Indian Engineering Journal
ISSN: 2984-9047
Vol 4 Issue 1 (2026)

A Rank-Consensus Ensemble Feature Selection and Soft-Voting Framework for Interpretable Breast Cancer Prediction

M. Priyadharshini[1], V. Murugesh[2]

[1]Department of CSE, Faculty of Science and Technology (IcfaiTech), ICFAI Foundation for Higher Education, Hyderabad, 501203, Telangana, India. mpriyadharshinimca@gmail.com

[2]School of Computer Science, Coventry University Kazakhstan, 3A,Korgalzhyn Highway, Astana, Kazakhstan. murugesh72@gmail.com

ABSTRACT:

Early and correct prediction of breast cancer is clinically relevant due to the complex, correlated and non-linear nature of patterns in fine-needle aspiration measurements. In this work, a rank consensus ensemble feature selection and soft-voting classification approach to classify the Breast Cancer Wisconsin Diagnostic data sets into distinction between Malignant and Benign classes is presented. The idea behind the proposed framework is to combine three complementary flavours of feature-selection: dependency in non-linear manner using MI, separability of classes using Fisher score and recursive feature elimination aimed at example-model guided relevance. The outputs of these selectors are then normalized and the score combined to create a consensus score and feature importance score for reduced redundancy and a small feature-set that can be clinically interpreted. Logistic regression, random forest and decision tree learners are then combined in a probabilistic soft-voting manner to enhance the stability of prediction. The most informative features found were: worst radius, worst concave points, mean concave points, mean radius and worst perimeter. Results from the hold-out evaluation set showed that the proposed model was able to correctly classify 109 out of 114 cases with 95.61% accuracy, 95.24% precision, 93.02% sensitivity, 97.18% specificity and 94.12% F1 score. Unconfined analysis of the confusion matrix also reveals low false-positive and false-negative rates showing this proposed framework offers good balance of clinical sensitivity and diagnostic specificity. The key novelty this will bring is that the pipeline is interpretable, it is low dimensional and ensemble based, and it will help computer aided screening of breast cancer, which also is feasible in the computational space of clinical decision support.

Keywords: Breast Cancer Prediction; Ensemble Feature Selection; Rank Consensus; Soft Voting Classifier; Machine Learning; Diagnostic Decision Support.

1. Introduction

Breast cancer is one of the most thoroughly investigated malignant diseases in computer-aided diagnosis where timely clinical intervention can be exercised through the early diagnosis of the disease. Traditional methods like mammography, magnetic resonance imaging, ultrasound examination, and biopsy to perform cytological analysis are clinically essential, but require trained experts, quality imaging and appropriate infrastructure for the diagnostics [1]. This may be achieved using machine learning, which may act as an additional decision support system as the machine learning model can be used to detect complex relationships in the quantitative measurements of morphology obtained from FNA images.

One of the main causes of difficulty in predicting the classification of breast cancer is not only because of the fact that it is a binary issue (malignant or benign), but also because several of its cytological attributes are strongly correlated. Various metrics including radius, perimeter, area, concavity, compactness and concave points, are measurements, which describe related properties of tumour morphology. When all features are plugged directly into a classifier, the model can become unnecessarily heavy, less readable, and have more computations. Hence, the ideal diagnostic model should be able to achieve high model classification performance along with a set of small no of clinically significant variables [2].

The present study extends on the previous manuscript, where the method is re-formulated as a rank-consensus based ensemble feature selection and soft-voting. The novelty in the approach is that it involves three

heterogeneous feature-selection logics (before the fusion of the classifiers). Mutual information considers non-linear dependency between features and the class label, Fisher score models favour those features with high between-class separation and low within-class variations, and recursive feature elimination brings a supervised model-based perspective. Their consensus decreases risk of depending on 1 selector and also stemming to a more steady base of features [3], [4]. Since a soft-voting classifier collects the probabilistic output decisions of multiple classifiers, it is employed to leverage the output of multiple classifiers in the decision.

The research is motivated with the following research objectives:

- To present a novel rank-consensus ensemble feature-selection method based on a mix of filter and wrapper-based relevance estimation.
- To build a compact prediction model for breast cancer, using the most important diagnostic features.
- To assess and analyze the proposed soft-voting classifier using clinically relevant Performance metrics like Accuracy, Precision, Sensitivity, Specificity, F1-score, False-positive rate, False-negative Rate, Balanced accuracy, Matthews correlation coefficient and Cohen kappa.
- To compare the proposed model with the already existing machine learning approaches, that have been reported in related studies.

2. Related Work

Previous research has shown the ability to distinguish breast cancer cases based on cytological and clinical data using machine learning algorithms. There have been plenty of studies done with Decision trees, Artificial neural networks, Support Vector Machine, Logistic Regression, K-nearest neighbours and Ensemble models for malignant - benign discrimination. There are some differences in assumption and learning mechanism of these models. For instance, a decision tree has clear rules but can overfit; logistic regression has consistent linear decision planes but may not be as effective when interactions are complicated; a support vector machine is good at handling high dimensional spaces, but the kernel and parameters have to be carefully selected, and a random forest has low variance using bagging but may not be as informative at the level of individual features [5], [6].

Feature selection has also been investigated around this due to the fact that descriptive morphological features of diagnostic datasets are frequently correlated. Efficient filter approaches, like mutual information and Fisher score, can be employed to rank the features and then use them for classification. There are methods superior in complexity, such as wrapper methods, which estimate the relevance in terms of the relevance of learning algorithm. That is, a single feature-selection method might only capture one aspect of the predictive structure. The drawback of the current framework is that it does not consider the multiple relevance signals, but rather compounds it into a consensus score for classification later [7], [8].

3. Methodology

3.1. Research Design

This study has a supervised binary-classification design. Each record in the data set provides quantitative measurements taken from images digitized from the fine needle aspiration images of the breast masses and each record is classified as malignant or benign. The stages of the modelling pipeline are sequentially: data screening, data preprocessing, train-test split, parameters of the rank consensus feature selection, model training, probability fusion by soft voting and evaluation of diagnostic performance. The methodological logic is shown in Figure 1.

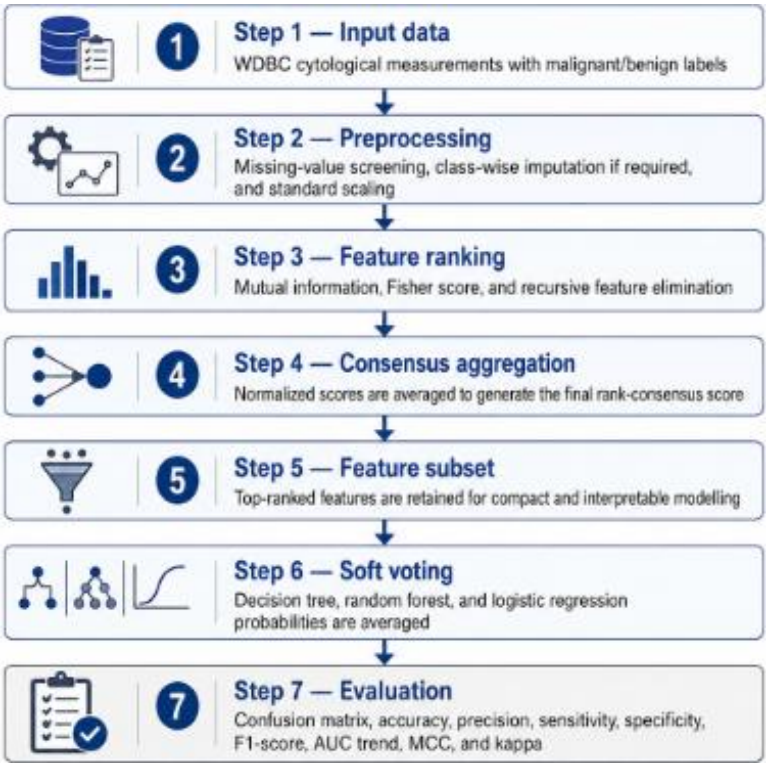

Figure 1. Proposed Rank – Consensus Ensemble Feature Selection and Soft Voting Diagnostic Framework

3.2. Dataset Description

The data set for the study is a Breast Cancer Wisconsin Diagnostic data set containing 569 labelled observations from breast mass images obtained by fine needle aspiration as mentioned in table 1. There are only two classes for the diagnostic task: "malignant" cases are considered as positive class since the goal of clinical screening is to avoid missing the diagnosis of a malignant case and "benign" cases are considered as a negative class. Morphometric variables for each observation were included and quantified the cell nuclei size, shape, texture, smoothness, compactness, concavity, concave points, symmetry, and fractal dimension. Given its structured, labelled, and widely-used nature as a benchmark data for breast cancer classification, the dataset is suitable for the evaluation of machine-learning models.

Item	Description
Dataset	Breast Cancer Wisconsin Diagnostic dataset
Number of observations	569
Diagnostic classes	Malignant and benign
Input information	Morphological measurements extracted from fine-needle aspiration images
Learning task	Binary supervised classification
Positive class in evaluation	Malignant
Reported hold-out test size	114 cases
Reported correct predictions	109 cases

Table 1. Dataset and Evaluation Summary.

3.3. Data Preprocessing and Leakage Control

Data preprocessing was not considered as a separate process, but as an integral part of the whole modelling process. This is important, to avoid any spilling of data in testing set to the training process, through scaling or imputation. Initial screening of the dataset was done by checking for missing and abnormal value [9]. In case of missing values, the values should be replaced by class-wise mean in the training partition and the same imputation parameters can be used in the test partition. Second, to overcome the sensitivity to feature scale of several classifiers, the continuous variables were standardized. The normalization was done as follows:

$$z_{ij} = (x_{ij} - \mu_j) / \sigma_j \quad (1)$$

where x_{ij} is the original value of feature j for observation i, μ_j is the mean of feature j for the training set and σ_j is the standard deviation of feature j for the training set. This transformation ensures the features have mean values around 0 and variance around 1 so that the feature-selection and classification algorithms can perform on similar scales.

3.4. Rank Consensus Ensemble Feature Selection

The feature selection stage proposed is aimed to overcome limitation of any particular feature selector. One filter method may give a feature a high ranking because it is very related to the target class; however, it does not consider the feature's overlap with other features [10]. A wrapper method can discover features valued for a certain model but can be more susceptible to the training partition and calculation expenses. This is to overcome these deficiencies, the proposed framework adopts the rank consensus approach with three complementary selectors.

3.4.1. Mutual Information

Each feature is dependent on the diagnostic class to a greater or less degree as measured by mutual information [11], [12]. It is appropriate for breast cancer data – the relation between morphological feature and malignancy is possibly non-linear. Mutual information between a feature X_{sample} j and a class label Y is:

$$MI(X_j,Y) = \sum_{x_j \in X_j} \sum_{y \in Y} p(x_j,y) \log \left[\frac{p(x_j,y)}{p(x_j), p(y)} \right] \quad (2)$$

The greater the MI the more information the feature provides to distinguish between malignant and benign cases.

3.4.2. Fisher Score

Fisher: the measure for discriminatory ability of a feature is based on comparing the separation between classes with the dispersion within the classes. The score is calculated for feature j as:

$$F_j = \frac{\sum_{c=1}^{C} n_c (\mu_{cj} - \mu_j)^2}{\sum_{c=1}^{C} n_c \sigma_{cj}^2} \quad (3)$$

where c represents the diagnosis class, n_c is the number of diagnosis samples belonging to class c, μ_{cj} is the sample mean of feature j belonging to class c, μ_j is the mean of the samples taken from the various classes and $\sigma_{cj}^2$is the within-class variance of feature j. The higher Fisher score signifies that a feature has an excellent effect in discriminating between malignant and benign cases.

3.4.3. Recursive Feature Elimination

The wrapper method of the proposed selection method was implemented by recursive feature elimination [13]. Feature ranking was done using a logistic regression estimator on the basis of their contribution to the

supervised decision boundary. Redundant characteristic was recursively deleted till the desired subset was found. This model driven process enables the feature selector to take into account the behavior of variables inside a classifier than just as independent statistical entities [14].

3.4.4. Consensus Aggregation Rule

All scores were normalized to the range [0, 1] after getting the three outputs of feature relevance. Since MI, Fisher score and RFE result in scores which are on individual scales, a normalization is needed prior to the aggregation. The final "rank-consensus" score for feature j is given by:

$$RC_j = (S_{MI}, j + S_{Fisher}, j + S_{RFE},j) / 3 \quad (4)$$

S_{MI}, j, S_{Fisher}, j, S_{RFE},j are the normalized scores of the j-th feature from the normalized score matrix of mutual information, Fisher score and recursive feature elimination, respectively. These features were then placed in a rank based on RC_j, in descending order. The top 5 features are selected as they can represent the diagnosis in a more compact and convenient way and still have good predictive power. Worst radius, worst concave points, mean concave points, mean radius and worse perimeter were selected as feature.

Selected feature	Diagnostic interpretation
worst radius	Captures the largest observed nuclear radius and reflects abnormal nuclear enlargement.
worst concave points	Represents severe boundary irregularity and local inward contour changes.
mean concave points	Summarizes the average extent of contour indentation across cell nuclei.
mean radius	Represents the central tendency of nuclear size across sampled cells.
worst perimeter	Captures the largest boundary length and is closely related to abnormal cell morphology.

Table 2. Diagnostic Interpretation andSelected rank-consensus features.

The diagnostic interpretation of the selected rank-consensus features is presented in Table 2.

3.5. Soft Voting Classification Model

The chosen features were then input to soft voting ensemble of classifiers, which consists of 3 classifiers, decision tree, random forest and logistic regression. These models were selected as they provide complementary decision behaviours [15], [16]. Decision tree is a non-linear split based rule based approach, random forest stabilizes it by ensemble averaging and logistic regression is a linear approach treated as a stable baseline probabilistic model [17]. Soft voting allows each classifier to give an estimate of class-probability as well as just the hard class label. The final probability of malignancy is obtained as:

$$\hat{P}(y=1|x) = \frac{P_{DT}(y=1|x) + P_{RF}(y=1|x) + P_{LR}(y=1|x)}{3} \quad (5)$$

The class that has the highest averaged diagnostic label is the final label. The benefit of this is that it has been clinically helpful, as it does not rely on a single classifier, and the outputs it gives are a smoother decision boundary than offered by prediction using a single model only. In order to keep transparency equal weights were used; however, if a large external validation set is available then future versions may use validation weights for classifiers.

3.6. Model Validation and Evaluation Metrics

The performance of the final model was then assessed on a hold out test set comprising of 114 samples. Malignant was considered as the positive class as it is clinically more important. Accuracy, Precision, Sensitivity, Specificity, F1-score, Negative predictive value, False-positive rate, False-negative rate, Balanced

accuracies, Matthews correlation coefficient and Cohen kappa were calculated using confusion matrix. The important formulas are:

Accuracy = (TP + TN) / (TP + TN + FP + FN) (6)

Precision = TP / (TP + FP) (7)

Sensitivity = TP / (TP + FN) (8)

is the accuracy of negative results.

In this, the F1-score is calculated as 2 * (Precision * Sensitivity) / (Precision + Sensitivity)

Balanced accuracy is the average of sensitivity and specificity and it can be calculated as sensitivity plus specificty divided by 2.

4. Results And Discussions

4.1. Feature Selection Results

Each of the feature-selection procedures yielded complementary lists. Mutual information highlighted attributes that were highly related to the diagnostic class label such as those concerning radius, perimeter and concave points. Fisher score was used to favour strongly class separative features, particularly variables that exhibited well separated mean values for their malignant and benign distributions. Recursive feature elimination proposed a model-dependent relevance estimate, mitigating the risk of picking features that are very strong but not so powerful when included in an classifier. The rank-consensus aggregation was able to fuse these three views together to find that the top five results were worst radius, worst concave points, mean concave points, mean radius and worst perimeter.

These seem to be clinically relevant feature subsets, capturing tumour morphology from two distinct aspects, namely size related abnormality and boundary irregularity. Both of the variables "radius" and "perimeter" describe abnormal enlargement of the cell nuclei and the variable "concave-points" describes irregularity in the nuclear contours. Both the "both" in the best level descriptors and the "both" in the mean descriptors suggest that average morphology and extreme abnormality both play a role in separating the diagnostics.

4.2. Confusion Matrix based Performance

A total of 114 observations in the reported test set are available. The proposed soft-voting classifier correctly classifies 109 and has misclassification of 5 cases. If the positive class is considered to be malignant then the same performance values have been reported which refer to 40 true positives, 69 true negatives, 2 false positives, and 3 false negatives. The following is the confusion matrix (Table 3).

Actual class	Predicted malignant	Predicted benign	Total
Actual malignant	40	3	43
Actual benign	2	69	71
Total	42	72	114

Table 3. Confusion matrix of the proposed soft-voting classifier.

The model had minimum number of false positives; hence, the model rarely classified benign cases as malignant cases. More important, only three cases were missed (false-negative), and the model was able to predict few malignant cases. In diagnostic use, false negatives may be of more clinical importance, as it can lead to a delay in further investigation or treatment. In this regard, the obtained sensitivity percentage is 93.02% which is notable. Simultaneously, the specificity of 97.18% demonstrates that the accuracy of the model which can accurately identify benign cases is still good.

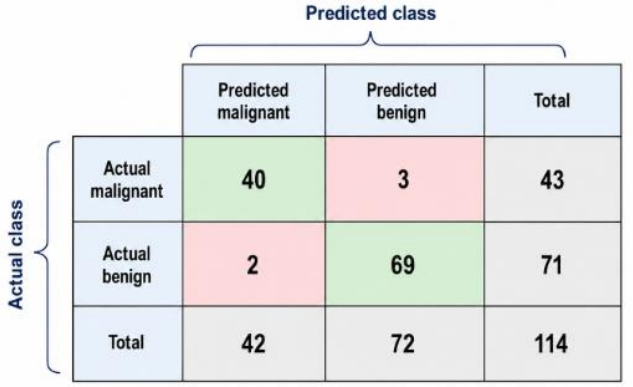

Figure 2. Confusion Matrix for Proposed Soft Voting Classifier

The confusion matrix for the proposed soft voting classifier is shown in figure 2. 40 malignant and 69 benign cases were correctly classified while only 3 malignant and 2 benign cases were misclassified as benign and malignant respectively. Overall, diagnostic performance was good with 109+114 samples correctly predicted.

4.3. Diagnostic Performance Matrics

Metric	Value	Interpretation
Accuracy	95.61%	109 correct predictions out of 114 test cases
Precision / positive predictive value	95.24%	Most predicted malignant cases were truly malignant
Sensitivity / recall	93.02%	The model detected most malignant cases
Specificity	97.18%	The model correctly recognized most benign cases
F1-score	94.12%	Balanced performance between precision and sensitivity
Negative predictive value	95.83%	Most predicted benign cases were truly benign
False-positive rate	2.82%	Low unnecessary malignant classification among benign cases
False-negative rate	6.98%	Few malignant cases were missed
Balanced accuracy	95.10%	Strong class-balanced performance
Matthews correlation coefficient	0.906	High-quality binary classification even under class imbalance
Cohen kappa	0.906	Very strong agreement beyond chance

Table 4. Extended diagnostic performance of the proposed model.

The proposed model obtained 95.61% accuracy thus validating that the 5-features tiny representation captured most of the information in the original model with diagnostic values. Precision was 95.24%, which indicates that most of the cases which were predicted as malignant were actually malignant. A high percentage of malignant cases were found with the model (sensitivity = 93.02%). The specificity was slightly higher (97.18%) indicating that the model was good for benign cases. The F1-score is 94.12%, which indicates that both the detection of malignant cases and the control of false alarms are well balanced.

The other measurements add more details than just accuracy to the results section. The multivariate model had a negative predictive value of 95.83% which means that the model was reliable in predicting a benign diagnosis. These were low false-positive and false-negative errors of 2.82% and 6.98%, respectively. The model's accuracy of 95.10% is balanced as there are: Both the Matthews correlation coefficient and Cohen kappa were about 0.906, showing that all results were good and highly compatible with the correctness of the labels is mentioned in Table 4.

4.4. Comparative Performance

Method	Accuracy (%)	Precision (%)	Sensitivity (%)	Specificity (%)	F1-score (%)
Decision Tree (C4.5) [15]	93.60	91.115	95.80	90.70	93.24
ANN (MLP) [15]	94.70	92.99	95.60	92.80	94.15
Logistic Regression [16]	92.10	95.31	91.00	93.61	93.06
KNN [16]	92.23	96.55	90.32	95.12	93.22
SVM [16]	92.78	95.94	91.07	95.14	93.37
Proposed RC-EFSV framework	95.61	95.24	93.02	97.18	94.12

Table 5. Comparative performance of the proposed method and existing methods.

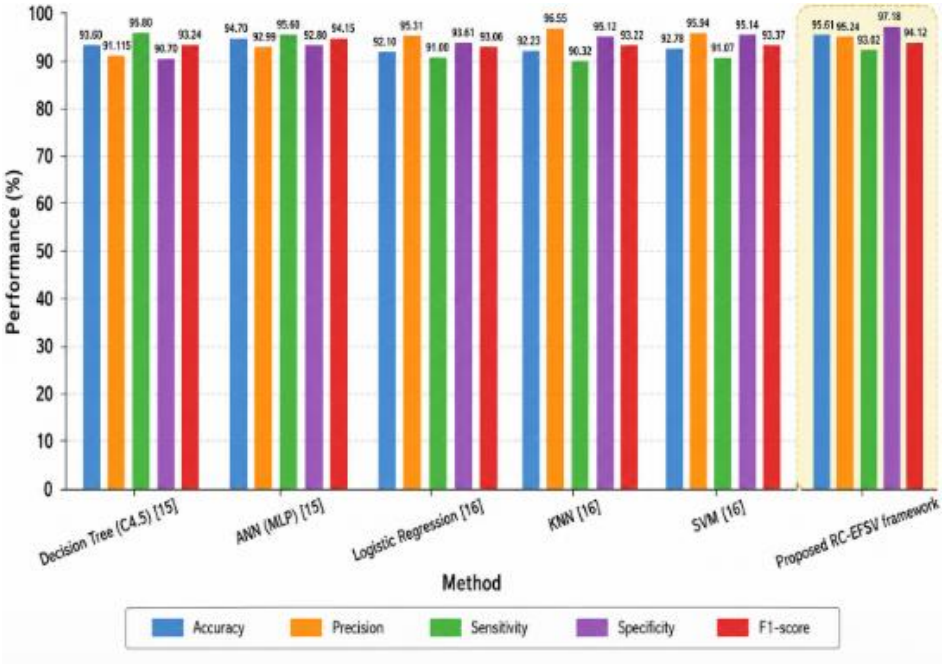

Figure 3. Comparative Performance of Breast Cancer Classification Methods

Figure 3 illustrates the comparison of the classifiers' performance with the proposed RC-EFSV framework in every classifier performance measurement including accuracy, precision, sensitivity, specificity, and F1-score reflecting the classifiers' overall diagnostic performance indicating the best performance for RC-EFSV framework, specifically in terms of accuracy and specificity.

The accuracy of the proposed RC-EFSV framework is the highest one reported in the compared methods. It was found to be more accurate than the Decision Tree, logistic regression, KNN, and SVM baselines and slightly better than the ANN one. As indicated in the comparison table, the ANN was slightly more sensitive and F1-score was slightly higher than the proposed method, but the specificity and overall accuracy was higher, and the feature subset was smaller and easier to interpret than the proposed method. This is crucial, as clinical decision support models should not only be as predictive as possible, but also be understandable to the domain experts.

Based on the comparison, it can be shown thatClassifier fusion is not responsible for the improvement of the classification results. The rank-consensus feature-selection step probably helped to discard redundant variables and keep variables that are consistently selected via multiple selection methods. It produces, thus, a shallower model than a full-featured classifier, and a more interpretable one than a black-box model utilizing all the variables without a feature-selection mechanism.

4.5. Clinical and Methodological Interpretation

From the selected features, it is noted that size is one of the most important aspects of the classification task in addition to the contour irregularity. Worst radius and worst perimeter are just extreme nuclear enlargement and mean radius is the determination of general size distribution of cell nuclei. Mean concave points and worst concave points quantify the 'indentations' of the boundary which are usually correlated to irregular tumour morphology. Thus the model conforms to the biological approach of "cytological diagnosis", that is, large or abnormal nucleoli are indicative of malignancy.

From methodological point of view there are three attributes of the method: Firstly it is interpretable, in the sense that it compresses the diagnostic space to five features which are clinically describable. Second, it is strong, since it is based on the fusion of various feature selection logics and classifiers results. Third, it is computationally feasible as just a small feature set is adopted for the final prediction. These properties make the framework an appropriate one for computer-aided diagnosis applications where the performance of prediction, interpretability, and simplicity of implementation are all of significance.

4.6. Novelty of the Proposed Approach

The novelty of the proposed article does not lie in the fact of adopting soft voting alone, but also in conjunction with a rank-consensus feature-selection layer. These breast cancer classifiers are directly compared or a single feature selector is applied in many breast cancer classification studies. Presently, this study views feature selection as an ensemble, Preclassification decision problem. The model first queries the information about the features that are consistently important for non-linear information gain, statistical class separation, and supervised elimination. The selected features are only handed over to the soft-voting classifier, when this consensus is reached. The ensemble advances at the two levels sets on improving the interpretations and prediction stability.

The article can thus be regarded as a short diagnostic toolbox, instead of just comparing two classifiers. This framing is stronger for publication as it gives a better understanding of why the model is likely to work well, why the features chosen are significant, and how the model can be generalized to other medical classification datasets.

4.7. Limitations

The proposed framework obtained good results, but there are several limitations that needs to be recognized. First, the WDBC data set is a benchmark data set and does not capture the complete diversity of hospital dataset. Second, it is based on measurements extracted from the cytological images, not imaging data or multi modal clinical measurements. Thirdly, external validation of the schema should be performed on an independent clinical data set to show claim to generalizability. Fourth, if such an ROC graph is obtained as the result of the code, a numerical value of the ROC-AUC curve should be reported in the final manuscript besides the visual description of it. Fifth, future studies could be conducted to compare the equal weight and validation weight soft voting to see if weighting of classifiers increases the sensitivity without compromise on the specificity.

5. Conclusion

This study presents an enhanced and publication-oriented version of a breast cancer prediction framework based on rank-consensus ensemble feature selection and soft-voting classification. By combining mutual information, Fisher score, and recursive feature elimination, the method identifies a compact set of five diagnostically meaningful features: worst radius, worst concave points, mean concave points, mean radius, and worst

perimeter. The soft-voting classifier achieved 95.61% accuracy, 95.24% precision, 93.02% sensitivity, 97.18% specificity, and 94.12% F1-score on the reported test set. Extended analysis further showed a negative predictive value of 95.83%, balanced accuracy of 95.10%, Matthews correlation coefficient of 0.906, and Cohen kappa of 0.906. These findings indicate that the proposed framework provides strong classification performance while maintaining interpretability and low dimensionality. With external validation and integration into clinical workflows, the method has potential as a practical decision-support tool for early breast cancer screening.

References

[1] Wu, J., & Hicks, C. (2021). Breast cancer type classification using machine learning. *Journal of personalized medicine*, *11*(2), 61.

[2] Nemade, V., &Fegade, V. (2023). Machine learning techniques for breast cancer prediction. *Procedia Computer Science*, *218*, 1314-1320.

[3] Islam, M. M., Haque, M. R., Iqbal, H., Hasan, M. M., Hasan, M., &Kabir, M. N. (2020). Breast cancer prediction: a comparative study using machine learning techniques. *SN Computer Science*, *1*(5), 290.

[4] Khalid, A., Mehmood, A., Alabrah, A., Alkhamees, B. F., Amin, F., AlSalman, H., & Choi, G. S. (2023). Breast cancer detection and prevention using machine learning. *Diagnostics*, *13*(19), 3113.

[5] Rabiei, R., Ayyoubzadeh, S. M., Sohrabei, S., Esmaeili, M., &Atashi, A. (2022). Prediction of breast cancer using machine learning approaches. *Journal of biomedical physics & engineering*, *12*(3), 297.

[6] Chaurasia, V., & Pal, S. (2020). Applications of machine learning techniques to predict diagnostic breast cancer. *SN Computer Science*, *1*(5), 270.

[7] Mohammed, S. A., Darrab, S., Noaman, S. A., &Saake, G. (2020, July). Analysis of breast cancer detection using different machine learning techniques. In *International conference on data mining and big data* (pp. 108-117). Singapore: Springer Singapore.

[8] Wei, Y., Zhang, D., Gao, M., Tian, Y., He, Y., Huang, B., & Zheng, C. (2023). Breast cancer prediction based on machine learning. *J. Softw. Eng. Appl*, *16*(8), 348-360.

[9] Carriero, A., Groenhoff, L., Vologina, E., Basile, P., &Albera, M. (2024). Deep learning in breast cancer imaging: state of the art and recent advancements in early 2024. *Diagnostics*, *14*(8), 848.

[10] Rahman, M. A., Khan, M. S. H., Watanobe, Y., Prioty, J. T., Annita, T. T., Rahman, S., ... &Bhuiyan, T. (2025). Advancements in breast cancer detection: A review of global Trends, risk Factors, imaging Modalities, machine learning, and deep learning approaches. *BioMedInformatics*, *5*(3), 46.

[11] Manikandan, P., Durga, U., &Ponnuraja, C. (2023). An integrative machine learning framework for classifying SEER breast cancer. *Scientific Reports*, *13*(1), 5362.

[12] Zizaan, A., &Idri, A. (2023). Machine learning based Breast Cancer screening: Trends, challenges, and opportunities. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, *11*(3), 976-996.

[13] Radak, M., Lafta, H. Y., &Fallahi, H. (2023). Machine learning and deep learning techniques for breast cancer diagnosis and classification: a comprehensive review of medical imaging studies. *Journal of Cancer Research and Clinical Oncology*, *149*(12), 10473-10491.

[14] Ghavidel, A., &Pazos, P. (2025). Machine learning (ML) techniques to predict breast cancer in imbalanced datasets: a systematic review. *Journal of Cancer Survivorship*, *19*(1), 270-294.

[15] Boddu, A. S., & Jan, A. (2025). A systematic review of machine learning algorithms for breast cancer detection. *Tissue and Cell*, *95*, 102929.

[16] Zuo, D., Yang, L., Jin, Y., Qi, H., Liu, Y., & Ren, L. (2023). Machine learning-based models for the prediction of breast cancer recurrence risk. *BMC Medical Informatics and Decision Making*, *23*(1), 276.

[17] Zhu, J., Zhao, Z., Yin, B., Wu, C., Yin, C., Chen, R., & Ding, Y. (2025). An integrated approach of feature selection and machine learning for early detection of breast cancer. *Scientific reports*, *15*(1), 13015.