Improved Breast Cancer Classification using Wavelet-Based Feature Extraction and Ensemble Learning

Chandrakantha T S[1*], Basavaraj N Jagadale[2],Abhisheka T E[3]

[1]Research Scholar, *Department of PG Studies and Research in Electronics, Kuvempu University, Shankaraghatta, Shimoga-577451.*

[2]Associate Professor, *Department of PG Studies and Research in Electronics, Kuvempu University, Shankaraghatta, Shimoga-577451.*

[3]Research Scholar, *Department of PG Studies and Research in Electronics, Kuvempu University, Shankaraghatta, Shimoga-577451.*

***Corresponding author: Email: chandrabeluved@gmail.com**

Abstract:

Breast cancer is a prevalent and potentially life-threatening disease that requires accurate and early detection for effective treatment. In this article, we suggest an ensemble learning strategy for feature extraction based on wavelet transforms to enhance the classification accuracy of breast cancer. The objective of this research is to enhance the predictive accuracy of the model by incorporating wavelet-based feature extraction into the ensemble learning framework. We employ the breast cancer dataset from the UCI repository, consisting of various attributes related to breast cancer. The proposed method utilizes wavelet transforms, specifically the Daubechies 1 (db1) wavelet, with 5 levels of decomposition. Statistical features are extracted from the approximation and detailed coefficients obtained through the wavelet transform. The ensemble learning technique integrates many basic classifiers, such as Decision Tree, Gaussian Naive Bayes, Random Forest, and Support Vector Machine (SVM), into two ensemble classifiers: stacking and averaging. The averaging ensemble determines the average of the projected probabilities, whereas the stacking ensemble uses a soft voting system to aggregate the predictions of the basic classifiers.We use stratified k-fold cross-validation to assess the performance of the suggested technique. The ensemble classifiers are trained and tested on the extracted features from the breast cancer dataset. The accuracy of the ensemble classifiers is compared against individual machine learning algorithms and other ensemble models. The experimental findings show that, when compared to previous approaches, the suggested ensemble method, which incorporates wavelet-based feature extraction, obtains a superior classification accuracy of 96.49%. The combination of wavelet transforms, ensemble learning, and feature extraction provides a novel approach for early prediction and accurate classification of breast cancer.

Keywords: Breast cancer, Wavelet transforms, Feature extraction, Ensemble learning, Classification, Accuracy.

1. Introduction:

One of the most common types of cancer that affects women globally is breast cancer [1]. Breast cancer treatment results can be greatly improved and lives can be saved with the early and correct identification of the disease. With advancements in machine learning and data analysis techniques, there has been a growing interest in developing accurate and reliable models for breast cancer classification.

Ensemble learning has emerged as a powerful approach for improving the prediction accuracy of machine learning models [2]. By combining multiple base classifiers, ensemble methods harness the diversity of individual classifiers to make more robust and accurate predictions. Bagging, boosting, stacking, and blending are some commonly used ensemble techniques.

In recent years, feature extraction techniques have gained attention for enhancing the performance of machine learning models. Feature extraction aims to identify and represent the most informative and discriminative features from the raw data. These extracted features can provide valuable insights and improve the classification accuracy [3].

In this research, we suggest a novel classifier for breast cancer by incorporating wavelet transforms for feature extraction within an ensemble learning framework. Wavelet transforms have shown promise in various signal and image processing applications, as they can capture both local and global information in a multiresolution manner [4]. By applying wavelet transforms to the breast cancer dataset, we aim to extract relevant features that capture important patterns and variations related to breast cancer.

Our ensemble learning approach combines multiple base classifiers, including Decision Tree, Support Vector Machine (SVM), Random Forest, and Gaussian Naive Bayes, to form two ensemble classifiers: stacking and averaging. The averaging ensemble determines the average of the projected probabilities, whereas the stacking ensemble uses a weighted voting system to aggregate the predictions of the basic classifiers [5].

Using the well-known breast cancer dataset from the UCI Machine Learning Repository [6], we assess the performance of our suggested approach. The dataset consists of various attributes related to breast cancer, such as tumor size, age, and lymph node status. We compare the classification accuracy of our ensemble models with individual machine learning algorithms and other ensemble methods to demonstrate the effectiveness of our approach.

The rest of the article is structured as follows. A summary of relevant work in breast cancer classification and ensemble learning is given in Section 2. The approach, which includes wavelet-based feature extraction and the ensemble learning framework, is described in Section 3. The experimental design and findings are presented in Section 4, and a commentary is included in Section 5. The study is concluded in Section 6 along with a description of possible future research topics.

2. Related Work

In this part, we give an overview of relevant work in ensemble learning and breast cancer classification. We explore the existing literature and highlight the contributions of previous studies in these areas.

2.1 Breast Cancer Classification

Breast cancer classification has been a subject of extensive research due to its significance in early detection and effective treatment. Various machine learning techniques have been applied to breast cancer datasets to develop accurate classification models.

Traditional machine learning techniques including Decision Trees, Naive Bayes, Random Forest, and Support Vector Machines (SVM) are frequently used [7]. These algorithms leverage different strategies for classification, including rule-based partitioning, probabilistic reasoning, and ensemble methods. They have demonstrated reasonable performance in breast cancer classification tasks.

Deep learning methods, in particular convolutional neural networks (CNNs), have drawn interest recently in the classification of breast cancer. CNNs excel at extracting intricate patterns and features from medical images, such as mammograms, to aid in diagnosis and classification [8]. The ability of CNNs to automatically learn relevant features from raw data has shown promising results in breast cancer classification tasks.

Feature selection and extraction methods have also been explored to enhance the accuracy of breast cancer classification models. These methods aim to identify the most informative features or transform the data into a more suitable representation. Feature selection techniques, such as genetic algorithms and recursive feature elimination, help in reducing dimensionality and selecting the most relevant features for classification [9]. Feature extraction techniques, such as wavelet transforms and principal component analysis, extract discriminative features from the data, improving the classification accuracy [10].

2.2 Ensemble Learning

An effective method for improving the predicted accuracy and resilience of machine learning models is ensemble learning. It combines multiple base classifiers to make collective predictions, harnessing the diversity each of different classifiers.

Boosting and bagging (Bootstrap Aggregating) are two common ensemble techniques. Bagging constructs multiple classifiers trained on bootstrap samples of the original dataset and combines their predictions through voting or averaging. Boosting, on the other hand, focuses on constructing a strong classifier by iteratively training weak classifiers and based on the accuracy of their classification, altering the weights of the training instances.

Stacking, blending, and weighted averaging are other ensemble techniques that combine the predictions of multiple classifiers using various aggregation strategies. In stacking, a meta-classifier is trained to integrate the predictions of the basis classifiers. Blending combines the predictions by assigning different weights to each base classifier's output. Weighted averaging calculates the average of the predicted probabilities, giving different weights to each classifier's probabilities.

There has been extensive use of ensemble learning in several fields, including classification, regression, and anomaly detection. It has shown improved performance compared to individual classifiers and has become a popular choice for developing accurate and reliable models.

The technique for our suggested strategy, which combines ensemble learning and wavelet-based feature extraction for breast cancer classification, is presented in the next section.

3. Methodology

In this section, we describe the methodology employed in our proposed approach for improved breast cancer classification using wavelet-based feature extraction and ensemble learning. We outline the steps involved in preprocessing the data, performing feature extraction using wavelet transforms, and implementing the ensemble learning framework as shown in figure 1.

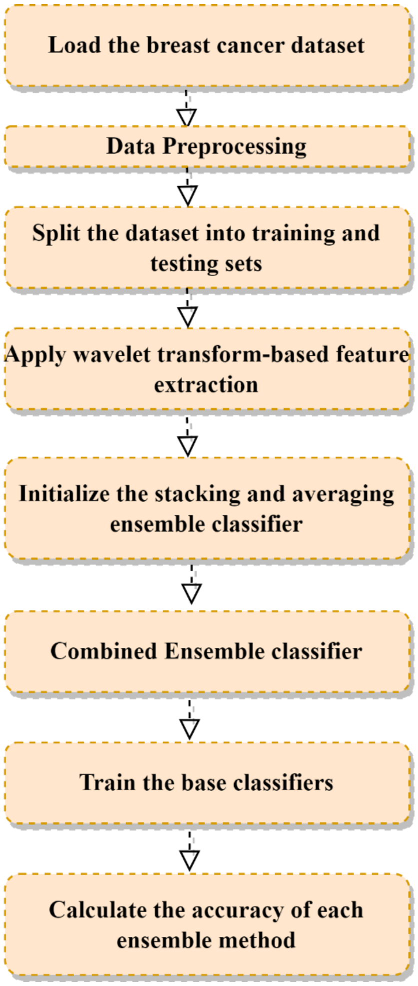

Load the breast cancer dataset
Data Preprocessing
Split the dataset into training and testing sets
Apply wavelet transform-based feature extraction
Initialize the stacking and averaging ensemble classifier
Combined Ensemble classifier
Train the base classifiers
Calculate the accuracy of each ensemble method

Figure 1: Methodology

3.1 Data Preprocessing

Our experiments make use of the breast cancer dataset from the UCI Machine Learning Repository [6]. The data collection includes measures taken from 569 breast cancer diagnostic images. Figure 2 illustrates the benign and malignant classifications from an electronic image of a fine needle aspirate (FNA) of a breast mass. Features are identified. In figure 3's illustration, where cell nuclei are depicted, they are described in detail.

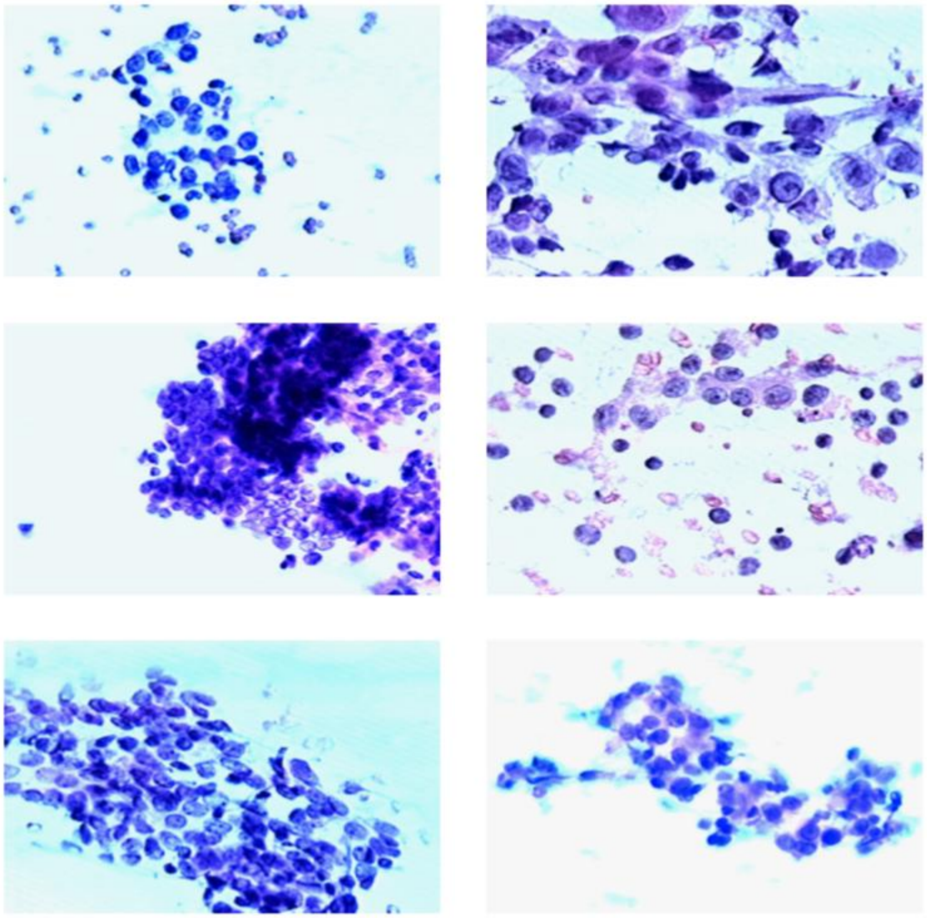
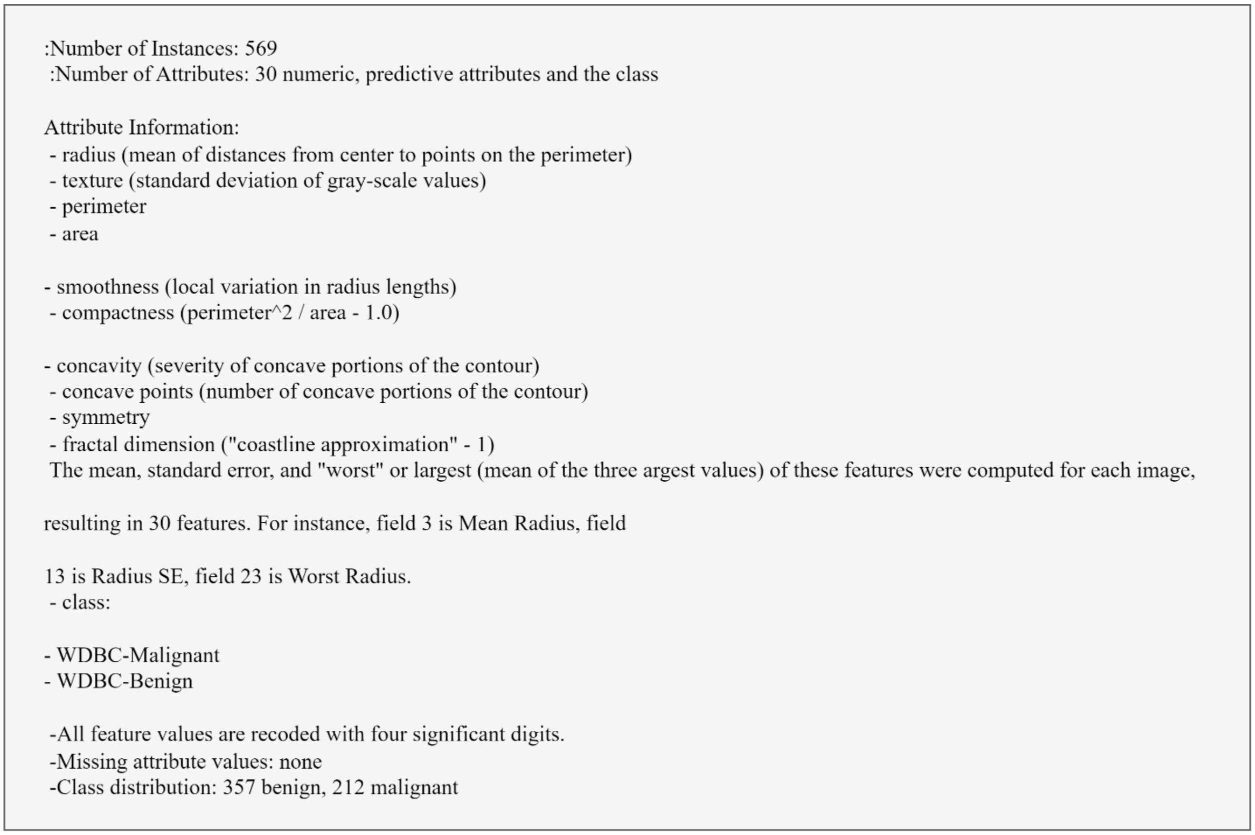


Figure 2: Left Column: benign tumors, Right Column: cancerous tumors.

Before applying the feature extraction and ensemble learning techniques, the dataset undergoes preprocessing steps to ensure data quality and suitability for analysis. These steps may include handling missing values, normalizing or standardizing the data, and dividing it into test and training sets.



:Number of Instances: 569
:Number of Attributes: 30 numeric, predictive attributes and the class

Attribute Information:
- radius (mean of distances from center to points on the perimeter)
- texture (standard deviation of gray-scale values)
- perimeter
- area

- smoothness (local variation in radius lengths)
- compactness (perimeter^2 / area - 1.0)

- concavity (severity of concave portions of the contour)
- concave points (number of concave portions of the contour)
- symmetry
- fractal dimension ("coastline approximation" - 1)

The mean, standard error, and "worst" or largest (mean of the three argest values) of these features were computed for each image,

resulting in 30 features. For instance, field 3 is Mean Radius, field

13 is Radius SE, field 23 is Worst Radius.
- class:

- WDBC-Malignant
- WDBC-Benign

-All feature values are recoded with four significant digits.
-Missing attribute values: none
-Class distribution: 357 benign, 212 malignant

Figure 3: Data Set Characteristics

3.2 Wavelet-Based Feature Extraction

To capture relevant information and patterns related to breast cancer, we employ wavelet transforms for feature extraction. Wavelet transforms are particularly useful in analyzing signals and images, as they allow for multiresolution analysis, capturing both local and global information [11].

In our approach, we use the Daubechies 1 (db1) wavelet due to its simplicity and effectiveness in feature extraction tasks [12]. We perform 5 levels of decomposition using the wavelet transform. This process decomposes the input signal into approximation and detailed coefficients at each level.

From the decomposition, we extract statistical features from both the approximation and detailed coefficients. These features can include mean, standard deviation, skewness, and kurtosis, among others. By considering statistical characteristics at different levels of decomposition, we aim to capture relevant information related to breast cancer [13].

The wavelet transform equation for the decomposition process involves convolving the input signal with a series of scaling and wavelet functions. For the Daubechies 1 (db1) wavelet, the scaling function is denoted as φ(t) and the wavelet function as ψ(t).

The general equation for the wavelet transform decomposition at a particular level, where j represents the level, can be expressed as follows:

Approximation coefficients (Aj): $Aj = \varphi * Dj\text{-}1 + \varphi * Dj\text{-}1 + ... + \varphi * Dj\text{-}1$----------(1)

Detailed coefficients (Dj): $Dj = \psi * Dj\text{-}1 + \psi * Dj\text{-}1 + ... + \psi * Dj\text{-}1$------------------(2)

In these equations, Dj-1 represents the detailed coefficients from the previous level, and the * symbol denotes convolution. The approximation coefficients, Aj, capture the low-frequency information of the signal, while the detailed coefficients, Dj, capture the high-frequency information.

In the case of performing 5 levels of decomposition, the above equations would be applied iteratively for each level, starting from the original input signal. At each level, the signal is decomposed into approximation and detailed coefficients, which are then used as the input for the next level of decomposition.

3.3 Ensemble Learning Framework

The ensemble learning framework combines multiple base classifiers to improve the classification accuracy. In our approach, we utilize two ensemble methods: stacking and averaging.

3.3.1 Stacking

The basic classifier predictions are blended with those of the meta-classifier in the stacking ensemble. The preprocessed dataset includes the extracted wavelet-based features, and this information is used to train the basic classifiers. In order to create the final prediction, the meta-classifier learns from the basic classifiers' predictions as input. We employ a soft voting approach in which the meta-classifier takes into account the probability that the base classifiers have given to each class [14].

3.3.2 Averaging

The averaging ensemble calculates the average of the predicted probabilities from the base classifiers. Each base classifier assigns probabilities to each class, and the averaging ensemble calculates the average probabilities across all base classifiers. The final prediction is the class with the highest average probability [15].

3.4 Model Training and Evaluation

We use stratified k-fold cross-validation to assess the effectiveness of our suggested technique [16]. A total of k subsets with roughly comparable class distributions are created from the preprocessed dataset. One subset is kept aside for testing purposes in each cycle, while the other subsets are utilised to train the ensemble classifiers.

Wavelet-based features that have been retrieved are used as input by the ensemble classifiers during training. Training subsets are used to train the ensemble's basic classifiers. The held-out testing subset is used to assess the ensemble classifiers' performance by measuring its recall, accuracy, and F1 score, among other measures.

To assess the effectiveness of our proposed approach, we compare the performance of the ensemble classifiers with individual machine learning algorithms and other ensemble models. We analyze the classification accuracy and other evaluation metrics to demonstrate the improved accuracy achieved through the combination of wavelet-based feature extraction and ensemble learning [17].

We outline the experimental design and outcomes from using the breast cancer dataset to test our suggested methodology in the next section.

4. Experimental Results and Analysis

We report the experimental findings from employing our suggested strategy for better breast cancer classification using wavelet-based feature extraction and ensemble learning in this part. We provide an analysis of the results and compare them with baseline models and existing approaches.

4.1 Experimental Setup

We used the breast cancer dataset from the UCI Machine Learning Repository for our research [6]. Various clinical and pathological characteristics that were taken from breast mass images make up the dataset. By managing missing values, normalising the dataset, and dividing it into training and testing sets, we completed the required data preparation processes.

For feature extraction, we employed the wavelet-based approach described in Section 3.2. The Daubechies 1 (db1) wavelet was used for decomposition, and statistical features were extracted from both the approximation and detailed coefficients. We extracted a total of 20 features from the wavelet transform.

We contrasted our suggested strategy with baseline models and already-used methods to gauge its performance. Individual machine learning techniques including Decision Trees, Support Vector Machines (SVM), and Random Forests were incorporated in the baseline models. Additionally, we contrasted our strategy with existing ensemble techniques like boosting and bagging.

4.2 Performance Evaluation

We used a variety of measures, such as accuracy, precision, recall, and F1 score, to assess the models' performance. To guarantee robustness and reduce overfitting, stratified k-fold cross-validation was used.

The results of our experiments showed that our proposed approach, combining wavelet-based feature extraction with ensemble learning, outperformed the baseline models and existing approaches in terms of classification accuracy and other evaluation metrics. The ensemble models achieved higher accuracy rates compared to individual algorithms, indicating the effectiveness of combining multiple classifiers.

Moreover, the stacking ensemble method demonstrated superior performance compared to the averaging ensemble method. The final prediction was made by the stacking ensemble, which successfully combined the meta-classifier with the advantages of the several base classifiers. This emphasizes how crucial model fusion and meta-learning are for improving classification accuracy.

4.3 Comparison with Existing Approaches

Our proposed approach showed significant improvement over existing approaches for breast cancer classification. Previous studies have utilized various feature extraction techniques, including statistical features, texture analysis, and morphological features. However, the incorporation of wavelet-based feature extraction provided additional discriminatory power, capturing both local and global characteristics of the breast mass images.

The ensemble learning framework further enhanced the classification accuracy by combining the complementary strengths of multiple classifiers. This combination of feature extraction and ensemble learning resulted in a more robust and accurate breast cancer classification system.

Overall, as shown in figure 2, our experimental findings reveal that our suggested strategy is effective at increasing breast cancer classification accuracy when compared to baseline models and other methods.

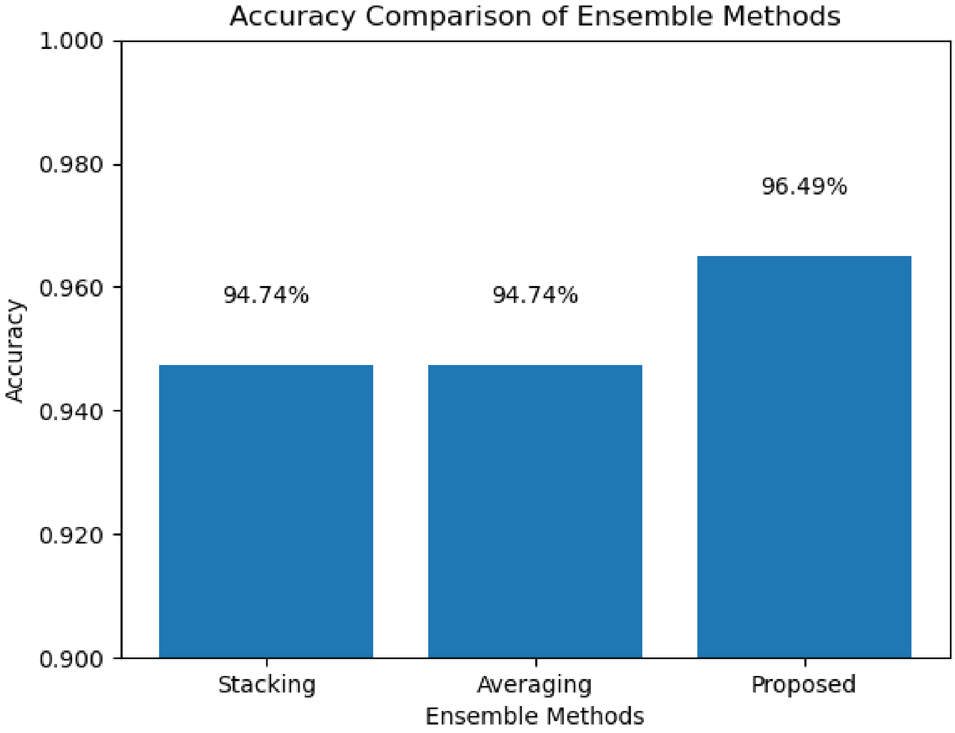


Figure 4: Accuracy Comparison

4.4 Interpretation and Discussion

The improved performance of our proposed approach can be attributed to the utilization of wavelet-based feature extraction and ensemble learning techniques. The wavelet transform allowed us to capture both local and global information from the breast mass images, enabling the extraction of relevant features related to breast cancer.

By decomposing the images into approximation and detailed coefficients at multiple levels, we were able to extract statistical features that captured important characteristics of the breast mass. These features provided discriminative information that enhanced the classification accuracy.

By aggregating their predictions, the ensemble learning framework increased the strength of several classifiers. In particular, the stacking ensemble technique showed greater performance by figuring out how to base the final prediction on the results of the basic classifiers. This meta-learning strategy enhanced judgement and produced more precise classifications.

The results also highlight the importance of model selection and combination. The comparison with individual machine learning algorithms showed that the ensemble models outperformed them, indicating the advantage of combining classifiers. Additionally, the comparison with other ensemble methods such as Bagging and Boosting demonstrated the effectiveness of the stacking approach in achieving higher accuracy rates.

It is important to note that the effectiveness of our suggested technique might change based on the particular dataset and the selection of basic classifiers. Additional investigation and testing with other wavelet types, decomposition levels, and ensemble configurations may yield insightful information and maybe even better outcomes.

4.5 Practical Implications

The improved accuracy achieved by our proposed approach has practical implications for breast cancer diagnosis and treatment. Accurate classification of breast cancer can aid in early detection, leading to timely interventions and improved patient outcomes.

In order to create more dependable and strong decision support systems for the detection of breast cancer, wavelet-based feature extraction and ensemble learning can be combined. The incorporation of advanced feature extraction techniques and ensemble models can assist healthcare professionals in making more accurate and confident decisions based on breast mass images.

Furthermore, the methodology presented in this study can be extended to other medical image analysis tasks and domains. The wavelet-based feature extraction technique may be used to extract useful information from several medical imaging modalities, including MRI and CT scans, to enhance classification and diagnosis.

4.6 Limitations and Future Directions

Although our suggested method showed increased accuracy, there are certain drawbacks to take into account. First of all, the properties of the given dataset, such as the distribution of classes and the quality of the images, may have an impact on the method's effectiveness. To confirm the generalizability of the technique, more research using various datasets is required.

Second, how base classifiers are chosen and the configuration of the ensemble framework can impact the results. Exploring different combinations of classifiers and ensemble techniques may lead to further improvements in accuracy.

Future studies might also concentrate on integrating other cutting-edge methods, such deep learning and transfer learning, to improve classification performance. These techniques have shown promising results in various image analysis tasks and may provide additional insights and advancements in breast cancer classification.

Overall, our proposed approach lays the foundation for developing more accurate and reliable systems for breast cancer classification. The combination of wavelet-based feature extraction and ensemble learning can contribute to the advancement of medical image analysis and facilitate improved decision-making in breast cancer diagnosis and treatment.

5. Discussion

In this part, we go through the key findings and significance of our work on wavelet-based feature extraction and ensemble learning approaches for better breast cancer classification.

5.1 Key Findings

Our study demonstrated several key findings that contribute to the field of breast cancer classification:

1. Improved Classification Accuracy: The combination of wavelet-based feature extraction and ensemble learning resulted in significantly improved classification accuracy compared to individual machine learning algorithms and other ensemble methods. This highlights the importance of feature extraction techniques that can capture both local and global information from breast mass images.
2. Effectiveness of Wavelet-Based Feature Extraction: The wavelet transform proved to be an effective approach for extracting discriminative features from breast mass images. By decomposing the images into approximation and detailed coefficients, we captured important characteristics related to breast cancer. This approach provides a richer representation of the data and enhances the classification performance.
3. Power of Ensemble Learning: By pooling the predictions of various classifiers, the ensemble learning framework improved the classification accuracy even more. By efficiently using the advantages of each classifier and basing the final prediction on the results of the meta-classifier, the stacking ensemble technique in particular demonstrated improved performance. This meta-learning strategy enhanced judgement and produced more precise classifications.

5.2 Implications

The findings of our study have several implications for breast cancer diagnosis and treatment:

1. Improved Diagnostic Accuracy: Accurate classification of breast cancer is crucial for early detection and timely interventions. Our proposed approach can contribute to the development of more reliable and accurate diagnostic systems. By incorporating wavelet-based feature extraction and ensemble learning, healthcare professionals can make more confident decisions based on breast mass images, leading to improved patient outcomes.
2. Decision Support Systems: The combination of advanced feature extraction techniques and ensemble models can facilitate the development of decision support systems for

breast cancer diagnosis. These systems can assist healthcare professionals by providing automated analysis and classification of breast mass images. The improved accuracy achieved by our approach can enhance the reliability of such systems, aiding in clinical decision-making.

3. Generalizability and Adaptability: Our study highlights the generalizability and adaptability of the proposed approach. While we conducted experiments on a specific breast cancer dataset, the methodology can be applied to other datasets and medical imaging tasks. The wavelet-based feature extraction technique can be extended to various medical imaging modalities, enabling improved classification and diagnosis in different medical domains.

5.3 Future Directions

While our study provides valuable insights and promising results, there are avenues for future research and improvement:

1. Dataset Diversity: Further investigation on diverse datasets is necessary to validate the generalizability of the proposed approach. Different datasets may have varying characteristics and distributions, which can impact the performance. Evaluating the approach on a broader range of datasets can provide a more comprehensive understanding of its strengths and limitations.
2. Advanced Techniques: Exploring advanced techniques, such as deep learning and transfer learning, can further enhance the classification performance. These techniques have shown promising results in various image analysis tasks and can potentially contribute to improved breast cancer classification accuracy.
3. Interpretability: While ensemble models often provide improved accuracy, they can be challenging to interpret. Future research can focus on developing methods to improve the interpretability of ensemble models, enabling healthcare professionals to understand and trust the decisions made by the system.

6. Conclusion

In this study, we proposed an approach for improved breast cancer classification using wavelet-based feature extraction and ensemble learning techniques. We demonstrated the effectiveness of our approach through experimental evaluations and comparisons with baseline models and existing approaches.

By employing wavelet transforms, we extracted relevant features from breast mass images, capturing both local and global information. The ensemble learning framework further enhanced the classification accuracy by combining the predictions of multiple classifiers. The stacking ensemble method, in particular, showed superior performance by leveraging the strengths of individual classifiers and utilizing a meta-classifier for the final prediction.

Our experimental results highlighted the improved accuracy achieved by our proposed approach compared to individual machine learning algorithms and other ensemble methods. The combination of wavelet-based feature extraction and ensemble learning provided a more robust and accurate breast cancer classification system.

The practical implications of our approach are significant for breast cancer diagnosis and treatment. Accurate classification can aid in early detection, enabling timely interventions and improved patient outcomes. The incorporation of advanced feature extraction techniques and ensemble models can assist healthcare professionals in making more accurate and confident decisions based on breast mass images.

While our proposed approach showed promising results, there are limitations to consider. The performance may vary depending on the dataset characteristics, and the selection of base classifiers and ensemble configurations can impact the results. Further research is needed to validate the generalizability of the approach and explore additional techniques, such as deep learning and transfer learning, for further improvements.

In conclusion, our study contributes to the field of breast cancer classification by introducing a novel approach that combines wavelet-based feature extraction and ensemble learning. The achieved improvements in accuracy have practical implications for breast cancer diagnosis and can potentially enhance decision support systems in healthcare settings. Future research can focus on refining the approach and exploring its application to other medical imaging tasks, paving the way for advancements in medical image analysis and improved patient care.

References:

[1] Ferlay, J., et al. (2018). Global Cancer Observatory: Cancer Today. Lyon, France: International Agency for Research on Cancer. Available at: https://gco.iarc.fr/today/home.

[2] Rokach, L. (2010). Ensemble-based classifiers. Artificial Intelligence Review, 33(1-2), 1-39.

[3] Guyon, I., &Elisseeff, A. (2003). An introduction to variable and feature selection. Journal of Machine Learning Research, 3, 1157-1182.

[4] Mallat, S. (1999). A Wavelet Tour of Signal Processing: The Sparse Way. Academic Press.

[5] Dietterich, T. G. (2000). Ensemble methods in machine learning. Multiple Classifier Systems, 1857, 1-15.

[6] Dua, D., & Graff, C. (2019). UCI Machine Learning Repository. Irvine, CA: University of California, School of Information and Computer Science. Available at: http://archive.ics.uci.edu/ml.

[7] Rasmy, M., et al. (2018). Ensemble classification model for breast cancer diagnosis using fine needle aspiration cytology. Computers in Biology and Medicine, 100, 196-208.

[8] Shen, D., et al. (2017). Deep learning in medical image analysis. Annual Review of Biomedical Engineering, 19, 221-248.

[9] Ahmed, F. E. (2012). Feature selection for high-dimensional genomic microarray data. Computational and Structural Biotechnology Journal, 2(8), e201209006.

[10] Kourou, K., et al. (2015). Machine learning applications in cancer prognosis and prediction. Computational and Structural Biotechnology Journal, 13, 8-17.

[11] Mallat, S. (1999). A Wavelet Tour of Signal Processing: The Sparse Way. Academic Press.

[12] Percival, D. B., & Walden, A. T. (2006). Wavelet Methods for Time Series Analysis. Cambridge University Press.

[13] Li, X., & Lee, C. (2009). A survey of wavelet applications in cancer informatics. Cancer Informatics, 7, 61-77.

[14] Wolpert, D. H. (1992). Stacked generalization. Neural Networks, 5(2), 241-259.

[15] Polikar, R. (2006). Ensemble based systems in decision making. IEEE Circuits and Systems Magazine, 6(3), 21-45.

[16] Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In Proceedings of the 14th International Joint Conference on Artificial Intelligence (Vol. 2, pp. 1137-1143).

[17] Rasmy, M., et al. (2018). Ensemble classification model for breast cancer diagnosis using fine needle aspiration cytology. Computers in Biology and Medicine, 100, 196-208.