Research Article

International Journal of Advanced Computer Research, Vol 14(66)
ISSN (Print): 2249-7277 ISSN (Online): 2277-7970
http://dx.doi.org/10.19101/IJACR.2024.1466003

A hybrid random forest and k-nearest neighbors approach for breast cancer detection

Om Prakash Kumar[1*] and Animesh Kumar Dubey[2]
Department of Computer Science and Engineering, Patel College of Science & Technology, Bhopal, Madhya Pradesh, India[1]
Assistant professor, Department of Computer Science and Engineering, Patel College of Science & Technology, Bhopal, Madhya Pradesh, India[2]

Received: 20-October-2023; Revised: 10-March-2024; Accepted: 14-March-2024

Abstract
In this paper, a novel hybrid approach was presented combining random forest (RF) and k-nearest neighbors (kNN) for the classification of breast cancer data. RF is selected for its robustness against overfitting and its ability to handle high-dimensional data effectively, providing a measure of feature importance and generalizing well due to its ensemble nature. kNN is chosen for its simplicity and effectiveness in capturing local data patterns. Our hybrid RF-kNN approach involves feature importance weighting in kNN, dynamic k selection, polynomial feature expansion, and ensemble output combination. The Wisconsin breast cancer database (WBCD) is used for experimentation, evaluated using 10-fold cross-validation. Performance metrics include accuracy, precision, recall, and F1-score. The results demonstrate that the hybrid RF-kNN model outperforms individual models, achieving superior performance across all metrics and data splits. This highlights the robustness and effectiveness of the hybrid model in reducing false positives and correctly identifying patients with breast cancer, making it a reliable model for breast cancer detection.

Keywords
Breast cancer detection, Random forest, k-nearest neighbors, Hybrid model, Machine learning.

1.Introduction
Breast cancer remains one of the most prevalent and life-threatening cancers affecting women worldwide [1–3]. It accounts for a significant portion of cancer-related morbidity and mortality, indicating its critical importance in public health. Despite considerable advancements in medical technology and healthcare practices, early detection and accurate diagnosis of breast cancer continue to pose significant challenges [4, 5]. Traditional diagnostic methods, such as mammography and clinical breast examinations, have been pivotal in screening efforts but are not without limitations [6, 7]. These limitations include reduced sensitivity in detecting cancer in dense breast tissue and a high rate of false positives, which can result in unnecessary stress and medical interventions [8, 9]. The emergence of data mining and machine learning (ML) technologies presents transformative potential in the early detection and diagnosis of breast cancer [10, 11].

Data mining, a crucial component of this technological advancement, involves extracting valuable insights and patterns from large datasets. Often referred to as knowledge discovery in databases (KDD), data mining encompasses various functions, including data cleaning to resolve inconsistencies, pattern recognition, visualization, and rule generation [12–14]. These functions can be classified based on their capabilities. Advanced computational techniques, including linear regression, logistic regression, support vector machines (SVM), Naïve Bayes (NB), decision trees (DT), k-nearest neighbors (kNN), clustering methods (such as k-Means and fuzzy c-Means), random forests (RF), and association rule mining (Apriori), have shown remarkable capabilities in analyzing complex and voluminous datasets [13–15]. These methods can extract meaningful patterns and make highly accurate predictions, often surpassing the diagnostic capabilities of traditional methods. The integration of data-driven approaches in medical diagnostics is motivated by the need for greater

*Author for correspondence

accuracy, objectivity, and the ability to leverage extensive health data to enable personalized medicine. The main objective of this paper is to apply the hybridization of computational intelligence algorithms, specifically RF and kNN, for the classification of breast cancer data. Hybridizing RF and kNN leverages the strengths of both algorithms to enhance classification performance. RF is known for its robustness and ability to handle high-dimensional data, offering excellent accuracy and interpretability through feature importance measures. kNN, on the other hand, is simple and effective in capturing local data patterns due to its instance-based learning approach. By combining these two methods, the hybrid model can benefit from RF's global learning capabilities and kNN's local learning strengths. This synergy aims to improve the overall predictive accuracy and reliability of breast cancer classification.

The rest of the paper has been organized as follows. Literature review has been discussed in section 2. Section 3 covers the methods used and dataset for experimentation. Results have been illustrated in section 4. Finally, it is concluded and summarized in section 5.

2.Literature review

In this section our focus is to discuss and analyzes the related work to explore the methodological interventions along with the advantages and disadvantages.

In 2023, Neelima et al. [16] stated that breast cancer was considered one of the most dangerous illnesses in medical history. Early detection, though complicated due to mammography challenges, was crucial for saving lives. A system using ML techniques was developed for early detection and diagnosis. Fuzzy-based SVM and DT classification detected breast cancer with 98.2% accuracy, 97.6% precision, 96.5% recall, and 97.8% specificity. In 2024, Jain et al. [17] emphasized the urgent need for developing accurate breast cancer prediction methods. The study evaluated different SVM kernel functions and SVM classifier ensembles for their effectiveness in predicting breast cancer across various dataset sizes. It was found that SVM ensembles with linear kernels and bagging were optimal for small datasets, while RBF kernels with boosting excelled in larger datasets. This comprehensive assessment provided crucial insights into improving SVM methods for better breast cancer prediction. In 2023, Khan et al. [18] highlighted the urgent need for reliable prognostic models to manage breast cancer, which significantly affects women's health globally. The study aimed to develop a robust machine-learning model capable of distinguishing between benign and malignant breast cancer types. Evaluating five ML algorithms—XGBoost, NB, DT, RF, and Logistic Regression—the study found that XGBoost delivered the highest accuracy at 95.42%, with outstanding sensitivity, specificity, and an F1-score. These results indicate XGBoost's potential in breast cancer prediction, paving the way for further research to refine and implement this method in clinical settings. In 2023, Sun and Yang [19] emphasized the global impact of breast cancer, highlighting its prevalence and the critical need for early detection to improve prognosis and survival rates. Their research focused on the use of ML methods to accurately differentiate between benign and malignant tumors, thus preventing unnecessary treatments. The study began with an overview of various ML techniques, including linear discriminant analysis (LDA), RF classifier, and principal component regression (PCR), using the Wisconsin breast cancer database (WBCD) for data analysis. The research aimed to explore the application of these ML techniques in diagnosing and prognosing breast cancer, further introducing a healthcare system model based on recent studies. Findings demonstrated that while ML models are valuable for breast cancer diagnosis, there is a continued need for enhancing their accuracy. In 2024, Sawant et al. [20] reported that breast cancer continues to pose a significant health risk with millions of new cases identified globally each year. Their research highlighted the importance of timely and accurate diagnosis for guiding treatment options and enhancing patient outcomes. They utilized a logistic regression model to demonstrate the practical application of ML in classifying breast cancer cases. The primary objective of this study was to develop a model capable of reliably categorizing breast tissue as benign, malignant, or normal using medical images. In 2023, Vasista et al. [21] reported that cancer accounts for nearly one in six deaths globally, with breast cancer being a significant contributor. The study explored the application of ML and deep learning algorithms to assess the risk of breast cancer. The RF classifier, achieved an accuracy of 92%, the highest among the others. However, deep learning algorithms demonstrated superior performance, with the recurrent neural network (RNN) model achieving a notable 98% accuracy. Tinao et al. [22] utilized advanced ML techniques to predict breast cancer in a study involving 112 women. The research highlighted the kNN model as the most accurate, successfully classifying 86.96% of cases, with significant correlations noted between stress, family history, and cancer risk. In 2024, Kumar et al. [23] explored the predictive capabilities of ML in the early diagnosis of breast cancer, which has become increasingly common among women. The study focused on the application of ensemble ML algorithms, which enhanced performance and generalization. A comparative analysis between light gradient boosting (LightGBM) and XGBoost was conducted using a labeled dataset of breast cancer cases. The findings revealed that the XGBoost algorithm outperformed LightGBM in terms of accuracy. In 2023, Tripathi et al. [24] utilized ML algorithms with open-source data to enhance breast cancer treatment precision. Employing the Wisconsin (Diagnostic) Dataset, they tested algorithms like logistic regression, SVMs, and others to classify cancer subtypes. Logistic regression excelled, achieving 99.12% accuracy, 100% recall, and 99.87% AUC-ROC. These results demonstrate the potential of ML to refine diagnosis and treatment, ensuring more personalized care in breast cancer management. In 2023, Latha and Mahesh [25] reported that breast cancer, with the highest incidence rate globally, disproportionately affects women and is a leading cause of cancer-related deaths. Early detection can significantly lower mortality rates and improve treatment outcomes. While earlier studies utilized ML techniques like decision trees and SVMs, recent research has shifted towards deep learning, notably convolutional neural networks (CNNs), which surpass traditional ML in performance. This work reviews the evolution from ML to deep learning in breast cancer research, emphasizing CNNs' effectiveness in image classification. In 2023, Manjunathan et al. [26] highlighted breast cancer as one of the most prevalent cancers in women globally, noting the need for improved detection methods and standardized data processes. Their research focused on developing and evaluating ML models using mammograms and other medical imaging for diagnosis, recurrence prediction, and treatment planning. ML algorithms have shown potential in detecting subtle changes in breast tissue indicative of cancer, enhancing diagnostic accuracy and specificity. This advancement aids clinicians in decision-making. However, the study suggests further research is needed to overcome ML limitations and standardize data collection and analysis in breast cancer diagnosis.

3.Methods

In this paper the combination of random forest (RF) and k-nearest neighbors (kNN) have been considered for the classification of the breast cancer data. RF was chosen for its robustness against overfitting and its ability to handle high-dimensional data effectively. It provides a measure of feature importance and generalizes well due to its ensemble nature, making it suitable for complex classification tasks like breast cancer detection. kNN has been selected for its simplicity and effectiveness in capturing local data patterns. Its performance in classification can be very intuitive and accurate, especially when combined with a weighting scheme that respects feature importance.

The working process of our hybrid RF-kNN approach are as under:

1. Feature importance weighting in kNN

The RF provides a measure of feature importance, I(fi), for each feature fi. These values are used to modify the traditional Euclidean distance in kNN. The weighted distance d between two data points x and z in the feature space is defined as (Equation 1):

$$d(x,z) = \sqrt{\sum I(fi) \times (x_i - z_i)^2} \qquad (1)$$

Where n is the number of features, x_i and z_i are the values of the i-th feature for data points x and z, respectively.

2. Dynamic k selection

The optimal number of neighbors, k, is selected based on minimizing the cross-validation error across different subsets of the training data. The error for a given k can be calculated using (Equation 2):

$$E(k) = \frac{1}{m}\sum_{j=1}^{m} Loss\ (y_j, \ddot{y}_j(k)) \qquad (2)$$

Where m is the number of validation samples, y_j is the actual label, and $\ddot{y}_j(k)$ is the predicted label using k neighbors.

3.Polynomial feature expansion

Selected features Xs undergo a polynomial transformation, expanding the feature set to include non-linear interactions. This can be expressed as (Equation 3):

$$X_{expanded} = [1, Xs, X_s^2, X_s^3, \dots \dots \dots . X_s^d] \qquad (3)$$

Where d is the degree of the polynomial expansion.

4.Ensemble output combination

The final prediction is an ensemble of outputs from both RF and kNN. If $p_{RF}(x)$ and $p_{kNN}(x)$ are the predicted probabilities of breast cancer for a data point x from Rf and kNN, respectively, then the combined prediction p(x) is (Equation 4):

$$p(x) = \alpha \times p_{RF}(x) + (1-\alpha) \times p_{kNN}(x) \qquad (4)$$

Where α is a weight that can be determined based on the relative validation accuracies of the RF and kNN models, often optimized via cross-validation.

The flowchart illustrates the working procedure of a hybrid RF and kNN approach (*Figure 1*). It begins with the selection of the dataset, followed by data normalization to ensure consistency in the scale of features. Next, feature importance and weighting are evaluated to identify and prioritize the most significant features. The optimal value of k for the kNN model is then determined. If the determination of the optimal k is unsuccessful, the process is rejected. Once the optimal k is found, ensemble predictions are created based on an optimized weight parameter α, which balances the contributions of the RF and kNN models.

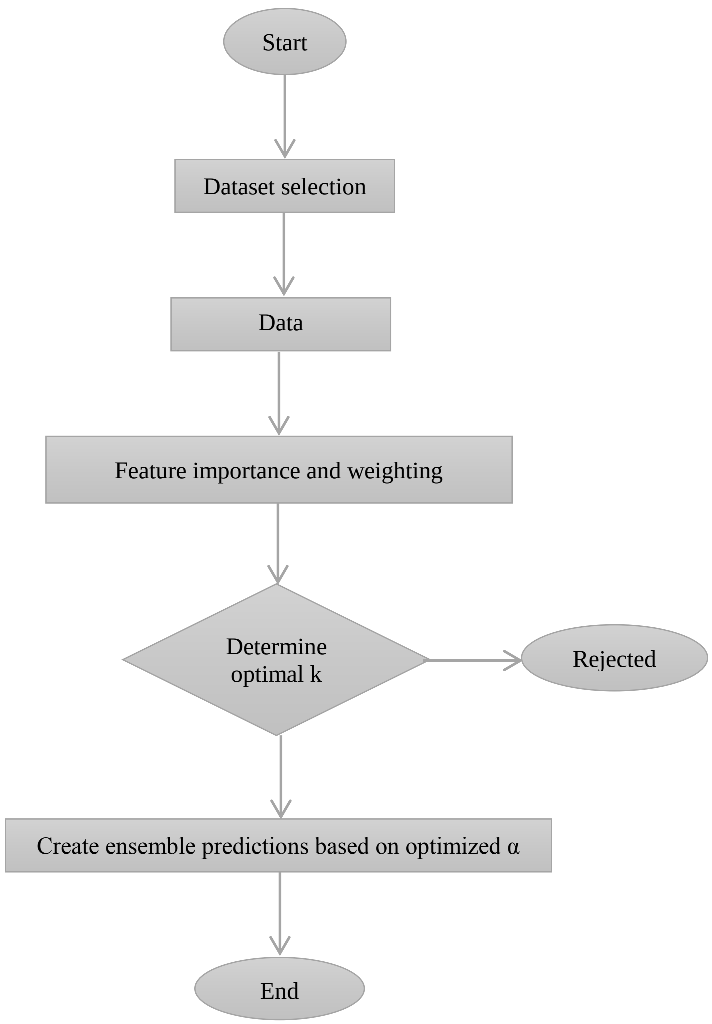

Figure 1 Flowchart of the working procedure of the hybrid RF-kNN approach

The algorithm steps are shown below.

Algorithm: Hybrid RF-kNN approach

Step1: Collect breast cancer datasets with features such as tumor size, shape, and cell characteristics.

Step 2: Normalize data to have zero mean and unit variance (Equation 5):

$$x_{norm} = \frac{x-\mu}{\sigma} \quad (5)$$

where μ is the mean and σ is the standard deviation.

Step 3: Split the dataset into training and testing sets, 70% of the data for training and 30% for testing.

Step 4: Train a RF model on the training set. Calculate feature importance for each feature.

Step 5: Select top features based on importance scores from RF.

Step 6: Apply feature weighting to kNN. Adjust the kNN distance metric to incorporate feature weights (Equation 6):

$$d(x,z) = \sqrt{\sum_{i=1}^{n} w_i \times (x_i - z_i)^2} \quad (6)$$

where w_i are weights derived from RF importance scores.

Step 7: Expand selected features using a polynomial transformation (Equation 7):

$$X_{expanded} = [X, X^2, X^3 \ldots\ldots X^d] \quad (7)$$

Step 8: Determine optimal k for kNN

Step 9: Train kNN using the modified distance metric and expanded features on the training set.

Step 10: Create ensemble predictions by averaging the outputs from both RF and kNN using Equation 4.

Step 11: Optimize α based on performance metrics on a validation set.

Step 12: Validate the hybrid model using the test dataset and perform analysis.

Step 13: Adjust parameters such as the number of trees in RF, k in kNN, and degree d in polynomial expansion based on validation feedback.

Step 14: End

4.Results

Wisconsin breast cancer database (WBCD), was used for the experimentation and evaluation. This dataset includes 699 patient records with molecular features from fine needle aspirations (FNAs), rated on a 1-10 scale for abnormality. It is available publicly on UCI repository [27]. To evaluate our analysis, 10-fold cross-validation was employed. The performance of the hybrid RF-kNN approach was evaluated and analysed using accuracy, precision, recall and F1-score. It has been computed as shown in (Equation 8, 9, 10 and 11).

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (8)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (9)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (10)$$

$$\text{F1-score} = 2\times \frac{Precision \times Recall}{Precision + Recall} \tag{11}$$

The results for precision are shown in *Figure 2*. It shows the precision results of different ML algorithms on different data splits. Precision measures the proportion of true positive predictions among all positive predictions made by the model. This indicates that the RF-kNN hybrid model outperforms the other models (RF, kNN, SVM, DT, NB) in terms of precision, regardless of the data split. The precision of the RF-kNN model remains above 98% for all splits, demonstrating its robustness and effectiveness in reducing false positives in breast cancer detection. The results for recall are shown in *Figure 3*. It shows the recall results of different ML algorithms on different data splits. Recall measures the proportion of actual positive cases that are correctly identified by the model. The RF-kNN model outperformed all in terms of recall for breast cancer detection across all tested data splits. It achieves the highest recall values, indicating its effectiveness in correctly identifying patients with breast cancer and minimizing false negatives in all splits. The results for F1-score are shown in *Figure 4*. It shows the F1-score results of different ML algorithms on different data splits. The F1-score is the harmonic mean of precision and recall, providing a balance between the two metrics. The RF-kNN model achieves the highest F1-scores for breast cancer detection across all tested data splits, indicating its superior balance between precision and recall. The results for accuracy are shown in *Figure 5*. The ensemble method used in RF-kNN helps in capturing the best aspects of both algorithms, leading to superior performance in diverse scenarios, making it less susceptible to overfitting compared to individual models. The RF-kNN model is the best performing in terms of accuracy, consistently achieving the highest values across all data splits due to its hybrid approach that effectively combines the advantages of RF and kNN. On the other hand, the kNN model is the worst performing due to its sensitivity to high-dimensional data and lack of robustness against noisy data and irrelevant features. A complete list of abbreviations is listed in *Appendix I*.

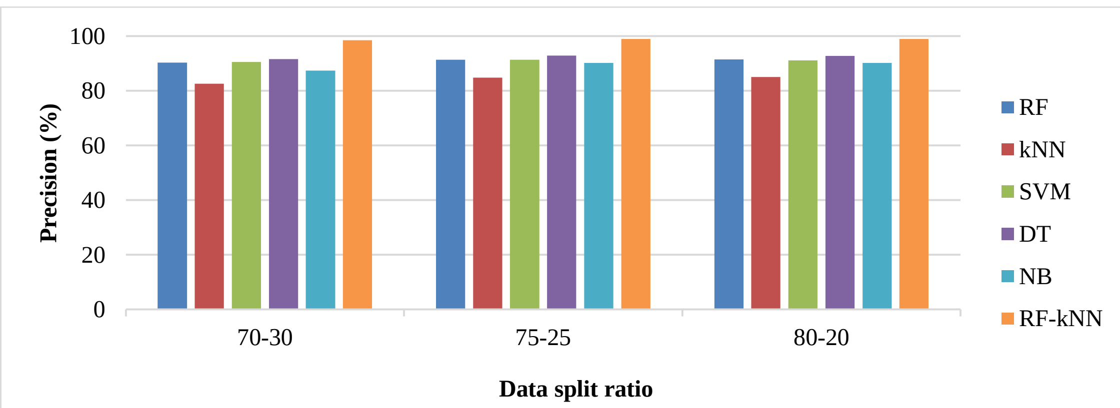

Figure 2 Precision results of different ML algorithms on various data splits for breast cancer detection

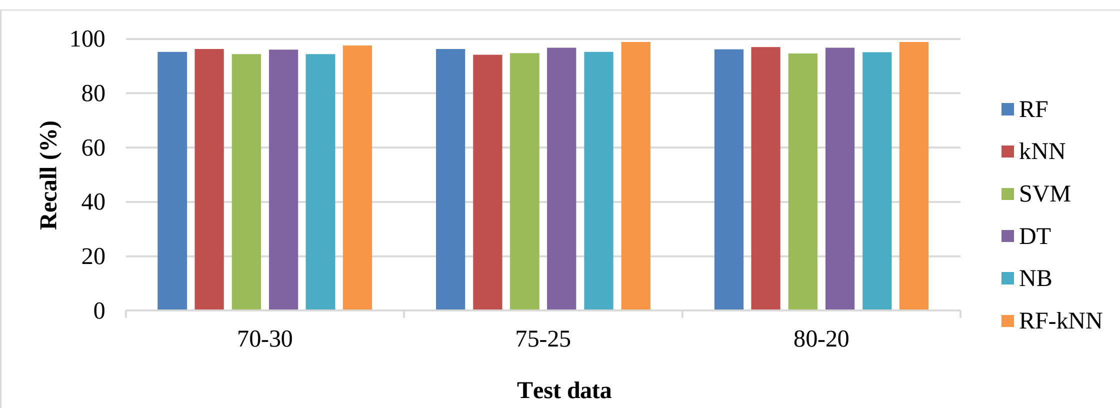

Figure 3 Recall results of different ML algorithms on various data splits for breast cancer detection

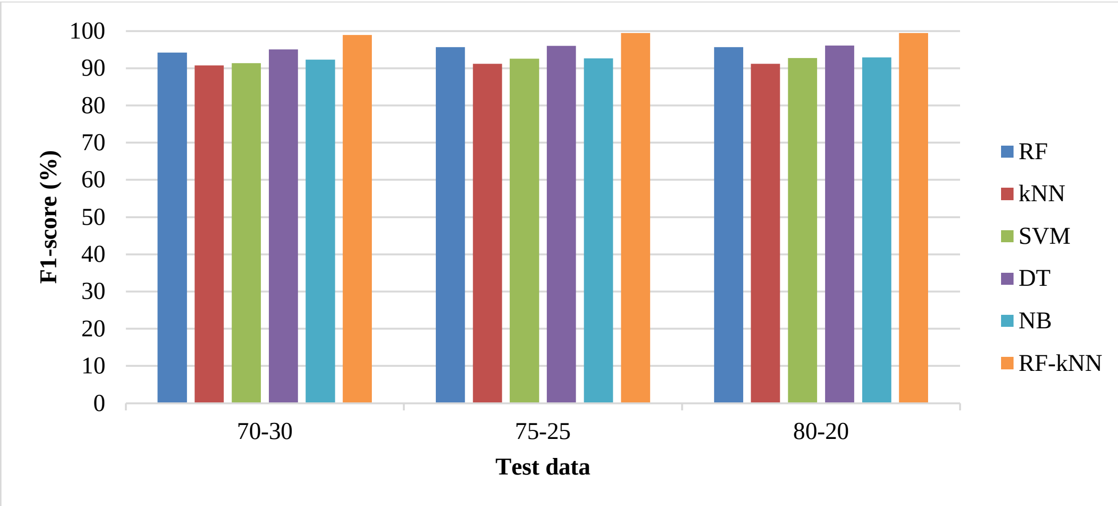

Figure 4 F1-score results of different ML algorithms on various data splits for breast cancer detection

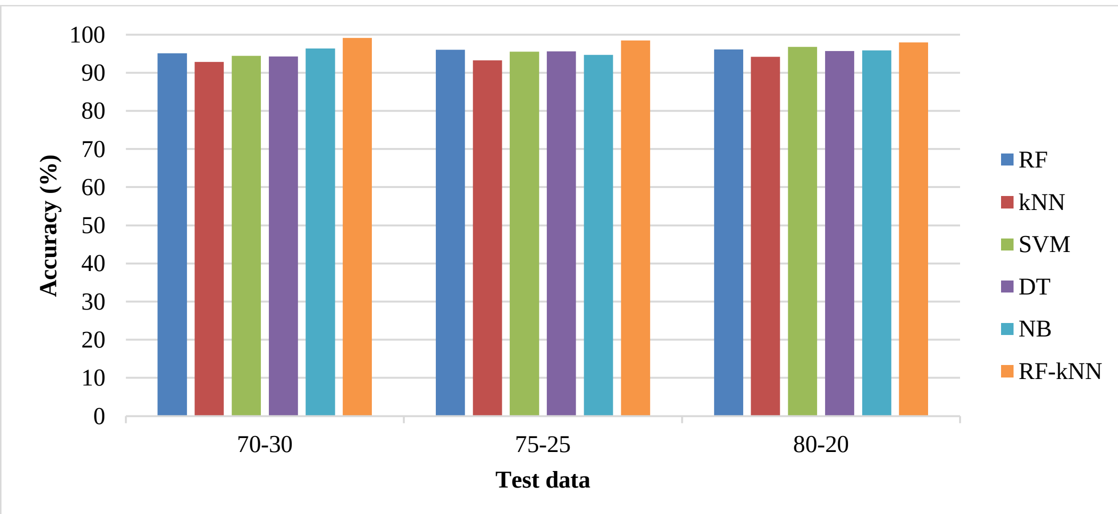

Figure 5 Accuracy results of different ML algorithms on various data splits for breast cancer detection

5.Conclusion

The hybrid RF-kNN approach for breast cancer detection combines the strengths of RF and kNN, achieving superior performance compared to individual models. RF's ability to handle high-dimensional data and provide feature importance is complemented by kNN's simplicity and local pattern recognition. The algorithm involves key steps such as feature importance weighting, dynamic k selection, polynomial feature expansion, and ensemble output combination, ensuring robust and accurate classification. Experimental results on the WBCD using 10-fold cross-validation demonstrate that the RF-kNN model consistently achieves the highest accuracy, precision, recall, and F1-score across various data splits. Specifically, the model's precision remains above 98%, recall values are highest among all tested models, and F1-scores indicate a superior balance between precision and recall. The ensemble method effectively captures the best aspects of both RF and kNN, making the hybrid model less susceptible to overfitting and more reliable in diverse scenarios. In contrast, the kNN model alone exhibits the lowest performance due to its sensitivity to high-dimensional data and lack of robustness against noisy data. The findings indicate the potential of the hybrid RF-kNN approach as a reliable and effective tool for breast cancer detection, paving the way for further research and development in medical diagnostics.

Acknowledgment

None.

Conflicts of interest

The authors have no conflicts of interest to declare.

References

[1] Ng AY, Oberije CJ, Ambrózay É, Szabó E, Serfőző O, Karpati E, et al. Prospective implementation of AI-assisted screen reading to improve early detection of breast cancer. Nature Medicine. 2023; 29(12):3044-9.

[2] Stergiopoulou D, Markou A, Strati A, Zavridou M, Tzanikou E, Mastoraki S, et al. Comprehensive liquid biopsy analysis as a tool for the early detection of minimal residual disease in breast cancer. Scientific Reports. 2023; 13(1):1258.

[3] Singh L, Alam A. An efficient hybrid methodology for an early detection of breast cancer in digital mammograms. Journal of Ambient Intelligence and Humanized Computing. 2024; 15(1):337-60.

[4] Abhisheka B, Biswas SK, Purkayastha B. A comprehensive review on breast cancer detection, classification and segmentation using deep learning. Archives of Computational Methods in Engineering. 2023; 30(8):5023-52.

[5] Gago A, Aguirre JM, Wong L. Machine learning system for the effective diagnosis and survival prediction of breast cancer patients. International Journal of Online & Biomedical Engineering. 2024; 20(2): 95-113.

[6] Hekal AA, Moustafa HE, Elnakib A. Ensemble deep learning system for early breast cancer detection. Evolutionary Intelligence. 2023; 6(3):1045-54.

[7] Khalid A, Mehmood A, Alabrah A, Alkhamees BF, Amin F, Alsalman H, et al. Breast cancer detection and prevention using machine learning. Diagnostics. 2023; 13(19):1-21.

[8] Dubey AK, Gupta U, Jain S. Analysis of k-means clustering approach on the breast cancer wisconsin dataset. International journal of computer assisted radiology and surgery. 2016; 11:2033-47.

[9] Abunasser BS, Al-hiealy MR, Zaqout IS, Abu-naser SS. Convolution neural network for breast cancer detection and classification using deep learning. Asian Pacific Journal of Cancer Prevention. 2023; 24(2):531-44.

[10] Prodan M, Paraschiv E, Stanciu A. Applying deep learning methods for mammography analysis and breast cancer detection. Applied Sciences. 2023; 13(7):1-18.

[11] Rachna, Choudhary C, Thakur J. A robust machine learning model for breast cancer prediction. Optimized Predictive Models in Healthcare Using Machine Learning. 2024:117-34.

[12] Kawina I, Amarendra K, Marapelli B. Deep learning and machine learning approach to breast cancer classification with random search hyperparameter tuning. International Journal of Intelligent Systems and Applications in Engineering. 2024; 12(16s):264-75.

[13] Saroğlu HE, Shayea I, Saoud B, Azmi MH, El-saleh AA, Saad SA, et al. Machine learning, IoT and 5g technologies for breast cancer studies: a review. Alexandria Engineering Journal. 2024; 89:210-23.

[14] Dianati-nasab M, Salimifard K, Mohammadi R, Saadatmand S, Fararouei M, Hosseini KS, et al. Machine learning algorithms to uncover risk factors of breast cancer: insights from a large case-control study. Frontiers in Oncology. 2024; 13:1276232.

[15] Lomboy KE, Hernandez RM. A comparative performance of breast cancer classification using hyper-parameterized machine learning models. International Journal of Advanced Technology and Engineering Exploration. 2021; 8(82):1080-101.

[16] Neelima G, Kanchanamala P, Misra A, Nugraha RA. Detection of breast cancer based on fuzzy logic. In international conference on advancement in data science, e-learning and information system 2023 (pp. 1-6). IEEE.

[17] Jain R, Kukreja V, Chattopadhyay S, Verma A, Sharma R. Radial basis function integrated with support vector machine model for breast cancer detection. In 2nd international conference on artificial intelligence and machine learning applications theme: healthcare and internet of things 2024 (pp. 1-5). IEEE.

[18] Khan RH, Miah J, Rahman MM, Tayaba M. A comparative study of machine learning algorithms for detecting breast cancer. In 13th annual computing and communication workshop and conference 2023 (pp. 647-52). IEEE.

[19] Sun F, Yang X. Advancing breast cancer diagnosis: a comprehensive study of machine learning algorithms on histological tumor characteristics. In international conference on data science & informatics 2023 (pp. 302-6). IEEE.

[20] Sawant A, Patil D, Khuman D, Pingle Y, Shinde V. Enhancing breast cancer detection: a machine learning approach for early diagnosis and classification. In 11th international conference on computing for sustainable global development 2024 (pp. 235-9). IEEE.

[21] Vasista VL, Sona K, Pedarla J, Sahithi B, Rao TK, Prakash KB. Predicting breast cancer using classical machine learning and deep learning algorithms. In international conference on intelligent and innovative technologies in computing, electrical and electronics 2023 (pp. 988-91). IEEE.

[22] Tinao MM, Rodriguez RB, Calibara ER. Breast cancer detection in the Philippines using machine learning approaches. In international conference on electronics, information, and communication 2024 (pp. 1-4). IEEE.

[23] Kumar R, Chaudhry M, Patel HK, Prakash N, Dogra A, Kumar S. An analysis of ensemble machine learning algorithms for breast cancer detection: performance and generalization. In 11th international conference on computing for sustainable global development 2024 (pp. 366-70). IEEE.

[24] Tripathi RP, Khatri SK, Van GD, Ather D. Unleashing the power of machine learning: a precision paradigm for breast cancer subtype classification using open-source data, with caution on dataset size and

interpretability. In 6th international conference on contemporary computing and informatics 2023 (pp. 1004-8). IEEE.
[25] Latha DU, Mahesh TR. Analysis of deep learning and machine learning methods for breast cancer detection. In international conference on computer science and emerging technologies 2023 (pp. 1-6). IEEE.
[26] Manjunathan N, Gomathi N, Muthulingam S. Early detection of breast cancer using machine learning. In international conference on sustainable computing and smart systems 2023 (pp. 165-9). IEEE.
[27] https://archive.ics.uci.edu/dataset/17/breast+cancer+wisconsin+diagnostic. Accessed 27 October 2023.

Om Prakash Kumar is currently pursuing an M.Tech in Software Systems at Patel College of Science and Technology, RGPV, Bhopal, Madhya Pradesh. He has completed his MCA at Indira Gandhi National Open University, New Delhi. His areas of interest include Machine Learning and Data Mining.
Email: kumarprakash1185@gmail.com

Dr. Animesh Kumar Dubey serves as an Assistant Professor in the Computer Science and Engineering Department at Patel College of Science and Technology Bhopal, Madhya Pradesh, India. He earned his PhD from JK Lakshmipat University, Jaipur, Rajasthan, and both his Bachelor of Engineering and M.Tech in Computer Science and Engineering from Rajiv Gandhi Technical University in Bhopal (Madhya Pradesh). Dr. Dubey has authored over 20 publications in well-regarded national and international peer-reviewed journals and conferences. His research interests encompass Data Mining, Optimization, Machine Learning, Cloud Computing, and Artificial Intelligence.
Email: animeshdubey123@gmail.com

Appendix I

S. No.	Abbreviation	Description
1	DT	Decision Trees
2	KDD	Knowledge Discovery in Databases
3	kNN	k-Nearest Neighbors
4	LDA	Linear Discriminant Analysis
5	LightGBM	Light Gradient Boosting
6	ML	Machine Learning
7	NB	Naïve Bayes
8	PCR	Principal Component Regression
9	RF	Random Forest
10	RNN	Recurrent Neural Network
11	SVM	Support Vector Machines
12	WBCD	Wisconsin Breast Cancer Database