Yugoslav Journal of Operations Research
(20##), Number #, #-#
DOI: https://doi.org/10.2298/YJOR250115026G

Research article

UNIFIED NEURAL NETWORK ENSEMBLE FOR ACCURATE BREAST CANCER CLASSIFICATION WITH ENGINEERED FEATURES

Yunus Emre GÖKTEPE*

Department of Computer Engineering, Necmettin Erbakan University, Konya, 42370, Turkey
ygoktepe@erbakan.edu.tr, ORCID: 0000-0002-8252-2616

Received: January 2025 / Accepted: June 2025

Abstract: This study presents the Unified Neural Network Ensemble for Breast Cancer Classification (UNNEBC), a novel ensemble model designed for breast cancer classification. An ensemble learning approach based on Convolutional Neural Networks (CNN) was used to improve the classification performance by utilizing three CNN models with different architectures and two fully connected dense networks. To enrich the dataset and improve classification performance, new features named Radius_Texture_Diff, Feature_Std_Dev, and Feature_Variance have been introduced. Radius_Texture_Diff captures the difference between the radius and texture features, potentially highlighting abnormalities in tumor shape and surface characteristics. Feature_Std_Dev represents the standard deviation of all features, providing a measure of variability within each tumor sample. Feature_Variance quantifies the variance across all features. By incorporating these features alongside the original dataset, we aim to enrich the feature space and enhance the model's ability to capture complex tumor patterns. The model was evaluated using the UCI breast cancer dataset and achieved an outstanding accuracy of 99.42%. It also showed strong performance in metrics such as specificity and sensitivity. The UNNEBC model tackles significant challenges in breast cancer classification, such as unequal class distribution and data variability. By using ensemble learning and integrating XGBoost as a meta-learner, the model leverages the strengths of individual networks and provides more reliable predictions. This study outperforms existing approaches. It also highlights the importance of feature engineering and ensemble learning in advancing breast cancer diagnosis.

Keywords: Breast cancer classification, Convolutional Neural Networks (CNNs), ensemble learning, feature engineering.

MSC: 68T07

*Corresponding author

1. INTRODUCTION

Breast cancer remains to be a major contributor to cancer-related mortality among women globally, emphasizing the importance of early detection for improving survival rates and patient outcomes. According to recent global cancer statistics, breast cancer accounted for nearly 11.6% of newly diagnosed cancer cases in 2022, highlighting the urgency for developing efficient diagnostic tools that can reliably distinguish between benign and malignant cases [1]. While various diagnostic approaches, including mammography and histopathological examination, provide valuable insights, the variability in human interpretation and the high volume of cases necessitate robust automated systems for reliable classification.

Recent progress in artificial intelligence (AI) has led to the development of sophisticated algorithms. These algorithms are capable of analyzing complex patterns in large datasets, making them suitable for medical diagnostics. Specifically, machine learning (ML) methods have proven their potential across various domains, with convolutional neural networks (CNNs) emerging as a cutting-edge approach for image and structured data classification [2]. CNNs offer the advantage of inherently discovering hierarchical features directly from raw data, reducing the reliance on extensive manual feature extraction. While CNNs have been predominantly utilized in image-based medical data, they also show promise in handling structured datasets, particularly when coupled with feature engineering methods [3].

The success of CNNs in cancer classification has been widely documented, with studies showing substantial improvements in diagnostic accuracy across various cancer types, including lung, skin, and breast cancer. However, a significant challenge of relying on a single CNN model lies in its tendency to overfit or underfit the data, especially in medical contexts where data imbalance and variance are common. Ensemble learning, which involves combining multiple models, offers a solution by aggregating predictions from multiple classifiers, thereby enhancing robustness, reducing bias, and capturing a broader spectrum of data patterns.

Breakthroughs in deep learning have yielded significant progress in breast cancer detection and classification. CNNs have demonstrated remarkable effectiveness in analyzing mammographic images and structured data, showcasing substantial potential for accurate diagnosis [4], [5]. In particular, ensemble models that integrate multiple CNN architectures have been shown to enhance accuracy and robustness in breast cancer classification tasks [6], [7]. These models frequently leverage transfer learning and pre-trained networks, which allow for improved performance, especially in cases where datasets are limited. Additionally, advanced feature engineering techniques along with the incorporation of diverse data types such as genetic markers, have further contributed to the precision of breast cancer predictions [8]. Although studies on the application of CNN-based mammography analysis for breast cancer diagnosis are still in their infancy, they hold significant potential to improve clinical applications and outcomes [4]. Current research attempts to address key challenges such as data imbalance, utilization of multimodal data, and development of more efficient neural architectures to further improve detection accuracy and patient follow-up [8].

Existing studies have made significant progress in the use of CNN-based and ensemble deep learning models in breast cancer classification. However, there are still some limitations that need to be addressed. Although existing methods are effective in certain cases, they have the risk of producing biased predictions, especially due to data imbalance

[9]. To address this problem, an ensemble learning method supported by XGBoost as a meta-learner was used [10], [11], [12]. The ensemble model combines the predictions of multiple CNNs and dense networks. In this way, it reduces the biases of individual models and takes advantage of diverse feature representations. In addition, XGBoost's mechanism allows minority class examples to be taken into account more in the training process by assigning higher weights to misclassified examples. This approach alleviates the data imbalance problem and increases the generalization ability of the model. In addition, the dependence on large labeled datasets poses a significant challenge due to the costly and time-consuming processing of medical images. In addition, existing models are not able to effectively integrate multimodal data or fully capture the complex and heterogeneous structure of cancerous tissues. These shortcomings highlight the need for innovative approaches with higher predictive accuracy, robustness, and generalizability. In this study, an innovative method is proposed that aims to overcome the limitations of existing methods and improve patient outcomes by increasing accuracy in breast cancer detection.

This research leverages the strengths of multiple CNN models and dense networks to improve the classification performance. In this way, an ensemble approach called Unified Neural Network Ensemble for Breast Cancer Classification (UNNEBC) is proposed. The ensemble model combines three CNN architectures with two dense, fully connected networks. Additionally, new feature engineering techniques are incorporated to enrich the dataset and improve the classification performance. In particular, features such as radius_texture_diff, feature_std_dev, and feature_variance, each designed to highlight different tumor features, are introduced. These features are integrated with the CNN architectures. Thus, the model increases its capacity to accurately distinguish between benign and malignant samples by leveraging a richer representation of features.

This research makes two important contributions. The first is the development of an ensemble CNN architecture adapted for structured breast cancer data. The other is the introduction of computational features that capture information in tumor features. Experimental studies have shown that the proposed ensemble model outperforms single CNN models and traditional machine learning classifiers in terms of accuracy, precision, and robustness. This work aims to provide a reliable, automated solution that can ultimately help clinicians make timely and accurate breast cancer diagnoses. Thus, it aims to contribute to the growing field of AI-enabled healthcare. Although deep learning models achieve successful results, a single CNN may have difficulty generalizing across different datasets. To overcome this problem, an ensemble model consisting of multiple CNNs can be used. By combining the predictions of different networks, a more accurate and robust classification is achieved. This method aims to provide a more balanced representation of the data by capturing different features to distinguish benign and malignant tumors.

2. MATERIALS AND METHODS

In this section, the structure of the proposed UNNEBC model is examined. The setup and the parts that make up the model are clearly introduced to show how the model works. Details of the techniques and methodologies employed in UNNEBC are given including data preprocessing, model architecture, ensemble learning, feature engineering, and evaluation metrics.

The UNNEBC model leverages an ensemble approach, integrating multiple convolutional and dense neural networks to create a unified system aimed at enhancing breast cancer classification accuracy. By combining the strengths of various network

architectures, UNNEBC is designed to provide a more robust and reliable diagnosis, addressing common challenges such as data imbalance and feature variability.

2.1. Problem Definition

Breast cancer continues to be a major global health issue and is the most frequently diagnosed cancer in women, with significant implications for mortality and morbidity rates worldwide. Timely and accurate diagnosis is essential in distinguishing between benign and malignant breast tumors, an aspect that can greatly influence treatment pathways and patient prognosis. Traditional diagnostic methods, including mammography, ultrasound, and histopathology, provide valuable insights but are often limited by variability in interpretation among radiologists and pathologists. Moreover, these methods are time-consuming. They are open to subjective bias, especially in cases where early-stage cancer symptoms are vague and difficult to distinguish.

With these limitations, it is evident that computer-aided diagnostic (CAD) systems are required to assist healthcare professionals in making more accurate and consistent diagnoses [13]. AI, and more particularly ML, has provided strong tools for such an end, with potential for automated analysis of complex medical data. Of the many of these, Convolutional Neural Networks (CNNs) have emerged to be particularly useful in image and structured data classification owing to their power to learn raw data-based hierarchical features automatically [14]. However, despite their strengths, standalone CNN models are marred by a series of constraints in medicine. One of the challenges is single CNN model overfitting or underfitting due to high variance and class imbalance prevalent in cancer data sets. Two, CNNs can represent not only intricate relations but also may be lacking in their representational capability over structured clinical data, potentially interfering with diagnostic accuracy and generalizability [15].

An alternative solution could be ensemble learning, which has some of the inherent shortcomings of standalone CNN models circumvented through a combination of outputs from a multitude of models for maximizing aggregate performance. Ensemble methods are well-known for robustness enhancement alongside reduction in bias and variance of prediction models, thereby creating an enhanced diagnostic device. By exploiting the different perspectives of disparate CNN and dense network structures, ensemble learning has the ability to pick up on more subtle nuances of tumor characteristics better, providing a better and more comprehensive picture of the data. Even ensemble models, however, must be carefully crafted so that they are able to capture salient features that reflect the intrinsic biological characteristics of the disease.

To further improve the robustness and interpretability of the ensemble model, feature engineering plays a critical role. Three new features—radius_texture_diff, feature_std_dev, and feature_variance—are developed here to enrich the dataset.

They are tailored specifically to emphasize tumor features in ways that can differentiate between benign and malignant cases. These features were chosen because they correspond to known patterns of tumor morphology that can make the model more sensitive and specific in detecting malignant cases.

Therefore, the problem addressed in this study is twofold:

1. Model Reliability and Accuracy: There is a need to develop an ensemble model that can accurately and reliably differentiate between benign and malignant breast tumors. The model should not only be highly accurate in classification but also reliable for different patient groups and presentations of tumors.

2. Enhanced Feature Representation: Traditional CNNs have their limitations when expressing structured data in medical diagnosis. Thus, introducing engineered features based on tumor characteristics is designed to offer a more comprehensive input representation, enhancing the model's diagnostic accuracy and make it more interpretable for medical professionals.

2.2. Dataset

The dataset that has been utilized for this study is devoted to the classification of breast cancer for the aim of distinguishing between benign and malignant instances of tumor cases. It is renowned UCI Breast Cancer Statistics Archive, generously provided from the Clinical Medicine Research Institute at the distinguished University of Wisconsin [16]. This dataset contains structured data representing various attributes of breast tumors, each of which reflects specific morphological and physical characteristics relevant to diagnosis. The features include measurements related to tumor size, shape, and texture, which are commonly used in clinical settings for cancer detection.

The dataset consists of 569 samples in total as seen in Table 1, with each sample representing an individual tumor case. Each sample includes 32 features (columns) that provide quantitative descriptors of the tumor.

Table 1: Explanation of numerical data related to the database.

	Count	Description
Number of features	32	ID, diagnosis, and 30 variables
Number of classes	2	benign / malignant
Number of instances	569	b = 357 / m = 212

These features can be generally divided into categories according to specific tumor characteristics. Radius refers to the average distance from the center to the points along the tumor's boundary, providing insights into the tumor's size. Texture measures the standard deviation of grayscale values, capturing variations in intensity across the tumor surface. Perimeter, area, and compactness relate to the tumor's boundary and shape, revealing structural irregularities that may indicate malignancy. Lastly, symmetry and fractal dimension assess the tumor's symmetry and shape complexity, as irregularities in these measurements are often associated with malignancy.

Since this data set is frequently used in scientific studies, it provides a good comparison opportunity for evaluating the proposed model [17]. Its extensive application in previous research provides a valuable benchmark, allowing for direct comparison of model performance and offering insights into advancements in breast cancer classification models. The feature-rich nature of the data, combined with the engineered attributes introduced here, aims to support precise and reliable model predictions, thereby contributing valuable insights into breast cancer diagnosis research.

2.3. Data Preprocessing

The dataset used in this study was preprocessed to ensure it was suitable for input into machine learning models. Initially, to facilitate efficient and stable training, all feature columns were scaled using the *StandardScaler* to standardize the dataset and improve convergence during model training.

For each feature X_i, the standardized value X_i' is computed as:

$$X_i' = \frac{X_i - \mu}{\sigma} \quad (1)$$

where μ is the mean and σ is the standard deviation of the feature within the dataset.

The class labels were transformed into numerical values using *LabelEncoder*. In machine learning, categorical variables must often be transformed into numerical representations for model compatibility [18]. For a binary classification problem where the target variable has two classes, such as benign and malignant, label encoding can be employed to convert these categories into a numerical format. Let y denote the target variable containing the binary classes $\{benign, malignant\}$. The transformation process can be described mathematically by defining a mapping from each category to a unique integer. For a binary classification:

$benign \rightarrow 0$

$malignant \rightarrow 1$

This mapping is applied to each instance in the target variable in order to transform the categorical values into their respective integer representations as:

$y' = LabelEncoder(y)$

The dataset was subsequently divided into training and testing sets, with 30% reserved for testing purposes. Additionally, the data was reshaped to meet the input requirements for the Convolutional Neural Network (CNN) architecture.

2.4. Feature Engineering

Feature engineering was performed to introduce additional data attributes that emphasize important tumor characteristics, aiming to improve the classifier's diagnostic accuracy. Three new features were generated: *Radius_Texture_Diff*, *Feature_Std_Dev*, and *Feature_Variance*.

Radius_Texture_Diff captures the difference between the radius and texture features, potentially indicating abnormalities in tumor shape and surface characteristics.

$$radius\_texture\_diff = X_{radius} - X_{texture} \quad (2)$$

Feature_Std_Dev: Represents the standard deviation of all features, providing a measure of variability within each tumor sample.

$$feature\_std\_dev = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(X_i - \bar{X})^2} \quad (3)$$

where N is the total number of features, X_i represents the individual feature values, and $\bar{X}$ is the mean of all features.

Feature_Variance: Measures the variance across all features, which may help identify subtle distinctions between benign and malignant cases.

These engineered features were included alongside the original dataset, adding informative dimensions that potentially assist the model in capturing complex tumor patterns. By enhancing feature representation, this step aimed to boost the classifier's predictive power and interpretability.

2.5. Dimensionality Reduction

Principal Component Analysis (PCA) was employed in this study as a dimensionality reduction technique to enhance computational efficiency and improve classification accuracy by scaling down the feature set while retaining essential variance in the data. PCA

aims to transform the original dataset into a set of uncorrelated variables, referred as principal components, that are ordered based on the variance they explain. This process helps reduce data complexity and eliminates redundancy, are ordered based on the variance they explain on the most informative features.

Let $X \in \mathbb{R}^{n*p}$ represent the standardized data matrix, where n signifies the total number of samples and p indicates the number of features. Each observation in X is centered by subtracting the mean of each feature, ensuring that PCA is unaffected by differing feature scales. PCA aims to derive a set of orthogonal vectors, or principal components, $w_1, w_2, \cdots, w_k$ that maximize the variance of the projections of X onto these vectors. Formally, the first principal component w_1 is obtained by solving the optimization problem:

$$w_1 = arg \max_{\|w\|=1} Var(Xw) = arg \max_{\|w\|=1} w^T S w \tag{4}$$

where S is the covariance matrix of X and calculated by:

$$S = \frac{1}{n-1} X^T X \tag{5}$$

Here, w_1 is the eigenvector corresponding to the largest eigenvalue λ_1 of S, as it represents the direction of maximum variance in the data.

Subsequent principal components $w_2, \cdots, w_k$ are determined iteratively by maximizing the variance in the orthogonal subspace, subject to the constraint of orthogonality to all previously identified components. Thus, the i-th principal component w_i is obtained as:

$$w_i = arg \max_{\|w\|=1,\ w \perp w_j \forall\, j<i} w^T S w \tag{6}$$

The solution involves decomposing the covariance matrix S into its eigenvalues and eigenvectors. The eigenvalues $\lambda_1, \lambda_2, \cdots, \lambda_p$ quantify the variance captured by each principal component, with larger eigenvalues indicating higher variance. By selecting the top k components, where $k<p$, PCA enables dimensionality reduction while retaining the features with the greatest variance. The transformed data matrix X_{PCA} can then be represented as:

$$X_{PCA} = XW \tag{7}$$

where $W = [w_1, w_2, \cdots, w_k]$ is the matrix of the top k eigenvectors. This reduction in dimensionality facilitates efficient training and enhances the model's capacity to make accurate predictions on unseen data by focusing on the most relevant features, ultimately contributing to improved classification performance.

2.6. Convolutional Neural Networks (CNNs)

Convolutional Neural Networks are a class of deep learning models widely used for image and structured data classification. CNNs autonomously learn hierarchical patterns of features via convolutional layers, which convolve input data with filters to extract specific features. For breast cancer classification, CNNs are particularly effective as they can capture intricate tumor patterns from imaging or structured tumor characteristics [19].

In convolutional layer of a CNN architecture, each filter W slides over the input matrix, generating feature maps where each map F at a spatial location (i, j) is given by:

$$F(i,j) = \sum_k \sum_l X(i+k, j+l).W(k,l) + b \tag{8}$$

where X is the input, W is the filter, and b is the bias term. Pooling layer reduces the spatial dimensions, typically using max-pooling to retain the most relevant features:

$$P(i,j) = max\{F(i.s + k, j.s + l) | k, l \in [0, p]\} \quad (9)$$

where s is the stride and p is the pooling size. The activation function (ReLU) applies non-linearity [20], enhancing model complexity:

$$f(x) = max(0, x) \quad (10)$$

In this study, three distinct CNN models are designed, each consisting of convolutional, max-pooling, and dense layers to automatically extract hierarchical feature representations from the data. The CNN architecture begins with a 2D convolutional layer with ReLU activation to learn spatial patterns, followed by a max-pooling layer that reduces dimensionality and prevents overfitting. The network also includes a dropout layer for regularization, enhancing the model's ability to generalize by preventing neuron co-adaptation. Each CNN outputs class probabilities via a final softmax layer. These CNNs are trained independently, and their outputs are later integrated into an ensemble model.

2.7. Dense Neural Networks

In addition to CNNs, two dense (fully connected) neural network models were implemented to complement the CNNs by focusing on the structured data aspects of the dataset. Dense networks are ideal for learning non-spatial features, as they establish connections between all neurons in successive layers. Each neuron in a dense layer interacts fully with all neurons in the preceding and subsequent layers, learning intricate feature interactions that are essential for fine-tuning the classification process.

In the structure of the dense layer forward pass calculation is executed. Each neuron output y is calculated as:

$$y = f(\sum_i w_i x_i + b) \quad (11)$$

where f is an activation function like ReLU or sigmoid, w represents weights, x is the input, and b is the bias. Dropout layers are applied to mitigate overfitting by selectively deactivating certain neurons at random during the training process. This promotes a more generalized feature learning.

The dense models consist of an input layer, two hidden layers utilizing ReLU activation, and dropout layers to avoid overfitting. The output layer employs softmax activation, allowing the dense networks to provide class probabilities for benign and malignant classifications. Similar to the CNNs, each dense network model was trained separately and then combined in the ensemble framework.

2.8. Ensemble Learning

Ensemble Learning aggregates outputs from several models to enhance robustness and minimize bias and variance. To enhance model performance, a strategy is implemented that merges outputs from various CNN and dense models. Ensemble learning helps improve the model's resilience and ability by aggregating insights from multiple classifiers. In this setup, predictions from the three CNN models and two dense models were obtained and concatenated to create a single, unified feature set. This aggregated feature set serves as input for a secondary meta-learning classifier, allowing the ensemble to leverage the strengths of both CNNs and dense networks. The ensemble learning process reduces overfitting and captures a broader range of data patterns, providing a more reliable classification of breast cancer cases.

Ensemble techniques used in this study are model averaging and stacked generalization. Model averaging combines outputs by averaging probabilities across models. Stacked

generalization trains a secondary (meta-learning) model on the output of base learners, improving final predictions:

$$z = h_1(x), h_2(x), \ldots, h_n(x) \tag{12}$$

where h_i are the base learners and z is given to the meta-learner for final classification.

2.9. Meta-Learning with XGBoost

To finalize the ensemble model, XGBoost, a gradient-boosting algorithm known for its performance and efficiency in handling structured data, is employed as a meta-learner. XGBoost is trained on the concatenated predictions from the CNN and dense models to make the final classification decision. With the objective set to binary: logistic, XGBoost optimizes for binary classification, where it differentiates between benign and malignant cases. Its regularization techniques further improve generalization, making it suitable for high-variance datasets such as those seen in medical applications. The inclusion of XGBoost as a meta-learner effectively combines the diverse strengths of the individual models in the ensemble, resulting in a robust diagnostic classifier.

XGBoost iteratively trains weak learners on residuals from previous rounds, gradually refining classification accuracy through weighted updates. The objective function for binary classification in XGBoost is:

$$f_{obj} = \sum_i l(y_i, \hat{y}_i) + \sum_k \Omega(f_k) \tag{13}$$

where ℓ is the loss function (e.g., logistic loss), and Ω is a regularization term controlling model complexity.

2.10. Architecture and Workflow of the UNNEBC

A novel ensemble model, named UNNEBC (Unified Neural Network Ensemble for Breast Cancer Classification), is proposed to enhance the accuracy and robustness of breast cancer classification tasks. The UNNEBC model integrates the strengths of various artificial intelligence architectures, aiming to achieve significant improvements in both performance and reliability.

A comprehensive pseudocode of UNNEBC is presented in Algorithm 1, which describes the sequential execution of the model.

Algorithm 1: Pseudocode for the UNNEBC model

```
Input: breast cancer dataset
Output: classification report
Load and preprocess data:
  X = all columns except the label column (features)
  y = target labels
Feature engineering:
  Calculate X['radius_texture_diff'], X['feature_std_dev'], and
X['feature_variance']
  y = transform y into numeric format using LabelEncoder
  X = scale X (zero mean, unit variance)
  Initialize PCA with n_components=nc
  X_pca = reduce dimensions of X using PCA
  Split data (train set / test set):
  Reshape X_train and X_test to fit CNN input shape
  Convert y_train and y_test to categorical format for multi-class
classification
Define CNN model:
```

```
  Add Conv2D layer, MaxPooling2D layer, Flatten layer, Dense layer,
Dropout layer
  Add output Dense layer with 2 neurons, softmax activation
Define Dense model:
  Add Dense layer, Dropout layer, Dense layer, Dropout layer
  Add output Dense layer with 2 neurons, softmax activation
Train CNN and Dense models in the first layer:
  Initialize cnn_models and dense_models
  for each model in cnn_models and dense_models do
    Compile model with optimizer=Adam(learning_rate=0.0001)
    Fit model on X_train, y_train with early stopping and validation
split of 0.25
  end for
Obtain predictions from each model in the first layer:
  for each model in cnn_models do
    cnn_predictions = [model.predict(X_train) for model in cnn_models]
  end for
  for each model in dense_models do
    dense_predictions = [model.predict(X_train) for model in
dense_models]
  end for
Train XGBoost model as meta-learner:
  stacked_train_input = concatenate cnn_predictions and
dense_predictions along axis=1
  Reshape stacked_train_input to fit XGBoost model input requirements
  Initialize xgb_model = XGBClassifier with parameters
  Fit xgb_model
Obtain predictions for test data in the first layer:
  Calculate cnn_test_predictions
  Calculate dense_test_predictions
  stacked_test_input = concatenate cnn_test_predictions and
dense_test_predictions along axis=1
  Reshape stacked_test_input to fit XGBoost model input requirements
Predict on test data using XGBoost model:
  y_pred = xgb_model.predict(stacked_test_input)
Print classification report
```

Additionally, the model's workflow, illustrating the integration of Convolutional Neural Networks (CNNs), dense neural networks, feature engineering techniques, and meta-learning strategies, is depicted in Figure 1.

This section delivers a comprehensive explanation of each component of the model, highlighting how they interact and contribute to the overall ensemble strategy, which combines predictions from multiple networks to improve diagnostic reliability. By incorporating advanced feature engineering methods, dimensionality reduction through PCA, and utilizing XGBoost as a meta-learner, the UNNEBC model is specifically designed to address common challenges in breast cancer diagnosis, such as class imbalance and data variability. These strategies work in concert to enhance the model's resilience and capacity to adapt across diverse datasets, ultimately offering a more reliable and accurate diagnostic tool for breast cancer classification.

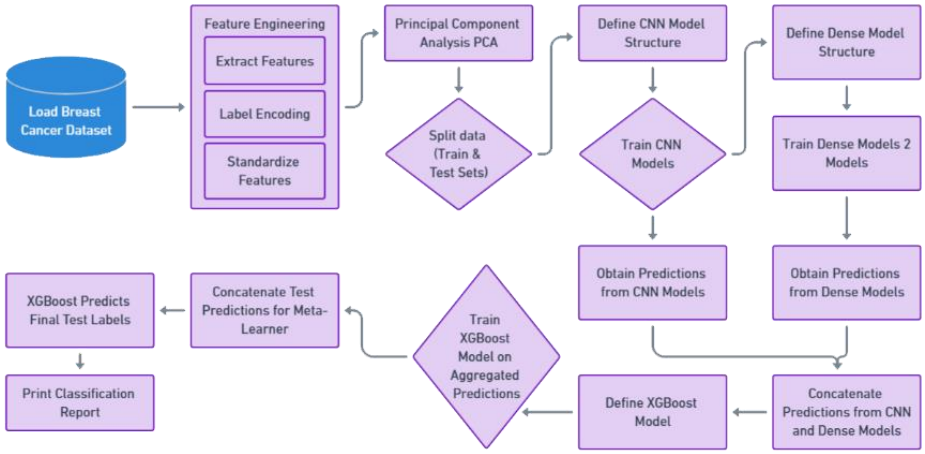

Figure 1: Workflow of the UNNEBC model

3. RESULTS AND DISCUSSION

The experimental analysis of the proposed model was conducted on a computer equipped with an Intel(R) Core (TM) i5-8250U CPU @ 1.60GHz processor and 12 GB of RAM. The implementation of the model was performed using the Python programming language on a Windows 11 Pro operating system.

3.1. Evaluation Metrics

The obtained experimental results were evaluated through a confusion matrix. It provides critical insights into the areas that need improvement for enhancing the model's accuracy and effectiveness. By highlighting the discrepancies between actual and predicted outcomes, the confusion matrix aids in pinpointing specific instances where the classifier may require refinement and adjustment.

The classification performance of the binary breast tumor classification model (benign vs. malignant) was evaluated using the following standard metrics:

True Positive (TP): Number of correctly predicted malignant cases.

True Negative (TN): Number of correctly predicted benign cases.

False Positive (FP): Number of benign cases incorrectly predicted as malignant.

False Negative (FN): Number of malignant cases incorrectly predicted as benign.

A confusion matrix was generated, providing insights into the distribution of correctly and incorrectly classified instances, including true positives, true negatives, false positives, and false negatives. This analysis provides a holistic evaluation of the model's effectiveness and reliability in a real-world diagnostic setting.

To assess the performance of the proposed ensemble model, several evaluation metrics were employed. The primary evaluation metric was accuracy, providing a general measure of correct classifications across benign and malignant cases. Additionally, a classification report detailing precision, recall, specificity, F1 score, and MCC for each class was calculated, allowing for a more nuanced understanding of the model's strengths and weaknesses.

Additionally, the model's performance was tracked using Receiver Operating Characteristic (ROC) and Precision-Recall (PR) curves. These curves are indispensable tools for performance evaluation the of classification models, particularly in scenarios involving imbalanced datasets. The ROC curve visualizes the balance between sensitivity and specificity by plotting the True Positive Rate (TPR) against the False Positive Rate (FPR) across different thresholds [21]. The area under the ROC curve (AUROC) serves as a quantitative measure of the model's ability to differentiate between classes. A higher AUROC value signifies superior performance, indicating that the model is more effective in accurately identifying both classes.

The PR curve, on the other hand, emphasizes the trade-off between Precision (true positive predictions relative to all predicted positives) and Recall (true positive predictions relative to all actual positives). The area under the PR curve (AUPR) is particularly relevant for imbalanced datasets, as it highlights the model's focus on the minority (positive) class. A high AUPR value suggests that the model effectively balances precision and recall, thereby reducing false positives and false negatives.

These curves and their associated AUC values complement each other in offering a well-rounded assessment of a classification model's performance. While the ROC curve is more suitable for evaluating models with balanced datasets, the PR curve is particularly insightful for imbalanced datasets. By analyzing both metrics, practitioners can gain a nuanced understanding of model behavior, facilitating informed decisions in algorithm selection, threshold optimization, and hyperparameter tuning for enhanced predictive accuracy.

3.2. Experimental Results

As depicted by the confusion matrix in Figure 2, the model accurately classified all 108 negative (benign) samples. Similarly, it correctly identified 62 out of 63 positive (malignant) samples.

Classification metrics were calculated based on the obtained results. Mean and maximum values of these metrics are summarized in Table 2. The model achieved a maximum precision and specificity of 100%, while other metrics also closely approached this value. When the model was repeatedly executed to compute average performance values, it achieved a high accuracy of 99.42%.

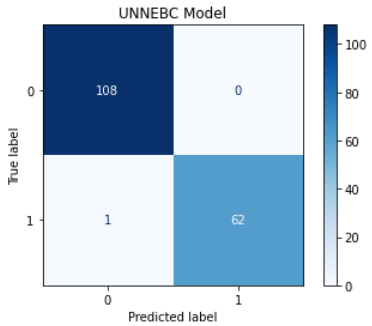

Figure 2: Confusion matrix of the proposed model

Table 2: Experimental results of the proposed model

Evaluation metrics	Obtained Results	
	Average	Maximum
Accuracy %	99,42	99,82
Precision %	99,08	100,00
Recall %	99,07	100,00
Specificity %	98,41	100,00
F1 Score %	99,07	99,53
MCC %	97,49	98,76

The average and maximum results of the model's performance metrics are illustrated in Figure 3. This figure comprehensively presents the evaluation metrics, highlighting the efficacy and robustness of the proposed model across different performance criteria.

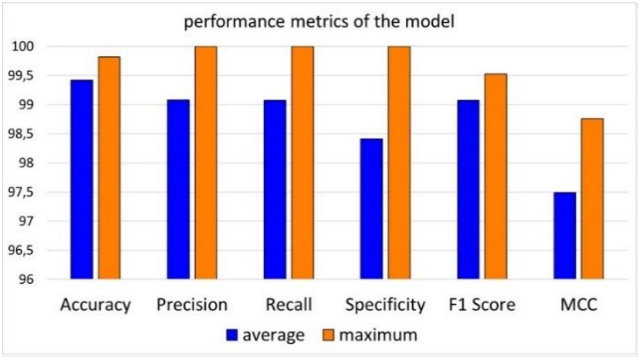

Figure 3: The average and maximum results of the model's performance metrics, including Accuracy, Precision, Recall, Specificity, F1 Score, and MCC

The ROC and PR curves of the model are given in Figure 4, which reflects the high performance of the model.

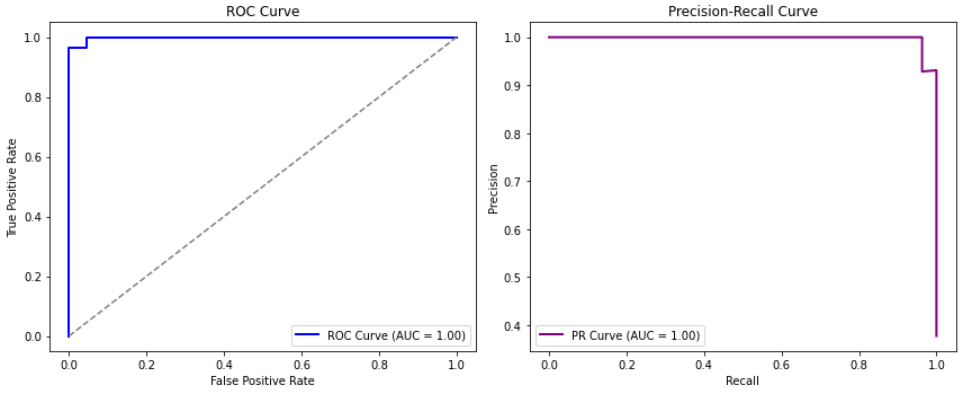

Figure 4: ROC and PR curves of the proposed model

The ROC and PR curves and corresponding AUROC and AUPR values of 1.00, underscore the model's excellent predictive capability. The results verify the model's

efficiency in effectively differentiating between classes and making firm and reliable predictions.

3.3. Comparison with previous models

The relative performance of the model compared to other approaches to breast cancer classification demonstrates its effectiveness. The proposed ensemble CNN model demonstrates a considerable improvement in accuracy with an average gain of 2,30% over the baseline CNN models. The improvement in accuracy can be attributed to the incorporation in design of engineered features and the ensemble architecture, which effectively integrates the best of a set of individual classifiers while addressing their limitations. Moreover, the enhanced precision and recall values indicate that the model not only excludes false positives but also captures malignant cases accurately, and thus it is a strong candidate for deployment in clinical practice. Table 3 provides a comparison of accuracy performance metric.

Table 3: Comparative analysis of the proposed method against cutting-edge research studies

Model	Reference	Accuracy %
GWO-SVM	[17]	98,07
SMO	[22]	96,20
BN+RBF	[23]	97,42
SVM	[24]	97,20
GONN	[25]	99,26
KNN	[26]	95,34
Proposed (UNNEBC)		99,42

In comparison to the conventional ML methods like support vector machines and random forests, the proposed model outperforms them by leveraging e hierarchical feature extraction capability of deep learning. The inclusion of engineered features further augments the classification process by providing domain-specific insights that complement the CNN's automatic feature learning.

The results highlight the strength of the new approach, particularly in addressing the challenges of imbalanced datasets and subtle feature variations in medical imaging. The findings suggest that the ensemble CNN model sets a new benchmark for breast cancer classification, paving the way for future research in integrating domain knowledge with advanced deep learning techniques.

3.4. Algorithm Complexity Analysis

The algorithm has O(n) complexity for the data loading stage, where n is the number of samples. The complexity in the data preprocessing and feature engineering steps is O(n*m) for the number of features m.

In the model training stage, the complexity can be measured as $O(k^2\ c\ n\ f)$, where k is the kernel size, c is the number of channels, and f is the number of filters. The algorithm complexity is O(n) for the MaxPooling layer and O(a b) (a: input, b: output size) for the Dense layer.

The Total Training Complexity is calculated as $O(e\ b\ (k^2\ c\ n\ f + a\ b))$, where e is the number of epochs and b is the number of batches.

4. DISCUSSION

The UNNEBC model shows how valuable using both advancements in feature engineering and dimensionality reduction techniques can be in the accurate and stable classification of breast cancer. PCA effectively assisted in reducing the dimensionality of the feature space from high-dimensional to a reasonable number without losing the most important variance in the data. The reduction resulted in fewer redundant predictor variables, increased computational efficiency, and enhanced generalization as visualized by the clear separation of classes in the PCA-reduced feature space. The ability to visualize features without PCA and then with PCA further demonstrates the value of dimensionality reduction to discriminate between benign and malignant cases, as shown in Figure 5.

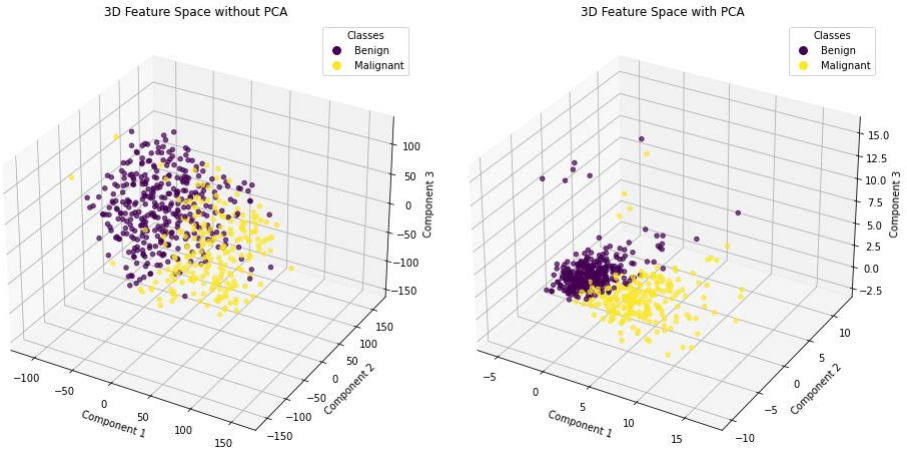

Figure 5: Feature space visualization without feature reduction and with feature reduction

In addition to the dimensionality reduction, engineered features such as *radius_texture_diff*, *feature_std_dev*, and *feature_variance* also had an important contribution in enriching the dataset. They were designed to describe subtle features of the tumor that might not be well represented in the data set by the original feature set. radius_texture_diff describes how the tumor size relates to the surface irregularity of the tumor, which might describe structural irregularity characteristics associated with aggressive tumors. feature_std_dev and feature_variance measure the variation of all of the features and demonstrate the inherent heterogeneity that is often evident in malignant tumors.

By incorporating engineered features, it improved not only the performance for discriminating between classes, but also improved interpretability as the predictions are closely representative of biological processes, allowing the predictions to be more interpretable and potentially clinically relevant.

While the engineered features contributed to improved performance, they also complemented PCA by providing biologically meaningful information that may not be captured solely through dimensionality reduction. This synergy between feature engineering and PCA underscores the significance of a holistic approach to feature representation in medical diagnostics.

Therefore, the complement of features engineered into the data and PCA demonstrates the need to combine approaches to represent the features for medical diagnosis. What is

important to observe here are the costs for each of these approaches. While PCA can reduce the feature space, it can also contribute to making the individual features less interpretable. Engineered features are domain-knowledge dependent and, although they may be biologically relevant, they may need to be further confirmed on other datasets to assess their generalizability.

Findings support a complementary contribution of PCA and engineered features in addressing the problem of high-dimensional, heterogeneous data in the breast cancer categorization. The approach improves classification performance and serves as a foundation for the development of more interpretable, more trustworthy diagnostic models. Future research will explore additional feature engineering methods, other dimensionality reduction methods, and feature engineering combined with multi-modal datasets to achieve better model performance and clinical relevance.

5. CONCLUSION

This paper introduces the Unified Neural Network Ensemble for Breast Cancer (UNNEBC) as a strong and effective ensemble model for breast cancer classification. Its model is an ensemble of many convolutional neural networks (CNN) and a dense neural network with engineered features that are purported to capture the intricate nature of tumor morphology. The results demonstrate that this approach can increase classification accuracy as the model has a 99.42% average accuracy, and both precision and recall demonstrate the model's effectiveness.

The UNNEBC model effectively addresses many breast cancer classification challenges including class imbalance, multi-feature variability, and limited data representation. The UNNEBC combines the capabilities of feature engineering and ensemble learning to develop an end-to-end diagnostic solution that is scalable and interpretable for potential clinical use. In addition, the model's ability to include a terminal decision layer on XGBoost using the principles of meta-learning allows it to use the knowledge collective of the individual members of the base level model to produce better performance rates.

Future work could see expansion of the model using multi-modal data sources such as imaging and genomic data and implement some more advanced feature selection algorithms to increase complexity of the feature space. Testing of the UNNEBC on other clinical datasets would also enable tests of generalizability and transfer to actual clinical diagnostic use. Collectively, this research illustrates that a deep learning ensemble models have the potential to revolutionize breast cancer diagnosis through the provision of a reliable, more automated, and potentially easier way of assisting in improving patient outcomes.

Availability of data and material

The dataset used in this study is publicly available from the UCI Breast Cancer Wisconsin (Diagnostic) dataset, as detailed in the manuscript [16]. The dataset is available for access at the following URL:

https://archive.ics.uci.edu/static/public/17/breast+cancer+wisconsin+diagnostic.zip.

The program codes developed for this study are available at the following GitHub repository: https://github.com/yegoktepe/breastCancerClassification.

Acknowledgements: The authors wish to acknowledge the publicly available UCI Machine Learning Repository for providing the dataset used in this study.

Funding: This research received no external funding.

REFERENCES

[1] F. Bray *et al.*, "Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: a cancer journal for clinicians*, vol. 74, no. 3, pp. 229–263, May 2024, doi: 10.3322/caac.21834

[2] J. Ker, L. Wang, J. Rao, and T. Lim, "Deep Learning Applications in Medical Image Analysis," *IEEE Access*, vol. 6, pp. 9375–9389, 2018, doi: 10.1109/ACCESS.2017.2788044

[3] G. Kourounis *et al.*, "Computer image analysis with artificial intelligence: a practical introduction to convolutional neural networks for medical professionals," *Postgraduate Medical Journal*, vol. 99, no. 1178, pp. 1287–1294, Nov. 2023, doi: 10.1093/postmj/qgad095

[4] L. Zou *et al.*, "A Technical Review of Convolutional Neural Network-Based Mammographic Breast Cancer Diagnosis," *Computational and Mathematical Methods in Medicine*, vol. 2019, pp. 1–16, Mar. 2019, doi: 10.1155/2019/6509357

[5] D. Abdelhafiz, C. Yang, R. Ammar, and S. Nabavi, "Deep convolutional neural networks for mammography: advances, challenges and applications," *BMC Bioinformatics*, vol. 20, no. S11, p. 281, Jun. 2019, doi: 10.1186/s12859-019-2823-4

[6] M. Zia Ur Rehman *et al.*, "An Efficient Automated Technique for Classification of Breast Cancer Using Deep Ensemble Model," *Computer systems science and engineering*, vol. 46, no. 1, pp. 897–911, 2023, doi: 10.32604/csse.2023.035382

[7] M. Yusoff *et al.*, "Accuracy Analysis of Deep Learning Methods in Breast Cancer Classification: A Structured Review," *Diagnostics*, vol. 13, no. 4, p. 683, Feb. 2023, doi: 10.3390/diagnostics13040683

[8] P. Bhaskar *et al.*, "Unlocking the Potential Thorough Analysis of Machine Learning for Breast Cancer Diagnosis," *International Journal of Innovative Science and Research Technology*, pp. 745–750, Apr. 2024, doi: 10.38124/ijisrt/IJISRT24APR707

[9] R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi, "Convolutional neural networks: an overview and application in radiology," *Insights Imaging*, vol. 9, no. 4, pp. 611–629, Aug. 2018, doi: 10.1007/s13244-018-0639-9

[10] R. Hoque, S. Das, and M. Hoque, "Breast Cancer Classification using XGBoost," *World Journal Of Advanced Research and Reviews*, vol. 21, no. 2, pp. 1985–1994, Feb. 2024, doi: 10.30574/wjarr.2024.21.2.0625

[11] A. Maleki, M. Raahemi, and H. Nasiri, "Breast cancer diagnosis from histopathology images using deep neural network and XGBoost," *Biomedical Signal Processing and Control*, vol. 86, p. 105152, Sep. 2023, doi: 10.1016/j.bspc.2023.105152

[12] D. Arora, R. Garg, and F. Asif, "BCED-Net: Breast Cancer Ensemble Diagnosis Network using transfer learning and the XGBoost classifier with mammography images," *Osong Public Health and Research Perspectives*, vol. 15, no. 5, pp. 409–419, Oct. 2024, doi: 10.24171/j.phrp.2023.0361

[13] J. Gao, Q. Jiang, B. Zhou, and D. Chen, "Convolutional neural networks for computer-aided detection or diagnosis in medical image analysis: An overview," *Mathematical*

Biosciences and Engineering, vol. 16, no. 6, pp. 6536–6561, 2019, doi: 10.3934/mbe.2019326

[14] F. Özen, R. O. Kabaoğlu, and T. V. Mumcu, "Deep Learning Based Temperature and Humidity Prediction," *Necmettin Erbakan Üniversitesi Fen ve Mühendislik Bilimleri Dergisi*, vol. 5, no. 2, pp. 219–229, 2023

[15] H.-P. Chan, R. K. Samala, and L. M. Hadjiiski, "CAD and AI for breast cancer—recent development and challenges," *British Journal of Radiology*, vol. 93, no. 1108, p. 20190580, Dec. 2019, doi: 10.1259/bjr.20190580

[16] O. M. William Wolberg, "Breast Cancer Wisconsin (Diagnostic)," *UCI Machine Learning Repository*, 1993, doi: 10.24432/C5DW2B

[17] S. Batool and S. Zainab, "A comparative performance assessment of artificial intelligence based classifiers and optimized feature reduction technique for breast cancer diagnosis," *Computers in Biology and Medicine*, vol. 183, p. 109215, Dec. 2024, doi: 10.1016/j.compbiomed.2024.109215

[18] A. Seveso *et al.*, "Ordinal labels in machine learning: a user-centered approach to improve data validity in medical settings," *BMC Medical Informatics and Decision Making*, vol. 20, no. S5, p. 142, Aug. 2020, doi: 10.1186/s12911-020-01152-8

[19] Z. Li *et al.*, "A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 12, pp. 6999–7019, Dec. 2022, doi: 10.1109/TNNLS.2021.3084827

[20] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015, doi: 10.1038/nature14539

[21] A. Pekgör and A. Karakoca, "ROC Analizinde Doğruluk Oranına Dayalı Uyarlanmış Kesme Değerleri ve Performans Karşılaştırmaları," *Necmettin Erbakan Üniversitesi Fen ve Mühendislik Bilimleri Dergisi*, vol. 6, no. 1, pp. 115–123, 2024.

[22] D. V. Chaurasia and S. Pal, "A novel approach for breast cancer detection using data mining techniques," *International Journal of Innovative Research in Computer and Communication Engineering*, vol. 2, 2017.

[23] M. A. Jabbar, "Breast cancer data classification using ensemble machine learning," *Engineering and Applied Science Research*, vol. 48, no. 1, 2021.

[24] M. A. Naji *et al.*, "Machine Learning Algorithms For Breast Cancer Prediction And Diagnosis," *Procedia Computer Science*, vol. 191, pp. 487–492, 2021, doi: 10.1016/j.procs.2021.07.062

[25] A. Bhardwaj and A. Tiwari, "Breast cancer diagnosis using Genetically Optimized Neural Network model," *Expert Systems with Application*, vol. 42, no. 10, pp. 4611–4620, Jun. 2015, doi: 10.1016/j.eswa.2015.01.065

[26] R. Shafique *et al.*, "Breast Cancer Prediction Using Fine Needle Aspiration Features and Upsampling with Supervised Machine Learning," *Cancers*, vol. 15, no. 3, p. 681, Jan. 2023, doi: 10.3390/cancers15030681