An Optimized Ensemble Machine Learning Framework for Multi-class Classification of Date Fruits by Integrating Feature Selection Techniques

V. V. R. Maheswara Rao[1](✉), N. Silpa[1], Shiva Shankar Reddy[2], S. Mahaboob Hussain[3], Sridevi Bonthu[3], and Padma Jyothi Uppalapati[3]

[1] Shri Vishnu Engineering College for Women (A), Bhimavaram, India
mahesh_vvr@yahoo.com, nrusimhadri.silpa@gmail.com
[2] S.R.K.R. Engineering College (A), Bhimavaram, India
[3] Vishnu Institute of Technology, Bhimavaram, India
mahaboobhussain.smh@gmail.com

Abstract. Date fruits are widely consumed and valued for their nutritional properties and economic importance. Accurate classification of date fruits is crucial for quality control, sorting, and grading processes in the food industry. However, the classification of date fruits poses several challenges due to variations in their external visible physical attributes and the presence of multiple cultivars. To address the issue of high-dimensional feature space, the authors, in this research study propose an optimized ensemble machine learning framework for multi-class classification of date fruits by integrating feature selection techniques. The objective is to develop a reliable and efficient system that can accurately classify date fruits into different classes based on their physical attributes. Initially, the framework applies various feature selection techniques to identify the most relevant and discriminative features for classification. Next, it employs variants of ensemble machine learning techniques to perform multi-class classification. We utilize popular ensemble methods such as Boosted Trees, Bagged Trees, RUSBoosted Trees and Optimized Ensemble to improve the accuracy and optimization of the classification model. By integrating feature selection techniques, the proposed optimized ensemble classifier effectively handles the challenges of high-dimensional feature space and variations in date fruit characteristics. The results confirm the superiority of the proposed framework, highlighting its potential for practical applications in the food industry.

Keywords: Date Fruit Classification · Machine Learning · Feature Selection · Ensemble Classifier · Neural Networks

1 Introduction and Related Work

Date fruits are widely consumed and recognized for their nutritional value and health benefits. The accurate classification of date fruits [1–4] into different categories is crucial for quality control, marketability, and efficient supply chain management. However,

P. Pareek et al. (Eds.): IC4S 2023, LNICST 536, pp. 12–27, 2024.
https://doi.org/10.1007/978-3-031-48888-7_2

the process of manually classifying date fruits can be labor-intensive, time-consuming, and prone to human errors. Therefore, there is a need for automated systems that can effectively classify date fruits into multiple classes.

In recent years, machine learning techniques [5–12] have shown great potential in solving complex classification problems [13–22]. These techniques enable computers to learn from data and make accurate predictions or decisions. In the context of date fruit classification, machine learning algorithms can analyze various features of date fruits, such as size, color, texture, and shape, to differentiate between different classes. This research paper presents a robust ensemble machine learning framework specifically designed for the multi-class classification of date fruits. The framework integrates feature selection techniques to enhance the classification performance by reducing the dimensionality of the data and focusing on the most informative features.

Feature selection [23–27] is a critical step in the classification process as it helps identify the most relevant features that contribute to accurate classification. By eliminating irrelevant or redundant features, the framework can improve computational efficiency and mitigate the curse of dimensionality of the classification models. Various feature selection techniques, including filter methods, wrapper methods, and embedded methods, are explored and integrated into the framework.

The ensemble learning approach [28, 29] is adopted to further enhance the classification accuracy and robustness of the framework. Ensemble learning combines multiple base classifiers to make collective predictions, leveraging the diversity of classifiers to achieve better overall performance. Different ensemble techniques are investigated and integrated into the framework.

To evaluate the effectiveness of the proposed framework, a comprehensive dataset of date fruit images is collected, encompassing different varieties, ripeness stages, lighting conditions, and imaging angles. The dataset is carefully labelled and pre-processed to ensure high-quality data for training and testing the classification models. The experimental results demonstrate the superiority of the proposed framework in accurately classifying date fruits into multiple classes. The integration of feature selection techniques significantly improves the efficiency and effectiveness of the classification process by selecting the most informative features. The ensemble learning approach further enhances the performance by leveraging the collective intelligence of multiple classifiers. The framework's robustness is also evaluated by conducting experiments with different base classifiers and ensemble techniques.

The application of this research extends beyond date fruit classification. The proposed framework can be adapted and applied to other multi-class classification problems in various domains. By providing an automated and accurate solution for date fruit classification, the framework contributes to improving quality control and management in the date fruit industry.

The remaining portion is organized as follows. Section 2 describes the proposed optimized date fruit multi-class classification framework. Section 3 discusses the experimental results and evaluation of our suggested framework. Finally, Sect. 4 finishes the work with concluding observations.

2 Proposed ML Based Optimized Ensemble Date Fruit Classifier (ML-OEDFS)

The primary focus of this research paper is the introduction of a novel Machine Learning based Optimized Ensemble Date Fruit Classifier (ML-OEDFC) for accurately classifying date fruits into different categories.

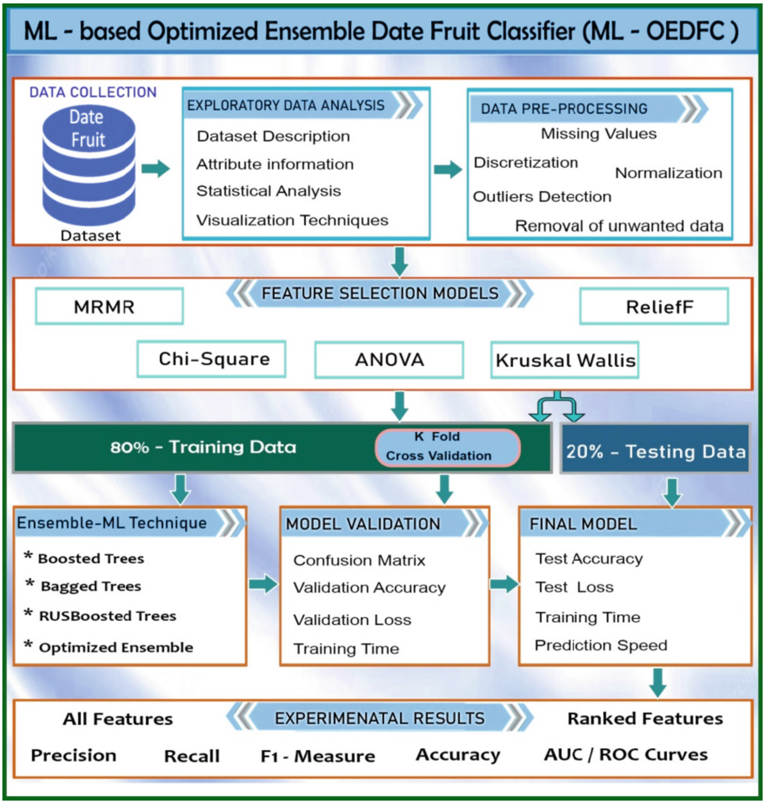

Fig. 1. Architecture of Proposed ML-RVR Model

The proposed system presented in Fig. 1, integrates feature selection techniques to enhance the classification performance by reducing data dimensionality and emphasizing the most informative features. The overarching goal of this work is to develop an automated system that effectively categorizes date fruits, leading to improved quality control, marketability within the date fruit industry.

2.1 ML-OEDFC: Data Collection and Exploration of Dataset

The proposed ML-RDFC leverages the Muratkoklu Date Fruit dataset repository [1] for research purposes to gain insights into the characteristics and relationships between different dry fruits. The dataset consists of 34 attributes that can be classified into seven distinct classes. In the research investigations, data collection is conducted to analyze the attributes associated with dry fruit category identification systems.

The model created visual representations, including Pie chart and correlation matrix to explore the data further. It is observed a normal distribution and most date fruits had a similar size range as presented in Fig. 2.

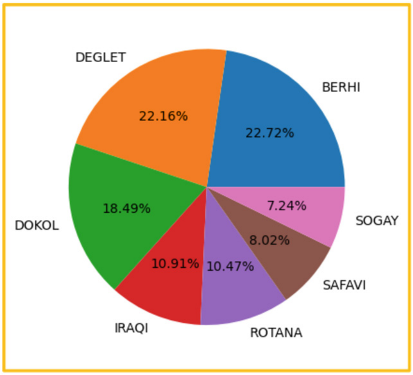

Fig. 2. Distribution of Data across all varieties of Date Fruits

Figure 3 displays the correlation matrix of the morphological attributes employed in this study. The matrix effectively visualizes the interconnectedness and interdependencies among the different morphological attributes utilized for identifying Date fruit

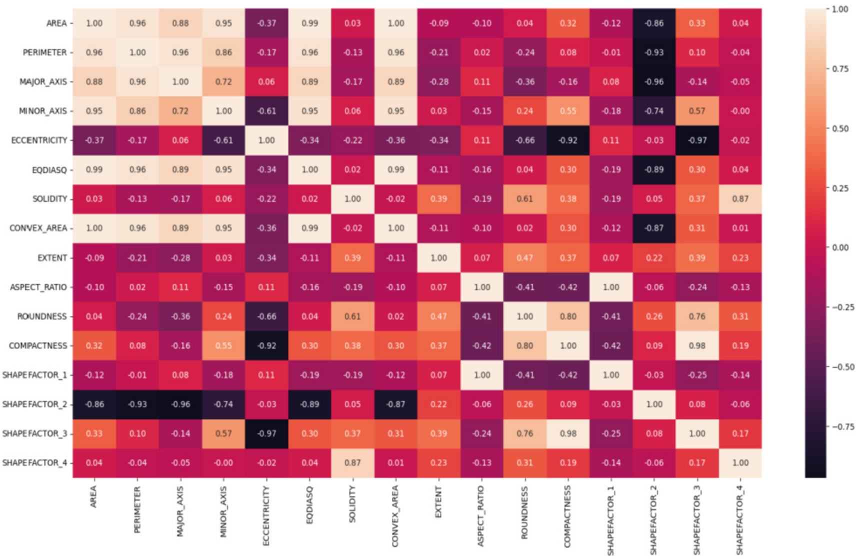

Fig. 3. Correlation Matrix of Morphological Attributes

varieties. Upon careful analysis of this matrix, it becomes evident that there exists a substantial and consistent association between the various qualities.

2.2 ML-OEDFC: Data Pre-processing

The collected dataset has already undergone stages of date fruits images collection, rotation, flipping and scaling to enhance the quality of dataset. In addition, the researchers performed resizing, normalization, noisy reduction, and edge detection to extract relevant features. For the proposed study, the processed dataset is taken as input and split the dataset into training, validation, and testing sets to evaluate the classifier's performance.

2.3 ML-OEDFC: Feature Selection Techniques

Feature selection plays a crucial role in optimizing the performance of a Date fruit classifier. In this study, various feature selection techniques were explored to identify the most relevant and informative attributes for accurate classification.

One commonly used technique is the Filter approach, which relies on statistical measures to assess the relevance of each feature independently of the classifier. These techniques evaluate the relationship between the attributes and the target class, allowing the selection of features that contribute the most to classification accuracy. Another approach is the Wrapper method, which utilizes the classifier's performance as the criterion for feature selection. This method involves creating subsets of features and evaluating their impact on the classifier's accuracy. Embedded methods integrate feature selection within the learning algorithm itself. They select features during the model training process based on their importance in improving the model's performance. Additionally, hybrid approaches that combine multiple feature selection techniques have been proposed. These approaches leverage the strengths of different methods to overcome their individual limitations, resulting in improved feature subsets and classification performance.

By applying these feature selection techniques to a Date fruit classifier, the proposed study aim to identify the most discriminative attributes that contribute significantly for accurate classification. An algorithm for performing feature selection using ML-OEDFC framework and ranking the features based on different feature selection methods. The algorithm also identifies the best feature selection method for a given classifier.

Algorithm: ML-OEDFC Feature Selection and Classifier Evaluation.

Input: Selected Date Fruits Dataset.

Classifiers - Boosted Trees(BT), Bagged Trees(BGT), RUSBoosted Trees (RUSBT), Optimized Ensemble (OE)

Output: Ranked Features by Feature Selection Methods and Best suitable Feature Selection Method for each classifier.

Step 1. Load the Selected Date Fruits Dataset into memory.
Step 2. Preprocess the dataset (e.g., handle missing values, discretization, normalization, encode categorical variables, outliers detection etc.).
Step 3. Split the dataset into input features (X) and target class(C).

Step 4. Initialize an empty dictionary to store the feature scores for each feature selection method.

Step 5. For each feature selection method in the list of feature selection methods (MRMR, Chi2, ReliefF, ANOVA, Kruskal Wallis), do the following steps:
   a. Apply the specific feature selection method to the input Date Fruits features X and target class C.
   b. Retrieve the feature scores or rankings obtained from the fea-ture selection method.
   c. Store the feature scores or rankings in the dictionary with the feature selection method as the key.

Step 6. Rank the features based on their scores obtained from each feature selection method.

Step 7. For each feature, calculate the average rank across all feature selec-tion methods.

Step 8. Sort the features based on their average rank in ascending order.

Step 9. Initialize an empty list to store the best feature selection methods for each classifier.

Step 10. For each classifier in the list of classifiers (BT, BGT, RUSBT, OE), do the following steps:
   a. Train the classifier using the Date Fruits input features selected based on the top-ranked features.
   b. Evaluate the classifier's performance using appropriate evalua-tion metrics (e.g., validation accuracy test accuracy, precision, recall, F1-score, AUC/ROC curve etc.).
   c. Store the classifier's performance metrics.

Step 11. Identify the best suitable feature selection method for each classifier based on their performance metrics.

Step 12. Return the ranked features and the best suitable feature selection method for each classifier as the output.

The algorithm begins by importing the dataset and, if necessary, preprocessing it. The system then applies each feature selection method to the dataset and stores the feature scores or rankings that result. The features are then ranked according to their average position across all selection methodologies. For each classifier, the algorithm trains the model with the most highly ranked features and evaluates its performance using the most pertinent metrics. Based on each classifier's respective performance metrics, the algorithm then determines the optimal feature selection method for each classifier.

2.4 ML-OEDFC: Training Ensemble Machine Learning Models

In ML-OEDFC, the training process involves utilizing variants of ensemble machine learning models. Ensemble models are renowned for their ability to enhance predictive performance by combining the predictions of multiple base models. ML-OEDFC takes advantage of ensemble techniques to create a robust and accurate ensemble.

ML-OEDFC employs multiple ensemble machine learning model variants to improve its performance. Bagging independently trains multiple models on distinct subsets of the training data, and then combines their predictions to produce the final ensemble out-put. Boosting trains models sequentially, with an emphasis on correcting prior errors. Combining random under-sampling with boosting, RUSBoosted Tree addresses class imbalance. It enhances classification performance and balances the dataset. Using techniques such as genetic algorithms, Optimized Ensemble optimizes the combination of base classifiers to improve overall accuracy and generalization. These variants enhance the capabilities of ML-OEDFC by minimizing overfitting, compensating for class imbalance, and optimizing ensemble performance.

The inclusion of all variants of ensemble learning models in ML-OEDFC expands the ensemble model repertoire, enabling the classifier to tackle specific challenges related to class imbalance and ensemble optimization. These variants provide additional flexibility and adaptability, allowing ML-OEDFC to deliver superior performance in diverse datasets and classification scenarios.

2.5 ML-OEDFC: K-Fold Cross Validation of Proposed Model

After training the ML-OEDFC model for classifying the date fruits with the chosen ensemble variants and feature selection techniques, it is necessary to assess its performance and generalization capabilities. One common strategy used for this purpose is k-fold cross-validation, where the dataset is divided into k subsets. In the case of ML-OEDFC, k is set to 5 and the percentage of training data is 80.

The k-fold cross-validation procedure begins with the dataset being divided into k folds of equal size. The model is then trained and evaluated k times, with a different fold serving as the validation set and the remaining folds serving as the training set each time. This enables a comprehensive evaluation of the performance of the model across various subsets of the data.

For each iteration of k-fold cross-validation, the ML-OEDFC model is trained on the training set and the corresponding validation set is used to calculate the performance metrics. Depending on the classification assignment, the metrics typically include accuracy, precision, recall, F1-score, or any other pertinent evaluation measures. At the conclusion of k-fold cross-validation, the performance metrics from each iteration are averaged to produce a more accurate estimation of the model's performance. This mitigates any bias imposed by a single validation set. It provides insight into the model's ability to generalize to new data and aids in choosing the opti-mal ensemble variant and feature selection method combination.

2.6 ML-OEDFC: Testing with Final Model

After completing the ML-OEDFC validation procedure using k-fold cross-validation, the model is evaluated using an independent test dataset. This dataset is distinct from the training and validation sets and is used to evaluate the efficacy of the model. The remaining 20% of the dataset is utilized to assess the ML-OEDFC model, and its predictions are compared against the actual or reference labels. Various evaluation metrics, including accuracy, precision, recall, and F1-score, are calculated to quantify the performance of the model on the test set. These metrics provide an objective evaluation of the model's classification accuracy for date fruits.

The ML-OEDFC model has effectively learned the underlying patterns and characteristics of the data if it achieves satisfactory results on the test dataset. The model benefits from the combined predictive potential of multiple base models and the ability to zero in on the most informative features by utilizing ensemble learners and feature selection techniques. This ultimately leads to a final model that can classify date fruits with high accuracy.

3 Experimental Results and Analysis

In order to evaluate the performance of the proposed ML-OEDFC in accurately classifying date fruits, a comprehensive series of experiments was conducted, and the significant results are presented in this section. The dataset utilized in this study consists of 34 distinct features derived from three primary categories: morphological, shape, and color characteristics. These characteristics were meticulously chosen to capture useful details about date fruits and to provide an exhaustive representation of their characteristics.

Four variants of ensemble machine learning models were devised in order to facilitate the classification process. These variants were created with the intent of enhancing the functionality of the ML-OEDFC algorithm. The objective of creating multiple variants was to compare their efficacy and identify the most effective model for accurately classifying date fruits.

K-fold cross-validation was utilized to ensure robustness and evaluate the generalization capabilities of the ML-OEDFC models. This method divides the dataset into K subsets or folds, where K is the number of subsets. This division is responsible for training and validating ML models using various combinations of training and testing data. By performing cross-validation, we hope to evaluate the consistency and reliability of the proposed ML-OEDFC models across diverse data subsets.

To assess the classification performance of the proposed ML-OEDFC variants on date fruits, a confusion matrix was developed as a fundamental tool for evaluating the classification models' accuracy. The confusion matrix shown in Table 1 provides an exhaustive examination of the predictions made by the models, allowing for a thorough analysis of the classification outcomes. Overall, the data show that different algorithms perform differently for each rice varieties, and the method used has a considerable impact on classification accuracy. The "BT" method produces good results for most date fruit varieties, and the "OE" algorithm performs very well across numerous date fruit types.

Table 1. Confusion Matrix for Classifying Date Fruit Types using ML-OEDFC Classifiers.

Date Fruit	Algorithm	BERHI	DEGLET	DOKOL	IRAQI	ROTANA	SAFARI	SOGAY
BERHI	BT	41	0	0	10	2	0	2
	BGT	41	0	0	8	1	0	1
	RUSBT	37	0	0	7	2	0	2
	OE	38	0	0	9	0	0	0
DEGLET	BT	0	51	9	0	2	0	8
	BGT	0	55	10	0	1	1	13
	RUSBT	0	55	13	0	1	0	10
	OE	0	54	8	0	1	0	9
DOKOL	BT	0	11	153	0	0	0	1
	BGT	0	9	153	0	0	0	0
	RUSBT	0	8	148	0	0	0	0
	OE	0	8	155	0	0	0	0
IRAQI	BT	9	1	0	44	1	0	2
	BGT	9	0	0	48	1	0	1
	RUSBT	12	0	0	49	1	0	2
	OE	11	0	0	48	0	0	0
ROTANA	BT	1	1	0	3	124	0	3
	BGT	1	1	0	1	125	0	3
	RUSBT	1	1	0	2	123	0	2
	OE	0	1	0	1	128	0	2
SAFARI	BT	0	0	2	1	0	156	1
	BGT	0	0	0	1	0	157	1
	RUSBT	0	0	1	0	0	156	1
	OE	0	0	0	0	0	158	1
SOGAY	BT	1	14	0	0	4	3	58
	BGT	1	13	1	0	5	1	56
	RUSBT	2	14	2	0	6	3	58
	OE	3	15	1	0	4	1	63

A number of tests were carried out in order to assess the efficacy of different ensemble machine learning approaches in categorizing date fruits. The findings, which are summarized in Table 2, reveal that the "OE" approach outperforms alternative ensemble ML techniques. It exhibits greater validation and test accuracy shown in Fig. 4, indicating that it is useful in properly classifying date fruits.

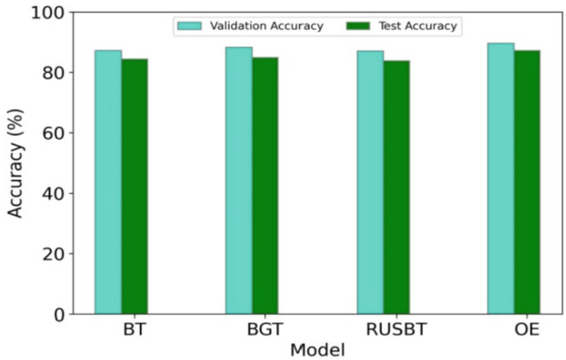

Fig. 4. Validation and Test Accuracy of various ensemble ML Techniques

Table 2. Performance of Ensemble ML Techniques

Ensemble ML technique	Training time (Sec)	Validation accuracy (Sec)	Test accuracy (Sec)
BT	6.72 s	87.2	84.4
BGT	3.64 s	88.3	84.9
RUSBT	2.76 s	87.1	83.8
OE	11.82 s	89.6	87.2

The ML-OEDFC algorithm performed an experiment using test data and generated ROC curve shown in Fig. 5. By plotting the true positive rate against the false positive rate, the ML-OEDFC algorithm provides a visual representation of its classification performance in distinguishing between different classes in the test data.

To leverage the advantages of feature selection methods for outperformed OE classification model, the proposed approach incorporates MRMR, Chi2, ReliefF, ANOVA, and Kruskal methods. Through experimentation, the rank of each feature is determined using these algorithms, and the results are tabulated in Table 3. This analysis enables the assessment of the significance and relevance of each feature, facilitating the optimization of classification performance.

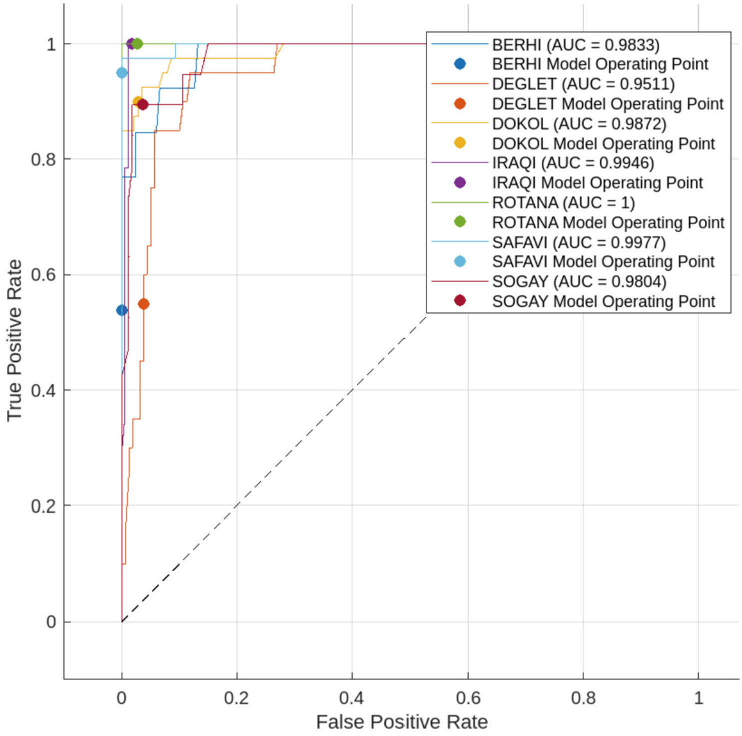

Fig. 5. ROC Curve on Test Data

With the employed feature selection methods, no. of experiments is conducted to find the impact on performance of classifiers mainly in terms of accuracy. The results are tabled in Table 4 and graph representation given in Fig. 6.

According to the observed results in the table, the MRMR feature selection method consistently performs well regardless of the size of the feature subset. It obtains a high degree of accuracy, ranging from 84.4% to 87.2%, demonstrating its effectiveness in selecting pertinent features. The Chi2 method, on the other hand, demonstrates variable efficacy, with 87.2% accuracy for 30 features and 77.7% accuracy for 10 features. With accuracies ranging from 82.1% to 86.6%, the ReliefF method exhibits a relatively stable performance. ANOVA also yields promising outcomes, with 89.0% accuracy for 30 features. The efficacy of the Kruskal method is comparable, ranging from 78.2 to 86.6%. Across a variety of feature subset sizes, the MRMR and ANOVA methods stand out as effective feature selection strategies.

Table 3. Ranking of Attributes by Feature Selection Algorithms

S. no	Feature	MRMR	Chi2	ReliefF	ANOVA	Kruskal
1	AREA	R28	R04	R02	R04	R04
2	PERIMETER	R17	R06	R05	R06	R07
3	MAJOR_AXIS	R11	R10	R08	R07	R08
4	MINOR_AXIS	R25	R01	R03	R02	R01
5	ECCENTRICITY	R33	R19	R24	R23	R22
6	EQDIASQ	R20	R05	R04	R01	R03
7	SOLIDITY	R15	R27	R33	R33	R30
8	CONVEX_AREA	R07	R03	R01	R03	R05
9	EXTENT	R08	R34	R32	R32	R34
10	ASPECT_RATIO	R23	R18	R21	R25	R21
11	ROUNDNESS	R05	R17	R26	R26	R26
12	COMPACTNESS	R12	R21	R20	R22	R24
13	SHAPEFACTOR_1	R01	R02	R09	R05	R02
14	SHAPEFACTOR_2	R34	R11	R16	R12	R09
15	SHAPEFACTOR_3	R30	R20	R19	R21	R23
16	SHAPEFACTOR_4	R03	R33	R34	R34	R33
17	MeanRR	R29	R08	R06	R09	R15
18	MeanRG	R32	R13	R10	R14	R18
19	MeanRB	R31	R23	R15	R16	R13
20	StdDevRR	R26	R30	R27	R29	R29
21	StdDevRG	R02	R32	R30	R31	R32
22	StdDevRB	R21	R31	R31	R30	R31
23	SkewRR	R19	R25	R23	R20	R20
24	SkewRG	R27	R15	R17	R10	R11
25	SkewRB	R06	R26	R25	R24	R25
26	KurtosisRR	R24	R29	R28	R28	R28
27	KurtosisRG	R10	R22	R18	R17	R19
28	KurtosisRB	R16	R28	R29	R27	R27
29	EntropyRR	R18	R07	R12	R11	R06
30	EntropyRG	R09	R12	R13	R18	R10
31	EntropyRB	R14	R16	R22	R19	R16
32	ALLdaub4RR	R13	R09	R07	R08	R14

(*continued*)

Table 3. (*continued*)

S. no	Feature	MRMR	Chi2	ReliefF	ANOVA	Kruskal
33	ALLdaub4RG	R22	R14	R11	R13	R17
34	ALLdaub4RB	R04	R24	R14	R15	R12

Table 4. Results of Classifiers on all Attributes

Feature selection method	All features	30 Features	25 Features	20 Features	15 Features	10 Features
MRMR	86.6	86.0	87.2	86.6	84.4	87.2
Chi2	86.0	87.2	84.4	86.0	84.4	77.7
ReliefF	86.0	86.6	85.5	82.7	83.8	82.1
ANOVA	88.3	89.0	86.0	82.1	81.6	81.0
Kruskal	86.0	85.5	86.6	83.8	83.8	78.2

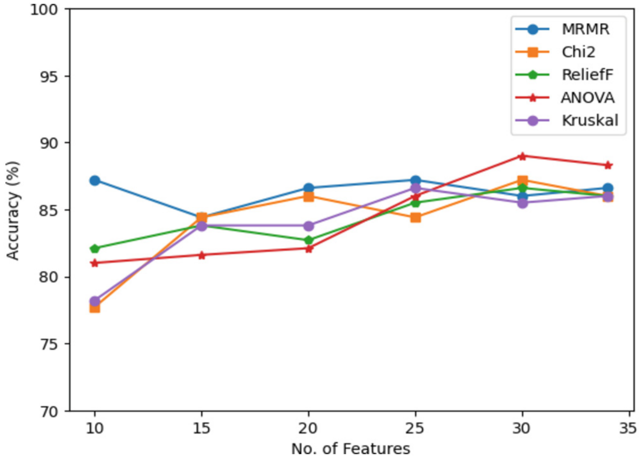

Fig. 6. Accuracy of Classifiers

The results indicate that a combination of the OE with MRMR and OE with ANOVA methods, can accurately classify date fruits. The MRMR method consistently achieves high performance across a variety of subset sizes, demonstrating its capacity to select pertinent features. In the meantime, ANOVA yields optimistic results, with an accuracy of 89.0% for 30 features. Moreover, the incorporation of feature selection methods can improve classification accuracy and also potentially reduce processing time by focusing on the most informative features, thereby enhancing the efficacy of the date fruit classification process.

4 Conclusion

In this study, a machine learning-based framework was proposed with the aim of automatically classifying date fruits with high accuracy. Specifically, variants of ensemble machine learning techniques were used, integrating feature selection techniques. The results indicated that the optimized ensemble variant performed better when incorporating the MRMR and ANOVA feature selection methods. These findings highlight the effectiveness of integrating these feature selection techniques into the ensemble framework. The proposed framework offers a promising solution for accurate and efficient classification of date fruits, with potential applications in the food industry and beyond. This resulted in the development of a software application that allowed users to categorize and obtain information about date fruits. The ML based framework could improve the success rates of classifying a variety of objects, such as date fruits, vegetables, fruits, and grains. This research endeavor not only contributes to the advancement of date fruit classification but also holds potential for broader applications in other domains requiring multi-class classification systems.

References

1. Koklu, M., Kursun, R., Taspinar, Y.S., Cinar, I.: Classification of date fruits into genetic varieties using image analysis. Math. Probl. Eng. **2021**, 1–13 (2021)
2. Khalid, S., Khalid, N., Khan, R.S., Ahmed, H., Ahmad, A.: A review on chemistry and pharmacology of AJWA date fruit and pit. Trends Food Sci. Technol. **63**, 60–69 (2017)
3. Aiadi, O., Kherfi, M.L.: A new method for automatic date fruit classification. Int. J. Comput. Vis. Robot. **7**(6), 692–711 (2017)
4. Hossain, M.S., Muhammad, G., Amin, S.U.: Improving consumer satisfaction in smart cities using edge computing and caching: a case study of date fruits classification. Futur. Gener. Comput. Syst. **88**, 333–341 (2018)
5. Liakos, K., Busato, P., Moshou, D., Pearson, S., Bochtis, D.: Machine learning in agriculture: a review. Sensors **18**(8), 2674 (2018)
6. Silpa, N., Maheswara Rao, V.V.R.: Machine learning-based optimal segmentation system for web data using a Genetic approach. J. Theor. Appl. Inf. Technol. **100**(11), 3552–3561 (2022)
7. Kovvuri, A.R., Uppalapati, P.J., Bonthu, S., Kandula, N.R.: Water level forecasting in reservoirs using time series analysis – auto ARIMA model. In: Gupta, N., Pareek, P., Reis, M. (eds.) Cognitive Computing and Cyber Physical Systems. IC4S 2022. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol. 472. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-28975-0_16
8. Maheswara Rao, V.V.R., Silpa, N., Gadiraju, M., Shankar, R.S., Vijaya Kumar, D., Brahma Rao, K.B.V.: An optimal machine learning model based on selective reinforced Markov decision to predict web browsing patterns. J. Theor. Appl. Inf. Technol. **101**(2), 859–873 (2023)

9. Reddy, S.S., Gadiraju, M., Maheswara Rao, V.V.R.: Analyzing student reviews on teacher performance using long short-term memory. In: Raj, J.S., Kamel, K., Lafata, P. (eds.) Innovative Data Communication Technologies and Application. Lecture Notes on Data Engineering and Communications Technologies, vol 96. Springer, Singapore (2022). https://doi.org/10.1007/978-981-16-7167-8_39
10. Ahmed, S.S., Uppalapati, P.J., Ayesha, S., Hussain, S.M., Narasimharao, K., Silpa, N.: Assessing public sentiment towards digital India through Twitter sentiment analysis: a comparative study. In: 7th International Conference on Intelligent Computing and Control Systems (ICICCS), pp. 955–959. IEEE (2023)
11. Silpa, N., Rao, V.M.: Enriched big data pre-processing model with machine learning approach to investigate web user usage behaviour. Indian J. Comput. Sci. Eng. [Internet] **12**(5), 1248–1256 (2021)
12. Reddy, S.S., Gadiraju, M., Preethi, N.M., Rao, V.M.: A novel approach for prediction of gestational diabetes based on clinical signs and risk factors. EAI Endorsed Trans. Scalable Inf. Syst. **10**(3), e8 (2023)
13. Cinar, I., Koklu, M.: Identification of rice varieties using machine learning algorithms. J. Agricult. Sci. **28**(2), 307–325 (2022)
14. Cinar, I., Koklu, M.: Classification of rice varieties using artificial intelligence methods. Int. J. Intell. Syst. Appl. Eng. **7**(3), 188–194 (2019)
15. Behera, S.K., Rath, A., Mahapatra, A., Sethy, P.K.: Identification, classification & grading of fruits using machine learning & computer intelligence: a review. J. Ambient Intell. Human. Comput. 1–11 (2020)
16. Tharwat, A.: Classification assessment methods. Appl. Comput. Inform. **17**(1), 168–192 (2020)
17. Oliveira, M.M., Cerqueira, B.V., Barbon, S., Barbin, D.F.: Classification of fermented cocoa beans (cut test) using computer vision. J. Food Compos. Anal. **97**, Article ID 103771 (2021)
18. Tahir, M.A.U.H., Asghar, S., Manzoor, A., Noor, M.A.: A classification model for class imbalance dataset using genetic programming. IEEE Access **7**, 71013–71037 (2019)
19. Koklu, M., Ozkan, I.A.: Multiclass classification of dry beans using computer vision and machine learning techniques. Comput. Electron. Agricult. **174**, Article ID 105507 (2020)
20. Maheswara Rao, V.V.R., Silpa, N., Mahesh, G., Reddy, S.S.: An enhanced machine learning classification system to investigate the status of micronutrients in rural women. In: Mahapatra, R.P., Peddoju, S.K., Roy, S., Parwekar, P., Goel, L. (eds.) Proceedings of International Conference on Recent Trends in Computing . Lecture Notes in Networks and Systems, vol. 341. Springer, Singapore (2022). https://doi.org/10.1007/978-981-16-7118-0_4
21. Jyothi, U.P., Dabbiru, M., Bonthu, S., Dayal, A., Kandula, N.R.: Comparative analysis of classification methods to predict diabetes mellitus on noisy data. In: Doriya, R., Soni, B., Shukla, A., Gao, X.Z. (eds.) Machine Learning, Image Processing, Network Security and Data Sciences. Lecture Notes in Electrical Engineering, vol. 946. Springer, Singapore (2023). https://doi.org/10.1007/978-981-19-5868-7_23
22. Maheswara Rao, V.V.R., Silpa, N., Gadiraju, M., Reddy, S.S., Bonthu, S., Kurada, R.R.: A plausible RNN-LSTM based profession recommendation system by predicting human personality types on social media forums. In: Proceedings of the 7th International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, pp. 850–855 (2023)
23. Vyas, A.M., Talati, B., Naik, S.: Colour feature extraction techniques of fruits: a survey. Int. J. Comput. Appl. **83**(15) (2013)
24. Korohou, T., Okinda, C., Li, H., et al.: Wheat grain yield estimation based on image morphological properties and wheat biomass. J. Sens. **2020**, Article ID 1571936, 1–11 (2020)

25. Subbarao, M.V., Sindhu, J.T.S., Reddy, Y.C.A.P., Ravuri, V., Vasavi, K.P., Ram, G.C. : Performance analysis of feature selection algorithms in the classification of dry beans using KNN and neural networks. In: Proceedings of the International Conference on Sustainable Computing and Data Communication Systems (ICSCDS), Erode, India, pp. 539–545 (2023)
26. Wang, G., Lauri, F., Hassani, A.H.E.: Feature selection by mRMR method for heart disease diagnosis. IEEE Access **10**, 100786–100796 (2022)
27. Wang, Z., et al.: Segmentalized mRMR features and cost-sensitive ELM With fixed inputs for fault diagnosis of high-speed railway turnouts. IEEE Trans. Intell. Transp. Syst. **24**(5), 4975–4987 (2023)
28. Lopes, J.F., Ludwig, L., Barbin, D.F., Grossmann, M.V.E., Barbon, S.: Computer vision classification of barley flour based on spatial pyramid partition ensemble. Sensors **19**(13), 2953 (2019)
29. Subbarao, M.V., Ram, G.C., Varma, D.R. : Performance analysis of pistachio species classification using support vector machine and ensemble classifiers. In: Proceedings of the International Conference on Recent Trends in Electronics and Communication (ICRTEC), Mysore, India, 2023, pp. 1–6 (2023)