification, ResNet18 achieved the highest accuracy of 99.7% while for the multi-class classification, ResNet18 achieved an accuracy score of 97.81%.

Mangukiya et al. [22] conducted a study that explored several techniques for achieving efficient, early, and accurate breast cancer diagnosis. The authors utilized various machine learning algorithms such as RF, DT, SVM, KNN, XGBoost, NB, and AB. The dataset used in the study includes features with highly varied units and magnitudes. To standardize all the features' magnitudes, they employed standard scaling. The findings demonstrate that the XGBoost machine learning algorithm attains an accuracy score of 98.24% with standard scaling. In the same way, [23] presented a deep ensemble learning model for detecting breast cancer using the whole slide image. They utilized various deep learning models such as CNN, deep neural network (DNN), long short-term memory (LSTM), gated recurrent unit (GRU), and Bidirectional LSTM (BiLSTM) and proposed the ensemble model CNN-GRU. Results reveal that the hybrid deep learning model CNN-GRU outperforms other learning models and achieved an accuracy score of 86.21%.

While the above-discussed studies utilize different machine and deep learning models for disease diagnosis, several studies focus only on using CNN models for the same purpose. For example, [24] employs the CNN model for mycobacterium tuberculosis detection from bright-field microscopy. The proposed system is a

Table 1
Summary of the discussed research works

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

computer-aided diagnosis system involving the use of image processing and deep learning that provides better disease detection accuracy than existing approaches. Similarly, the study [25] investigates the performance of various ensemble models regarding the prediction of tuberculosis using chest X-rays. The authors use the U-Net model for regions of interest from chest X-rays which are later used with deep learning models. Different variants of CNN are implemented in the study; the best results are obtained by the proposed stacked ensemble with a 98.38% accuracy. In the same vein, several other works deploy customized CNN models for disease detection. For example, [26] uses CNN for bleeding image detection, [27] uses CNN for pneumonia classification, and [28] uses CNN for cardiovascular disease prediction.

Several studies have been conducted to detect breast cancer using machine learning models, to improve classification performance and reduce pathological errors in automatic diagnosis. Table 1 summarizes some of the literature on breast cancer detection using machine learning models.

3. Materials and methods

The dataset used for the detection of breast cancer, the proposed approach, and the steps taken for the proposed methodology are discussed in this section. This section also presents a brief description of the machine learning classifiers used in this study.

3.1. Dataset for experiments

In this study, supervised machine learning models are utilized for breast cancer detection, with a focus on evaluating their performance. The study follows a series of steps, starting with the collection of dataset [29]. In this study, the "Breast Cancer Wisconsin Dataset" is obtained from the UCI machine learning repository, which is publicly accessible. 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). The dataset consists of two target classes, namely benign and malignant. The distribution of the samples shows that 45% of the data belong to the malignant class, while 55% are from the benign class. The 32 features in the dataset are classified into different types such as numeric, nominal, binary, etc. It is important to note that the target class is categorical, while the remaining attributes are numeric.

3.2. Data preprocessing

This study performs two steps in data preprocessing to improve the training process of machine and deep

Table 2
Dataset description

Feature name	Description
ID	Unique identification number assigned to each sample
Diagnosis	Whether the sample is benign (B) or malignant (M)
Radius mean	Mean of distances from center to points on the perimeter
Texture mean	Standard deviation of gray-scale values
Perimeter mean	Mean size of the core tumor
Area mean	Mean size of the area occupied by the tumor
Smoothness mean	Mean of local variation in radius lengths
Compactness mean	Mean of perimeterˆ2 / area – 1.0
Concavity mean	Mean severity of concave portions of the contour
Concave points mean	Mean number of concave portions of the contour
Symmetry mean	Mean symmetry of the tumor
Fractal dimension mean	Mean "coastline approximation" – 1
Radius SE	Standard error of distances from center to points on the perimeter
Texture SE	Standard error of gray-scale values
Perimeter SE	Standard error of the size of the core tumor
Area SE	Standard error of the size of the area occupied by the tumor
Smoothness SE	Standard error of local variation in radius lengths
Compactness SE	Standard error of perimeterˆ2/area – 1.0
Concavity SE	Standard error of severity of concave portions of the contour
Concave points SE	Standard error for number of concave portions of the contour
Symmetry SE	Standard error for symmetry of the tumor
Fractal dimension SE	Standard error for "coastline approximation" – 1
Radius worst	"Worst" or largest mean value for distances from center to points on the perimeter
Texture worst	"Worst" or largest value for standard deviation of gray-scale values
Perimeter worst	"Worst" or largest value for the size of the core tumor
Area worst	"Worst" or largest value for the size of the area occupied by the tumor
Smoothness worst	"Worst" or largest value for local variation in radius lengths
Compactness worst	"Worst" or largest value for perimeterˆ2/area – 1.0
Concavity worst	"Worst" or largest value for severity of concave portions of the contour
Concave points worst	"Worst" or largest value for number of concave portions of the contour
Symmetry worst	"Worst" or largest value for symmetry of the tumor
Fractal dimension worst	"Worst" or largest value for "coastline approximation" – 1

learning models. The missing values in the data may lead to bias. Deleting missing values can help avoid errors and reduce the probability of bias. However, if the number of records containing missing values is high, it may distort relationships between various attributes. In our case, the number of missing values is not high and they can be deleted to avoid error and bias. In addition, label encoding is also performed as the dataset contains categorical values. For training machine learning models, converting categorical data into numerical data is essential.

3.3. Machine learning models for breast cancer prediction

Machine learning classification is a supervised learning method where the system learns from a specific dataset and uses that knowledge to classify new observations. The dataset can be binary or multi-class. In this section, we discuss machine learning classifiers for breast cancer detection. The sci-kit-learn library is used to implement the machine learning models. All models

are implemented in the Python environment using the sci-kit module.

3.3.1. Random forest

RF is a widely used ensemble learning approach for classification and regression problems in machine learning [30,31]. It is a decision tree combination method in which several decision trees are generated and their outputs are merged to form the final prediction. The fundamental concept behind this technique is to train numerous decision trees, each on a unique subset of the data, and then combine their predictions to create the final prediction. This approach helps to reduce the overfitting problem that can arise when training a single decision tree. Mathematically, the random forest can be represented as

$$p = \text{mode}\{T_1(y), T_2(y), T_3(y), \ldots, T_m(y)\} \quad (1)$$

$$p = \text{mode}\left\{\sum_{m=1}^{m} T_m(y)\right\} \quad (2)$$

where p is final prediction,and $T_1(y), T_2(y), \ldots T_m(y)$ are the decision trees taking part in the prediction process.

3.3.2. Decision tree

Currently, the DT is one of the most widely used techniques for classification and prediction [32]. A DT is presented as a tree-like structure, similar to a flowchart, that displays logical steps. In this structure, an internal node signifies an attribute test, a branch represents the result of an attribute test, and a leaf node indicates a class label. Decision trees are highly beneficial in data classification as they can accomplish it in a short period with minimum computational resources. These trees can process both categorical and continuous data. Furthermore, decision trees can identify the essential data points that are required for accurate classification and forecasting.

3.3.3. K-nearest neighbour

The k-NN algorithm is a non-parametric approach in machine learning and is used for both regression and classification tasks. This algorithm uses lazy learning or instance-based learning, where it identifies the k number of closest training instances to a new data point and determines the majority class among those k nearest neighbors to classify the new data point [33]. The algorithm is based on the concept of similarity between the input data and training data, where it stores all available cases and uses a similarity measure, such as the distance function, to classify new cases. The k-NN algorithm is simple and easy to implement.

In the field of pattern recognition, k-NN is frequently employed for classification issues and has been used for tasks such as medical diagnosis, image recognition, and video recognition. One of the primary benefits of k-NN is its simplicity and versatility in handling both regression and classification tasks. However, it is vulnerable to the scale of the data and extraneous features, and the optimal value of k must be chosen with care.

3.3.4. Logistic regression

LR is a statistical model used for binary classification problems in supervised learning. It is commonly used when the outcome variable is binary, such as predicting whether a patient has a disease or not, or whether an email is spam or not. LR is used to estimate the probability of a binary outcome based on certain inputs, and then use that estimate to make a prediction. The logistic function (also called the sigmoid function) is used to model the probability of a binary outcome, and the out-

put of the logistic function is then used to make a prediction [34,30]. The logistic function or sigmoid function is commonly 'S' shaped curve as in the equation below

$$f(x) = \frac{L}{1 + e^{-m(v-v_o)}} \tag{3}$$

LR can be used for binary classification problems, as well as multi-class classification problems (when more than two classes are present) using one-vs-all or softmax regression.

3.3.5. Support vector machine

SVM is a well-known supervised learning algorithm [35] used for classification and regression problems in machine learning. SVM's main principle is to determine the optimal boundary (or hyperplane) that divides data points into different classes. The border is designed to maximize the margin, which is the distance between the boundary and the nearest data points from each class, also known as support vectors. SVM is suitable for both linear and non-linear classification tasks. A linear border (or hyperplane) is used to separate the data points in the case of linear classification. In the case of non-linear classification, a technique known as the kernel trick is employed to convert the input data into a higher dimensional space with a linear border to separate the data points. SVM is also effective in cases where there is a clear margin of separation in the data. However, it can be less effective when the data is noisy or when the classes are highly overlapping.

3.3.6. Gradient boosting machine

GBM is a machine learning algorithm used for both classification and regression problems, and it is part of the ensemble learning family called boosting [36]. GBM combines the predictions of multiple weak models, such as decision trees, to create a strong model. The idea behind gradient boosting is to iteratively train weak models, such as decision trees, and add them to the ensemble one at a time. New trees are trained to correct the mistakes of the previous trees by focusing on the training instances that were misclassified. The predictions of all trees are then combined to make the final prediction. This process is repeated until a predetermined number of trees is reached or the performance of the ensemble on a validation set stops improving. GBM has many advantages such as being able to handle a wide range of data types like categorical and numerical features and modeling non-linear interactions between features and the target. Additionally, it often performs well on large datasets with a large number of features and instances.

3.3.7. Extra tree classifier

ETC is an ensemble learning method that uses randomized trees [37] to generate a final classification output by combining uncorrelated trees in a forest of decision trees. The underlying concept of ETC is similar to RF but the method of constructing decision trees in the forest is different. In ETC, for the decision making some random samples of the K best features are used, and the optimal solution is found using the Gini index. This method gives the development of the uncorrelated tree in the ETC. Gini feature importance plays a vital role in the feature selection.

3.3.8. Gaussian naive Bayes

GNB is a popular machine learning algorithm used for classification tasks that are based on the Bayes Theorem. According to this theorem, the probability of a hypothesis (class label) given some evidence (feature values) is equal to the probability of the evidence given the hypothesis multiplied by the hypothesis's prior probability [38]. The 'naive' component of the term refers to the algorithm's strong assumption, known as class conditional independence, which stipulates that all features are independent given the class label. This assumption is rarely true in real-world situations but it still performs well in practice.

The GNB is used for continuous data, specifically for normally distributed data, it estimates the probability density function of each feature for each class, assuming a Gaussian distribution. It is a fast and simple algorithm that is easy to implement, and it does not require a lot of memory. It also works well with high-dimensional data, making it a good choice for text classification and sentiment analysis. However, it can perform poorly when there are a lot of irrelevant features or when the features are highly correlated.

3.3.9. Stochastic gradient decent

SGD is an optimization algorithm used to minimize a function, particularly for training models in machine learning such as linear regression, logistic regression, and neural networks [39]. It is a variant of the gradient descent (GD) algorithm and is called stochastic because it uses a random sample of the data, called a mini-batch, to estimate the gradient at each iteration. The main of the SGD algorithm is to update the parameters of the model in the opposite direction of the gradient of the loss function with respect to the parameters, with a fixed step size, called the learning rate.

The advantage of SGD is that it is computationally efficient and can handle large datasets, as it only uses a small subset of the data (mini-batch) at each iteration. Additionally, it can converge to a good solution even with a noisy or non-convex loss function. However, the solution found by SGD is sensitive to the choice of the learning rate, and it can converge to a local minimum or even oscillate around the optimal solution.

3.4. Deep learning models for breast cancer prediction

An expanding area of research in the field of artificial intelligence is deep learning. The modeling of data in deep learning gives promising results. The adoption of an automated process by medical professionals has shown to be a highly useful and successful tool for disease diagnosis. Deep learning is a common method for processing enormous amounts of data. It eliminates the need for manual feature extraction, it is being employed widely in medical data analysis.

3.4.1. Multilayer perceptron neural network

When we are talking about not large-sized training sets, easy implementation, speed, and quick results Multi-Layer Perceptron is the best choice [35]. The internal structure of MLP comprises three layers, input, output, and hidden layers. The hidden layer is an intermediate layer to connect the input layer with the output layer during neuron processing. The internal working of MLP is simply based on the multiplication of input neurons with weights w_{ij} and output y_j is the sum. Mathematically, it is computed as:

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

In this equation, the gradient descent algorithm is assigned weights w and O represents hidden layers.

3.5. RNN

When we are talking about sequential neural networks Recurrent Neural Network (RNN) is the best choice [40]. During processing, the input sequence of one neuron is fed to other neurons in the same weighted sequence of words in a sentence. RNN sequences are designed in a manner that generates the sequence and predicts the next word coming in the loop.

3.5.1. Convolutional neural network

CNN is an effective neural network model that can learn complex relations among different data attributes. A CNN is a deep learning model that can analyze the input image, rank various features and objects within

the image, and distinguish between them. CNN is made of a hidden layer, node layer, input, and output layer. To obtain better results, this study uses a customized CNN architecture [41]. The proposed 8-layer architecture includes 2 dense layers, 2 max-pooling layers, and 2 convolution layers. For classification purposes in the medical field, CNN performance is the best and most accurate. In the CNN model, the Sigmoid is used as the error function and it is a backpropagation algorithm. CNN has been used for the classification of multiple diseases i.e. brain tumors, lung disease, and cardiac disease. Nowadays, it is extensively used in the medical field and deals with large amounts of data. The pooling layer in CNN can be maximum and average pooling, maximum pooling is mostly used for sharp feature extraction while the average is used for flat feature extraction.

3.6. Long short term memory

An improved RNN called LSTM is more operative for long-term sequences [40]. LSTM overcame the vanishing gradient issue that RNN faces. It outperforms RNN and can memorize certain patterns. The input gate, output gate, and forget gate are the three gates that make up an LSTM. The word sequence is shown in Eqs (4) to (6).

$$i_t = \sigma(x_t U^i + h_{t-1} W^i + b_i) \tag{4}$$

$$o_t = \sigma(x_t U^o + h_{t-1} W^o + b_o) \tag{5}$$

$$f_t = \sigma(x_t U^f + h_{t-1} W^f + b_f), \tag{6}$$

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.

3.6.1. Architecture of convolutional neural network for feature extraction

In this study for breast cancer detection, the deep learning model CNN is used as a feature extraction technique [41]. CNN is a widely used deep learning system mostly used for classification tasks. As a deep learning system can extract features, the convoluted features are used for breast cancer detection. There are four layers in the conventional CNN model including the pooling layer, embedding layer, convolutional layer, and flattening layer. For breast cancer detection, the first layer of CNN used is an embedding layer and it has an embedding size of 20,000 and an output dimension of 300. The second layer is the convolutional layer which has 5000 filters, a kernel size of 2×2, and a rectified

linear unit (ReLU) as an activation function. The third layer is the max pooling layer; for the significant feature maps max pooling layer with 2×2 sizes is used from the output of the convolutional layer. In the end, a flatten layer is used to convert the output into a 1D array for the learning models.

For example, a tuple set (fs_i, tc_i) is from the breast cancer dataset, where the fs is the feature set, tc is the target class column, and I shows the index of the tuple. For the transformation of the training set into the required input, the embedding layer is employed as

$$EL = embedding\_layer(Vs, Os, I) \tag{7}$$

$$EOs = EL(fs) \tag{8}$$

where EL denotes the embedding layer and EO_s shows the embedding layers output. This output is the input of the conventional layer. There are three different parameters for the EL: V_s vocabulary size, I input length and Os is the dimension of the output.

In this study for breast cancer detection, the EL size is set at 20,000. It means that the EL can take the inputs from 0 to 20000. The input length is 32 and the output dimension Os is set to 300. EL processes all the input data and gives the output for the CNN for further processing. EL output dimension is $EO_s = $ (None, 32, 300)

$$1D - Convs = CNN(F, Ks, AF) \leftarrow EOs \tag{9}$$

The convolutional layer output is extracted from the EL output. CNN is implemented with the 500 filters, i.e., $F = 500$ and a kernel size of 2×2. The ReLU activation function is used for setting all negative values to zero and all the other values remain unchanged.

$$f(x) = max(0, E)s \tag{10}$$

For the significant feature extraction, the map max pooling layer is used. For this purpose, a 2×2 pool is used. *Fmap* shows the features after max-pooling, $Ps = 2$ is the size of the pooling window and S-2 is the size of the stride. In the end, the flattened layer is used for the data transformation. By using the above-mentioned steps we obtained the 25000 features for the training of the machine learning models.

$$Cf = Fmap = \lfloor (1 - P_s)/S \rfloor + 1 \tag{11}$$

To convert the 3D data into 1D, a flattened layer is used. The main reason behind this conversion is that the machine learning models work well on the 1D data. For the training of the models, the above-mentioned step is implemented for the training. The architecture of the used CNN along with the predictive model is shown in Fig. 1.
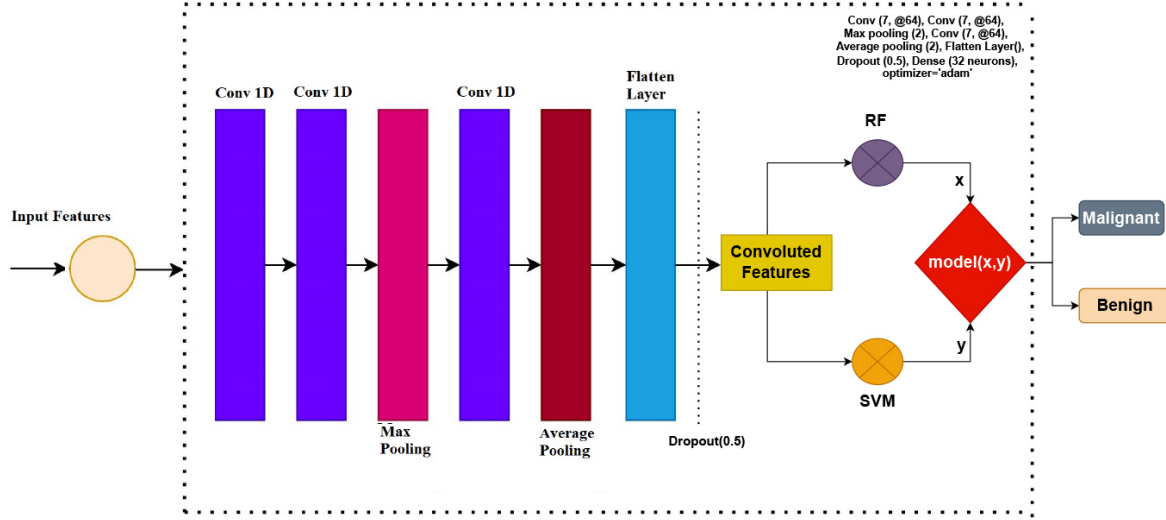
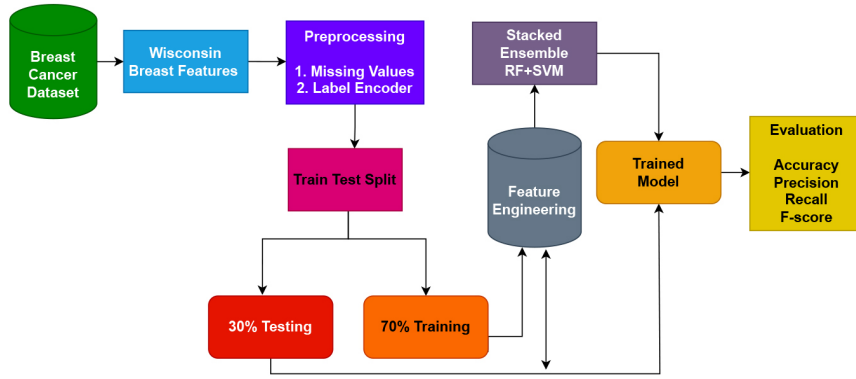Fig. 1. Architecture diagram of the CNN with voting classifier (LR+SGD) model.

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

3.7. Proposed methodology

Ensemble models are becoming more prevalent and have led to greater accuracy and efficiency for classification tasks. By merging multiple classifiers, it is possible to enhance the performance beyond what individual models can achieve. In this study, an ensemble learning approach is employed to enhance breast cancer detection. The proposed method involves a voting classifier that unites the RF and SVM through the soft voting criterion.

The ultimate output is determined by the class that receives the most votes. The proposed ensemble model, as outlined in Algorithm 1, operates as follows:

$$\widehat{p} = argmax \sum_i^n RF_i, \sum_i^n SVM_i \qquad (12)$$

Algorithm 1: Ensembling RF and SVM.

Input: input data $(x, y)_{i=1}^N$
$M_{RF} = $ Trained RF
$M_{SVM} = $ Trained SVM

for $i = 1$ to M **do**
  if $M_{RF} \neq 0 \& M_{SVM} \neq 0 \& training\_set \neq 0$ **then**
    $P_{SVM_1} = M_{SVM}.probability(class1))$
    $P_{SVM_2} = M_{SVM}.probability(class2))$
    $P_{RF_1} = M_{RF}.probability(class1))$
    $P_{RF_2} = M_{RF}.probability(class2))$
    Decision function $= \max \left(\frac{1}{n} \sum_{classifier} \right.$
    $\left. (Avg(P_{SVM_1}, P_{RF_1}), Avg(P_{SVM_2}, P_{RF_2}) \right)$

  end if
  return final label $\widehat{p}$
end for

The prediction probabilities for each test sample are provided by $\sum_i^n LR_i$ and $\sum_i^n SGD_i$. These probabili-
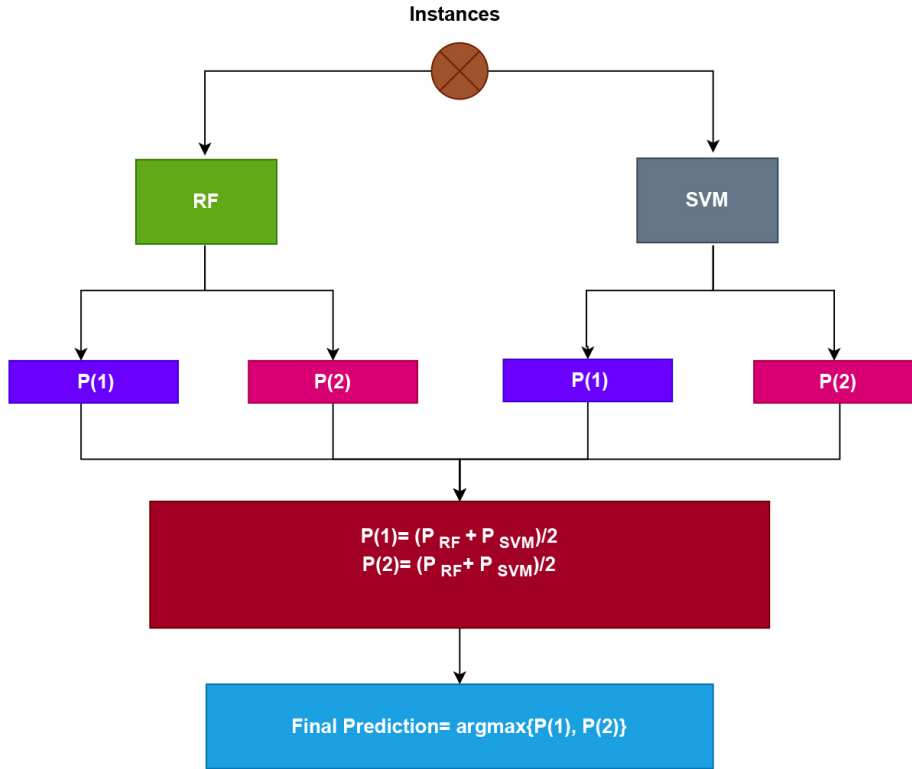
Instances

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

ties, as illustrated in Fig. 2, pass through the soft voting criterion, which yields the probabilities for each test case using the RF and SVM. The voting process is illustrated in Fig. 3.

$$VC(RF+SVM) = argmax(g(x)) \qquad (13)$$

To evaluate the proposed model VC(RF+SVM), it is tested on the 'Breast Cancer Wisconsin Dataset' in two stages. In the first stage, breast cancer is detected using all 32 features of the dataset. In the second stage of the experiments, the dataset is processed for machine-learning models using convolutional features. The data is divided into two parts, with 70% allocated for training and 30% reserved for testing. This approach, known as the training-testing split, is a common method in machine learning used to assess the accuracy of the model on new and unseen data.

3.8. Experiment set up

The experiments are conducted using a Python 3.8 programming environment. The study's experimental environment includes the software libraries (Scikit-learn and TensorFlow), programming language (Python

3.8), available RAM (8GB), operating system (64-bit Windows 10), CPU (Intel Core i7, 7th Gen, 2.8 GHz processor), and GPU (Nvidia GTX 1060 with 8 GB memory). This information is essential for understanding the technical specifications of the experimental setup and the computational resources used in the research.

3.9. Evaluation metrics

The performance of the machine learning models used in this study is measured in terms of accuracy, precision, recall, and F1 score. All these metrics are based on the values from the confusion matrix. These matrices have a minimum value of 0 and a maximum value of 1.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \qquad (14)$$

$$Precision = \frac{TP}{TP + FP} \qquad (15)$$

$$Recall = \frac{TP}{TP + FN} \qquad (16)$$

$$F1\ score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \qquad (17)$$

Table 3
Hyperparameter details of all classifiers

Classifier	Hyperparameter
LR	C = 10, class_weight='balanced', l1_ratio = 0.7, max_iter = 3000, penalty = 'elasticnet', solver = 'saga'
SVM	C = 300, class_weight = 'balanced'
RF	n_estimators = 300, criterion = 'entropy', max_depth = 30,
DT	criterion= 'entropy', max_depth = 30,
ETC	n_estimators = 300, max_depth = 30, criterion = 'entropy'
SGD	Larning_rate='optimal', epsilon = 0.2
GBM	n_estimators = 300, learning_rate = 0.2, max_depth = 30,
KNN	n_neighbors = 5, leaf_size = 35
GNB	var_smoothing = 1e-9, n_classes = default
VC	criteria='soft', n_jobs = −1
CNN	Stride = (1 × 1), pool size= (@ 2), filter= (@ 256), Dense neuron (60), activation ='Relu'

Table 4
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 5
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

4. Results

For breast cancer detection extensive experiments are carried out. Machine learning models are applied using the original features, as well as, the convoluted features. Hyperparameter tuning values of the models are presented in Table 3. Results are investigated and an ensemble of the top four individual machine learning models is also used in the experiments on both feature sets.

4.1. Results of individual machine learning models on original features and convoluted features

The present study uses nine machine learning models with optimized hyperparameters to obtain better results. To attain high accuracy, these parameters are set empirically. RF, for example, performs the best when it works with the original features. RF attains an accuracy score of 91%, followed by the ETC which achieves an accuracy score of 89%. The k-NN is the least performer and it achieves an accuracy score of 81%. The accuracy score of all the classifiers when used with original features is displayed in Table 4.

Table 5 shows the classification accuracy of the machine learning models when used with convoluted features. Experimental results depict that the RF and ETC outperform other models and achieved accuracy scores

of 93.75% and 93.74%, respectively. Similarly, the SVM and LR give a higher accuracy score than the other classifiers.

4.2. Performance of ensemble models using original features

At first, the individual models are applied to the original features and convoluted features and the results of the models are shown in Tables 4 and Table 5. Out of 9 machine learning models four models RF, ETC, LR, and SVM achieve the best results on both feature sets. In this part of the experiments, the ensembles of these machine learning models are tested on the original features. Results of the ensemble learning models show that the proposed ensemble model RF+SVM outperforms other models in terms of accuracy which is 95%; approximately 2% higher among all the ensemble learning models. It is followed by the SVM+ETC which achieves an accuracy score of 92%. However, the RF+SVM achieves 95% precision, 98% recall, and 96% F1 scores for breast cancer detection. The results of the ensemble learning models on the original feature set are shown in Table 6.

4.3. Performance of ensemble model on convoluted features

The ensemble models are also tested using the fea-

Table 6
Ensemble model results using original features set

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

Table 7
Ensemble model results using convoluted features set

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

Table 8
Results for k-fold cross-validation of the proposed ensemble model

Fold number	Accuracy	Precision	Recall	F-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

tures extracted by the customized CNN model and experimental results are given in Table 7. Results show that the proposed RF+SVM surpasses other models with 99% accuracy and 99% each for precision, recall, and F1 score. ETC+LR has shown the lowest results with a 94% accuracy. Ensemble learning model results are better when used with the features from the CNN model as compared to using the original features.

4.4. Results of k-fold cross-validation

K-fold cross-validation is also performed to verify the performance of the proposed model. Cross-validation aims at validating the results from the proposed model and verifying its robustness. Cross-validation is performed to analyze if the model performs well on all the sub-sets of the data. This study makes use of 5-fold cross-validation and results are given in Table 8. Cross-validation results reveal that the proposed ensemble model provides an average accuracy score of 0.996 while the average scores for precision, recall, and F1 are 0.998, 0.998, and 0.997, respectively.

4.5. Performance comparison with existing studies

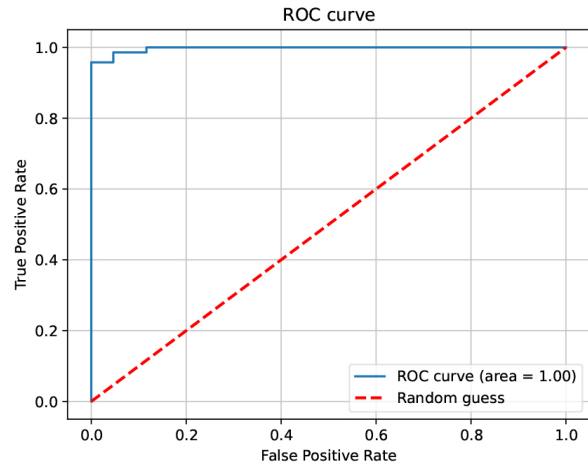In order to show the performance of the proposed

Fig. 4. ROC-AUC of the proposed model.

model over previous state-of-the-art models, results are compared with existing models. For this purpose, this research work selects the 9 most related research works. For instance, [43] used the PCA features with the machine learning model SVM for breast cancer detection and achieved an accuracy score of 96.99%. The study [45] used the autoencoder and achieved the highest accuracy score of 98.40%. Quadratic SVM is used by the [44] thereby reporting an accuracy score of 98.11%. For the same task, [47] used the XgBoost and achieved an accuracy score of 97.11%. In a similar fashion, [49, 50] used the Chi-square features and machine learning model LR with 98.21% and 98.10% accuracy scores, respectively. Table 9 shows the performance comparison between the proposed and existing studies. Results exhibit a better performance of the proposed model.

5. Discussion

The results presented in the study are focused on evaluating the performance of various machine learning models on both original and convoluted features, as well as the effectiveness of ensemble models. The dataset appears to be related to breast cancer detection, and the goal is to achieve high accuracy and other relevant metrics such as precision, recall, and F1 score. Figure 4 presents the AUC-ROC (Area Under the Receiver Operating Characteristic Curve) curve of the proposed approach. The AUC-ROC curve is both a visual representation and an important performance measure for models designed for binary classification tasks. This curve provides insights into a model's capacity to distinguish between two classes. The curve's shape, proximity to the top-left corner, and the AUC value indicate the

Table 9
Performance comparison with state-of-the-art studies

Ref.	Technique	Accuracy
[42]	K-means clustering	92.01%
[43]	PCA features with SVM	96.99%
[44]	Quadratic SVM	98.11%
[45]	Auto-encoder	98.40%
[46]	GF-TSK	94.11%
[47]	XgBoost	97.11%
[48]	Five most significant features with LightGBM	95.03%
[49]	Chi-square features	98.21%
[50]	LR with all features	98.10%
Proposed	Deep convoluted features with voting classifier (RF+SVM)	99.99%

Table 10
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

model's discriminatory ability and overall performance. It's a valuable tool for comparing models, selecting classification thresholds, and assessing model robustness. Figure 4 shows the higher AUC values, which are associated with better classification performance. The ROC-AUC curve shown in Fig. 4 indicates the superior performance of the proposed ensemble model for breast cancer detection.

However, the use of original features achieved slightly lower accuracy compared to convoluted features, which could be indicative of the potential of feature engineering or extraction methods to improve model performance. Ensemble models were also tested using features extracted from a customized Convolutional Neural Network (CNN) model. The results showed that RF+SVM outperformed all other models with an impressive accuracy of 99.99%. This highlights the significance of convolutional features for breast cancer detection, potentially indicating the importance of image analysis in this context. These findings could be valuable in the context of medical image analysis and disease detection, emphasizing the importance of feature engineering, model selection, and ensemble methods in improving the performance of machine learning systems.

To prove the effectiveness of the proposed approach experiments are performed on three deep learning models (MLP, RNN, and LSTM) and two other datasets. RNN and LSTM are versatile neural network architectures that have found applications beyond language processing. In this study, their inclusion might be motivated by their ability to model sequential dependen-

cies and capture temporal patterns in data. While MLP, CNN, and LSTM have been effectively employed in a wide range of applications, including medical diagnosis [51], medical image analysis [52], and breast cancer diagnosis [53,54].

5.1. Performance of deep learning models using original features

Deep learning models are applied to the original features and convoluted features and the results of the models are shown in Table 10. Out of 3 deep learning models CNN achieved the best results on both feature sets. In this part of the experiments, the significance of the proposed model is validated with state-of-the-art deep learning models. Still, the proposed ensemble model beats the deep learning models in terms of accuracy. The accuracy of MLP is reduced using CNN features while LSTM accuracy is improved because it gets more significant features to generate sequences. The accuracy of CNN remains almost the same because it receives the same convoluted features and an extra layer to make predictions.

5.2. Significance of proposed model

In order to validate the performance of the proposed model, we tested it on two further independent datasets. The first dataset [55] 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 [56] 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 (μU/mL), HOMA, Leptin (ng/mL), Adiponectin (μg/mL), Resistin (ng/mL), MCP-1(pg/dL), and Labels (absence or presence). The proposed model obtained 97.34% accuracy on the first dataset and 96.67% accuracy on the second dataset which greatly shows the stability of the proposed model on all kinds of datasets.

6. Conclusions

The goal of this study is to provide a framework that accurately classifies benign and malignant breast cancer patients and lowers the risk associated with this leading cause of death in women. For this purpose, an ensemble model is proposed owing to the reported superior performance of ensemble models in the existing literature. However, instead of manual feature extraction, the features from a customized CNN model are used for training. The proposed model classifies cancerous patients from normal ones with an accuracy of 99.99%. In addition, models tend to yield superior results when used with CNN-based features. K-fold cross-validation and performance comparison with existing state-of-the-art models also prove the effectiveness and robustness of the proposed model. In the future, we intend to apply this model on multi-domain datasets like breast cancerous images and microscopic feature numeric values obtained from those images.

Funding

Author contributions

Conception: Hanen Karamti, Muhammad Umer, and Hadil Shaiba.

Interpretation or analysis of data: Muhammad Umer, Abid Ishaq, Nihal Abuzinadah.

Preparation of the manuscript: Hanen Karamti, Muhammad Umer, and Imran Ashraf.

Revision for important intellectual content: Shtwai Alsubai, Raed Alharthi, and Hadil Shaiba.

Supervision: Hanen Karamti, Raed Alharthi, Shtwai Alsubai, and Imran Ashraf.

References

[1] WHO, World Health Organization. Cancer: Key Facts. 2022, July 2022, Online; accessed 10 January 2023.

[2] Y.-S. Sun, Z. Zhao, Z.-N. Yang, F. Xu, H.-J. Lu, Z.-Y. Zhu, W. Shi, J. Jiang, P.-P. Yao and H.-P. Zhu, Risk factors and preventions of breast cancer, *International Journal of Biological Sciences* **13**(11) (2017), 1387.

[3] WHO, World Health Organization. Breast Cancer, July 2022, Online; accessed 10 January 2023.

[4] WHO, World Health Organization. Breast Cancer, March 2021, Online; accessed 10 January 2023.

[5] F. Bray, J. Ferlay, I. Soerjomataram, R.L. Siegel, L.A. Torre, A. 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] F.M. Robertson, M. Bondy, W. Yang, H. Yamauchi, S. Wiggins, S. Kamrudin, S. Krishnamurthy, H. Le-Petross, L. Bidaut, A.N. Player et al., Inflammatory breast cancer: the disease, the biology, the treatment, *CA: A Cancer Journal for Clinicians* **60**(6) (2010), 351–375.

[7] S. Masciari, N. Larsson, J. Senz, N. Boyd, P. Kaurah, M.J. Kandel, L.N. Harris, H.C. Pinheiro, A. Troussard, P. Miron et al., Germline E-cadherin mutations in familial lobular breast cancer, *Journal of Medical Genetics* **44**(11) (2007), 726–731.

[8] A.R. Chaudhury, R. Iyer, K.K. Iychettira and A. Sreedevi, Diagnosis of invasive ductal carcinoma using image processing techniques, in: 2011 International Conference on Image Information Processing, IEEE, 2011, pp. 1–6.

[9] American Cancer Society, American Cancer Society. Breast Cancer, July 2022, Online; accessed 10 January 2023.

[10] A.P. Pandian, Identification and classification of cancer cells using capsule network with pathological images, *Journal of Artificial Intelligence* **1**(01) (2019), 37–44.

[11] A. Chekkoury, P. Khurd, J. Ni, C. Bahlmann, A. Kamen, A. Patel, L. Grady, M. Singh, M. Groher, N. Navab et al., Automated malignancy detection in breast histopathological images, in: Medical Imaging 2012: Computer-Aided Diagnosis, Vol. 8315, SPIE, 2012, pp. 332–344.

[12] F. Rustam, A. Ishaq, K. Munir, M. Almutairi, N. Aslam and I. Ashraf, Incorporating CNN Features for Optimizing Performance of Ensemble Classifier for Cardiovascular Disease Prediction, *Diagnostics* **12**(6) (2022), 1474.

[13] V. Rupapara, F. Rustam, A. Ishaq, E. Lee and I. Ashraf, Chi-Square and PCA Based Feature Selection for Diabetes Detection with Ensemble Classifier., *Intelligent Automation & Soft Computing* **36**(2) (2023).

[14] S.S. Yadav and S.M. Jadhav, Thermal infrared imaging based breast cancer diagnosis using machine learning techniques, *Multimedia Tools and Applications* (2022), 1–19.

[15] H. Dhahri, E. Al Maghayreh, A. Mahmood, W. Elkilani, M. Faisal Nagi et al., Automated breast cancer diagnosis based on machine learning algorithms, *Journal of Healthcare Engineering* **2019** (2019).

[16] S.A. Alanazi, M. Kamruzzaman, M.N. Islam Sarker, M. Alruwaili, Y. Alhwaiti, N. Alshammari and M.H. Siddiqi, Boosting breast cancer detection using convolutional neural network, *Journal of Healthcare Engineering* **2021** (2021).

[17] M. Umer, M. Naveed, F. Alrowais, A. Ishaq, A.A. Hejaili, S. Alsubai, A. Eshmawi, A. Mohamed and I. Ashraf, Breast Cancer Detection Using Convoluted Features and Ensemble Machine Learning Algorithm, *Cancers* **14**(23) (2022), 6015.

[18] M.F. Ak, A comparative analysis of breast cancer detection and diagnosis using data visualization and machine learning applications, *Healthcare* **8**(2) (2020), 111.

[19] Y.J. Suh, J. Jung and B.-J. Cho, Automated breast cancer de-

tection in digital mammograms of various densities via deep learning, *Journal of Personalized Medicine* **10**(4) (2020), 211.

[20] J. Zheng, D. Lin, Z. Gao, S. Wang, M. He and J. Fan, Deep learning assisted efficient AdaBoost algorithm for breast cancer detection and early diagnosis, *IEEE Access* **8** (2020), 96946–96954.

[21] H. Aljuaid, N. Alturki, N. Alsubaie, L. Cavallaro and A. Liotta, Computer-aided diagnosis for breast cancer classification using deep neural networks and transfer learning, *Computer Methods and Programs in Biomedicine* **223** (2022), 106951.

[22] M. Mangukiya, A. Vaghani and M. Savani, Breast cancer detection with machine learning, *International Journal for Research in Applied Science and Engineering Technology* **10**(2) (2022), 141–145.

[23] X. Wang, I. Ahmad, D. Javeed, S.A. Zaidi, F.M. Alotaibi, M.E. Ghoneim, Y.I. Daradkeh, J. Asghar and E.T. Eldin, Intelligent Hybrid Deep Learning Model for Breast Cancer Detection, *Electronics* **11**(17) (2022), 2767.

[24] E. Kotei and R. Thirunavukarasu, Computational techniques for the automated detection of mycobacterium tuberculosis from digitized sputum smear microscopic images: A systematic review, *Progress in Biophysics and Molecular Biology* **171** (2022), 4–16.

[25] E. Kotei and R. Thirunavukarasu, Ensemble technique coupled with deep transfer learning framework for automatic detection of tuberculosis from chest x-ray radiographs, in: Healthcare, Vol. 10, MDPI, 2022, p. 2335.

[26] F. Rustam, M.A. Siddique, H.U.R. Siddiqui, S. Ullah, A. Mehmood, I. Ashraf and G.S. Choi, Wireless capsule endoscopy bleeding images classification using CNN based model, *IEEE access* **9** (2021), 33675–33688.

[27] M. Mujahid, F. Rustam, R. Álvarez, I.d.l.T. Díez and I. Ashraf, Pneumonia classification from X-ray images with inception-V3 and convolutional neural network, *Diagnostics* **12**(5) (2022), 1280.

[28] F. Rustam, A. Ishaq, K. Munir, M. Almutairi, N. Aslam and I. Ashraf, Incorporating CNN Features for Optimizing Performance of Ensemble Classifier for Cardiovascular Disease Prediction, *Diagnostics* **12**(6) (2022), 1474.

[29] U. Repository, UCI Machine Learning Repository. https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+%28Diagnostic%29.

[30] L. Breiman, Bagging predictors, *Machine learning* **24**(2) (1996), 123–140.

[31] G. Biau and E. Scornet, A random forest guided tour, *Test* **25**(2) (2016), 197–227.

[32] M. Manzoor, M. Umer, S. Sadiq, A. Ishaq, S. Ullah, H.A. Madni and C. Bisogni, RFCNN: traffic accident severity prediction based on decision level fusion of machine and deep learning model, *IEEE Access* **9** (2021), 128359–128371.

[33] A. Juna, M. Umer, S. Sadiq, H. Karamti, A. Eshmawi, A. Mohamed and I. Ashraf, Water Quality Prediction Using KNN Imputer and Multilayer Perceptron, *Water* **14**(17) (2022), 2592.

[34] E. Besharati, M. Naderan and E. Namjoo, LR-HIDS: logistic regression host-based intrusion detection system for cloud environments, *Journal of Ambient Intelligence and Humanized Computing* **10**(9) (2019), 3669–3692.

[35] S. Sarwat, N. Ullah, S. Sadiq, R. Saleem, M. Umer, A. Eshmawi, A. Mohamed and I. Ashraf, Predicting Students' Academic Performance with Conditional Generative Adversarial Network and Deep SVM, *Sensors* **22**(13) (2022), 4834.

[36] I. Ashraf, M. Narra, M. Umer, R. Majeed, S. Sadiq, F. Javaid and N. Rasool, A Deep Learning-Based Smart Framework for Cyber-Physical and Satellite System Security Threats Detec-

tion, *Electronics* **11**(4) (2022), 667.

[37] M. Umer, S. Sadiq, M. Nappi, M.U. Sana, I. Ashraf et al., ETCNN: Extra Tree and Convolutional Neural Network-based Ensemble Model for COVID-19 Tweets Sentiment Classification, *Pattern Recognition Letters* **164** (2022), 224–231.

[38] R. Majeed, N.A. Abdullah, M. Faheem Mushtaq, M. Umer and M. Nappi, Intelligent Cyber-Security System for IoT-Aided Drones Using Voting Classifier, *Electronics* **10**(23) (2021), 2926.

[39] M. Umer, S. Sadiq, M.M.S. Missen, Z. Hameed, Z. Aslam, M.A. Siddique and M. Nappi, Scientific papers citation analysis using textual features and SMOTE resampling techniques, *Pattern Recognition Letters* **150** (2021), 250–257.

[40] L. Cascone, S. Sadiq, S. Ullah, S. Mirjalili, H.U.R. Siddiqui and M. Umer, Predicting Household Electric Power Consumption Using Multi-step Time Series with Convolutional LSTM, *Big Data Research* **31** (2023), 100360.

[41] A. Hameed, M. Umer, U. Hafeez, H. Mustafa, A. Sohaib, M.A. Siddique and H.A. Madni, Skin lesion classification in dermoscopic images using stacked Convolutional Neural Network, *Journal of Ambient Intelligence and Humanized Computing* (2021), 1–15.

[42] A.K. Dubey, U. Gupta and S. Jain, Analysis of k-means clustering approach on the breast cancer Wisconsin dataset, *International journal of computer assisted radiology and surgery* **11**(11) (2016), 2033–2047.

[43] D. Lavanya and D.K.U. Rani, Analysis of feature selection with classification: Breast cancer datasets, *Indian Journal of Computer Science and Engineering (IJCSE)* **2**(5) (2011), 756–763.

[44] O.I. Obaid, M.A. Mohammed, M. Ghani, A. Mostafa, F. Taha et al., Evaluating the performance of machine learning techniques in the classification of Wisconsin Breast Cancer, *International Journal of Engineering & Technology* **7**(4.36) (2018), 160–166.

[45] S.J. Singh, R. Rajaraman and T.T. Verlekar, Breast Cancer Prediction Using Auto-Encoders, in: International Conference on Data Management, Analytics & Innovation, Springer, 2023, pp. 121–132.

[46] A. Murphy, Breast Cancer Wisconsin (Diagnostic) Data Analysis Using GFS-TSK, in: North American Fuzzy Information Processing Society Annual Conference, Springer, 2021, pp. 302–308.

[47] P. Ghosh, Breast Cancer Wisconsin (Diagnostic) Prediction.

[48] S. Akbulut, I.B. Cicek and C. Colak, Classification of Breast Cancer on the Strength of Potential Risk Factors with Boosting Models: A Public Health Informatics Application, *Medical Bulletin of Haseki/Haseki Tip Bulteni* **60**(3) (2022).

[49] R.K. Sachdeva and P. Bathla, A Machine Learning-Based Framework for Diagnosis of Breast Cancer, *International Journal of Software Innovation (IJSI)* **10**(1) (2022), 1–11.

[50] M.F. Ak, A Comparative Analysis of Breast Cancer Detection and Diagnosis Using Data Visualization and Machine Learning Applications, *Healthcare* **8**(2) (2020). doi: 10.3390/healthcare8020111. https://www.mdpi.com/2227-9032/8/2/111.

[51] C. Liu, H. Sun, N. Du, S. Tan, H. Fei, W. Fan, T. Yang, H. Wu, Y. Li and C. Zhang, Augmented LSTM framework to construct medical self-diagnosis android, in: 2016 IEEE 16th International Conference on Data Mining (ICDM), IEEE, 2016, pp. 251–260.

[52] Y. Gao, J.M. Phillips, Y. Zheng, R. Min, P.T. Fletcher and G. Gerig, Fully convolutional structured LSTM networks for joint 4D medical image segmentation, in: 2018 IEEE 15th

international symposium on biomedical imaging (ISBI 2018), IEEE, 2018, pp. 1104–1108.

[53] S. Karimi Jafarbigloo and H. Danyali, Nuclear atypia grading in breast cancer histopathological images based on CNN feature extraction and LSTM classification, *CAAI Transactions on Intelligence Technology* **6**(4) (2021), 426–439.

[54] M. Desai and M. Shah, An anatomization on breast cancer detection and diagnosis employing multi-layer perceptron neu-

ral network (MLP) and Convolutional neural network (CNN), *Clinical eHealth* **4** (2021), 1–11.

[55] K. RAJANI, Breast Cancer Survival Dataset. https://www.kaggle.com/datasets/kreeshrajani/breast-cancer-survival-dataset/code.

[56] A.K. BARAI, UCI Machine Learning Repository. https://www.kaggle.com/datasets/ankitbarai507/breast-cancer-dataset/code.