Journal Homepage: -www.journalijar.com

INTERNATIONAL JOURNAL OF ADVANCED RESEARCH (IJAR)

ISSN NO. 2320-5407

Article DOI:10.21474/IJAR01/18164
DOI URL: http://dx.doi.org/10.21474/IJAR01/18164

RESEARCH ARTICLE

COMPARATIVE ASSESSMENT OF MACHINE LEARNING MODELS IN RICE CROP STAGE CLASSIFICATION

Shital Solanki
Assistant Professor, L.D. College of Engineering, Ahmedabad, Gujarat, India.

Manuscript Info

Manuscript History
Received: 15 November 2023
Final Accepted: 19 December 2023
Published: January 2024

Key words: -
Rice Crop Prediction, K-Nearest Neighbor (KNN), Support Vector Machine (SVM), Random Forests (RF), Decision Trees (DT), and XG Boost

Abstract

The effective classification of rice crops plays a crucial role in optimizing agricultural management and enhancing yield forecasts. In this paper, we explored the efficacy of various machine learning (ML) techniques in advancing the classification of rice crops. Four machine learning classification algorithms, namely k-nearest Neighbor (KNN), Support Vector Machine (SVM), Random Forests (RF), Decision Trees (DT), and XG Boost, are assessed using a dataset comprising rice crop images and environmental parameters. The study's findings reveal that XG Boost significantly outperforms other models, achieving an impressive accuracy of 96.78%, along with high precision and F1-Score. The Support Vector Machine also demonstrates strong performance with an accuracy of 93.83%. These findings emphasize the potential of ML in advancing agricultural practices and decision-making, highlighting the role of precision agriculture in food security.

Introduction:-

Rice, as a staple food for a significant portion of the world's population, holds paramount importance in global agriculture. As the world faces the challenges of feeding a growing population, the role of advanced agricultural technologies, particularly in rice cultivation, becomes increasingly crucial.

The accurate classification of rice crops is crucial for effective agricultural management, disease detection, and yield prediction. This study delves into the realm of machine learning (ML) techniques to explore their efficacy in classifying rice crops based on distinct growth stages, identifying the presence of diseases, and predicting yields. In this context, the classification of rice crops, a key aspect of agricultural management, plays a vital role in optimizing production, managing diseases, and predicting yields.

Recent advancements in machine learning (ML) have opened new avenues in agricultural practices, offering innovative solutions for crop classification. ML techniques, with their ability to process and analyze large datasets, are particularly well-suited for addressing the complexities of rice crop classification.

Among the ML techniques employed in this study, algorithms like k-nearest Neighbor (KNN), Support Vector Machine (SVM), Random Forests (RF), Decision Trees (DT), and XG Boost have been rigorously evaluated for their effectiveness in rice crop classification. These algorithms have shown promising results in various agricultural applications, as evidenced by their ability to handle large and complex datasets with high accuracy and efficiency. For example, XG Boost, known for its high performance in classification tasks, has demonstrated remarkable

Corresponding Author:- Shital Solanki
Address**:-** Assistant Professor, L.D. College of Engineering, Ahmedabad, Gujarat, India.

accuracy in categorizing rice crops, making it a valuable tool for precision agriculture [8]. Similarly, SVM has been widely used for its effectiveness in classification problems, particularly in scenarios with high-dimensional data (Cortes &Vapnik, 1995) [5]. This research aims to contribute to the advancement of precision agriculture and enhance global food security.

The structure of the paper is organized as follows: Section 2 provides a literature review on rice crop classification. Section 3 details the methodology and Section IV presents the results of the implementation. Finally, the paper concludes with Section V, offering concluding remarks and insights.

Literature Review:-

The field of agricultural technology, particularly rice crop classification, has witnessed significant advancements due to the integration of machine learning (ML) techniques. Recent studies in this domain have demonstrated the transformative impact of ML, overcoming the limitations of traditional methods and offering more precise, efficient solutions. Recent advancements in machine learning (ML) have significantly transformed the field of rice crop classification, as evidenced by several key studies:

Ze Lim Lee et al. presented a breakthrough in agricultural AI, particularly in rice grain classification. Utilizing a small dataset, they effectively trained convolutional neural networks (CNNs), challenging the usual need for large datasets in machine learning. Their method accurately classifies both homogeneous and heterogeneous rice grains, achieving 100% accuracy for homogeneous samples and 98.6% for heterogeneous ones. The implementation of this model on a Raspberry Pi system highlights its practicality for real-world agricultural use, offering an efficient and cost-effective solution for grain sorting. This study demonstrates the potential of neural networks in complex classification tasks with limited data, marking a significant step forward in the application of machine learning in agriculture, especially in resource-constrained environments. [1]

Moussa Diallo and colleagues analyze the impact of climate on rice yields in Mali's Office du Niger region using machine learning. They address unbalanced datasets by innovatively applying the Synthetic Minority Over-Sampling Technique (SMOTE) in stages, significantly improving classifier performance, especially the Multilayer Perceptron. This approach offers a new strategy for handling dataset imbalances in agricultural data analysis, enhancing rice yield predictions, and contributing significantly to agricultural data science. [2]

Felix Pherry and his team developed an innovative machine learning-based method for classifying the health of rice plants, addressing the limitations of traditional visual and lab inspections. Using a dataset from Kaggle, which includes images of both healthy and diseased rice plants, they tested various machine-learning models like VGG16, VGG19, ResNet50, and InceptionV3. The VGG19 model emerged as the most effective, achieving an accuracy of 84.4% and a loss rate of 55.1%. This advancement in machine learning offers a quicker and more accurate approach to disease detection in rice plants, significantly benefiting crop protection and agricultural practices. The research not only enhances disease management and yield rice improvement but also holds potential for application to other crop types, underscoring the broad impact of machine learning in agriculture. [3]

SomsawutNindam and colleagues utilized deep neural network architecture, specifically Convolutional Neural Networks (CNNs), for the multi-label classification of Jasmine Rice seed germination. They organized a dataset into three germination quality categories: excellent, good, and poor, consisting of 970 training images and 194 validation images. The team focused on classifying the seeds' germination stages and evaluating the model's performance through precision, recall, and F1 scores. The results demonstrated high accuracy in identifying the germination stages of Jasmine Rice seeds, highlighting the effectiveness of deep learning techniques in agricultural applications, particularly for accurately assessing seed germination quality. [4]

Methodology:-

The implementation details for the Cammeo and Osmancik rice varieties are presented in this section.

Dataset

The dataset used in the present study is the Rice Dataset (Cammeo and Osmancik) from Kaggle. [10] The dataset contains information and morphological features extracted from two different rice species: Osmancik and Cammeo.

Osmancik:
This species has been planted since 1997 and is described as having wide, long, glassy, and dull grains. The average weight of 1000 pieces of grain for Osmancik is in the range of 23-25 grams.

Cammeo:
This species has been grown since 2014 and is characterized by wide, long, glassy, and dull grains. The average weight of 1000 pieces of grain for Cammeo is in the range of 29-32 grams.

Among the certified rice grown in our country, the Osmancik species, which has had a large planting area since 1997, and the Cammeo species grown since 2014 have been selected for the study. When looking at the general characteristics of Osmancikpecies, they have a wide, long, glassy, and dull appearance. One thousand pieces of grain weigh 23–25 gr [11].

The dataset is comprised of the following components:
Morphological Features: These features were extracted from images of rice grains and are intended to capture various aspects of the grains' shape and characteristics. There are a total of 7 morphological features for each rice grain, which include:

1. Area: The number of pixels within the boundaries of the rice grain.
2. Perimeter: The circumference is calculated by measuring the distance between pixels around the boundaries of the rice grain.
3. MajorAxisLength: The length of the longest line that can be drawn within the rice grain.
4. MinorAxisLength: The length of the shortest line that can be drawn within the rice grain.
5. Eccentricity: A measure of how round the ellipse with the same moments as the rice grain is.
6. ConvexArea: The pixel counts of the smallest convex shell that encloses the rice grain.
7. Extent: The ratio of the region formed by the rice grain to the bounding box pixels.

Figure 1 provides a visual representation of this distribution for all 3810 rice grains obtained through image processing, offering a quick overview of distribution characteristics.

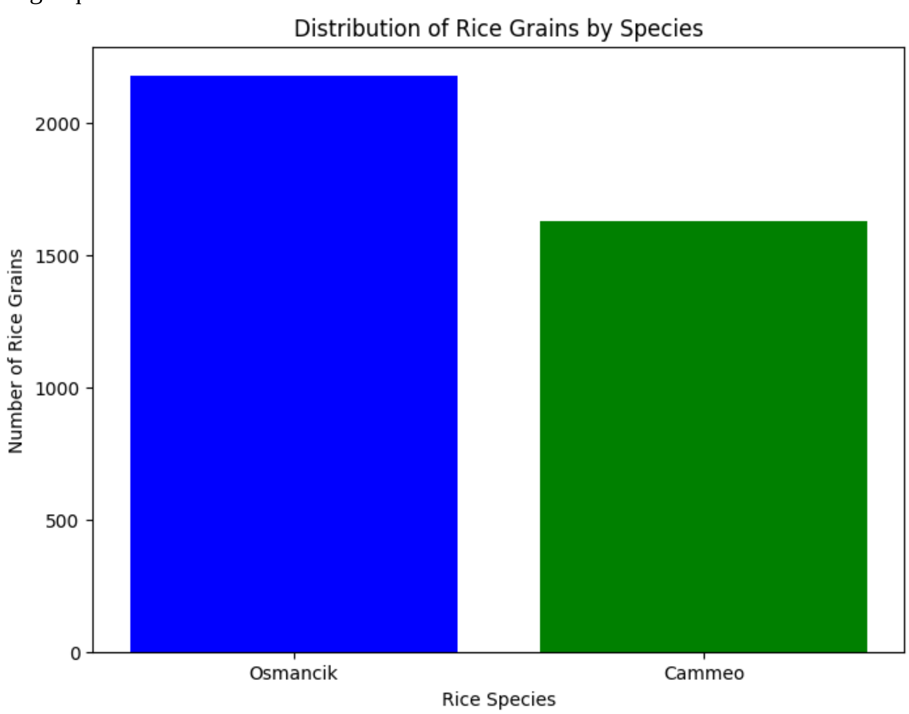

Figure 1:- Distribution of rice grains by species.

Data Preprocessing
The following preprocessing steps were applied to the rice dataset:

Data Cleaning:
An initial examination- exploratory data analysis (EDA) was conducted on the dataset to gain valuable insights and a deeper understanding of its characteristics. The dataset was found to be largely complete, with no significant instances of missing data.

Feature Selection:
The analysis focused on seven key morphological attributes of the rice grains: Area, Perimeter, Major Axis Length, Minor Axis Length, Eccentricity, Convex Area, and Extent. These features were chosen for their relevance in differentiating between the Cammeo and Osmancik rice varieties.

Data Splitting:
The dataset was split into training (70%) and testing (30%) sets

Feature Scaling:
To ensure uniformity and improve the performance of the machine learning algorithms, the feature set was normalized using the StandardScaler from the sklearn library. This step is crucial for models that are sensitive to the scale of input data.

Implementation Details
In this study, we employed five distinct classification models to analyze and classify rice grains based on their characteristics. The models used include Support Vector Machine (SVM), Decision Tree (DT), Random Forest (RF), Extreme Gradient Boosting (XG Boost), and K-Nearest Neighbor (k-NN). 10-fold cross-validation is employed for a robust evaluation of the machine learning models.

Support Vector Machine (SVM): SVM was utilized for its ability to classify data by finding the optimal hyperplane with the largest margin between classes. The model was trained using the radial basis function (RBF) kernel. Hyperparameters such as C and gamma were fine-tuned using GridSearchCV. Evaluation metrics included accuracy, precision, recall, and F1-score. [5]

Decision Tree (DT):
DT, structured like a tree, was used for classification, with branches representing classification queries and nodes indicating outcomes. The 'gini' criterion was selected for measuring the quality of splits. The tree's maximum depth was optimized. The model's performance was assessed using metrics similar to those used for SVM.[6]

Random Forest (RF):
RF, an ensemble method comprising multiple decision trees, was chosen for its enhanced accuracy. The classifier was trained with a specified number of trees. Parameters like the number of trees and maximum features were optimized. Accuracy, precision, recall, and F1-score were used for evaluation.[7]

Extreme Gradient Boosting (XG Boost):
XG Boost, known for its high performance, employs gradient boosting and regularization. It effectively handles missing data, identifies important features, and supports various loss functions. The algorithm's speed and effectiveness make it suitable for both competition and practical applications. [8]

K-Nearest Neighbour (k-NN):
k-NN, ideal for large datasets, classifies data points based on the majority class among their k nearest neighbours. Different values of K were tested to find the optimal one. The model was evaluated using the same metrics as the other classifiers. [9]

Each model was carefully selected for its unique strengths and suitability for the classification task at hand. The performance of each model was evaluated using performance metrics like accuracy, precision, and F1-score to ensure the most accurate classification of the rice grains into the Cammeo and Osmancik varieties.

Results and Discussion:-
The analysis of classification results, as illustrated in Figure 2 and elaborated in Table 1, reveals significant variations in the performance of different machine learning models. This comprehensive overview showcases the

diverse capabilities of each classifier in the task of rice grain classification. Through this detailed examination, we can draw valuable comparisons and gain deeper insights into the strengths and limitations of each model, thereby understanding their effectiveness in the specific context of classifying rice varieties.

Table 1:- Performance comparison on Rice dataset.

Sr No	Classifier	Accuracy(%)	Precision(%)	F1 Score (%)
1	XG Boost	96.78	95.43	95.98
2	Support Vector Machine	93.83	91.53	92.60
3	Random Forest	93,39	90.80	90.80
4	Decision Tree	92.49	91.29	91.00
5	K Nearest Neighbor	89.58	87.06	87.71

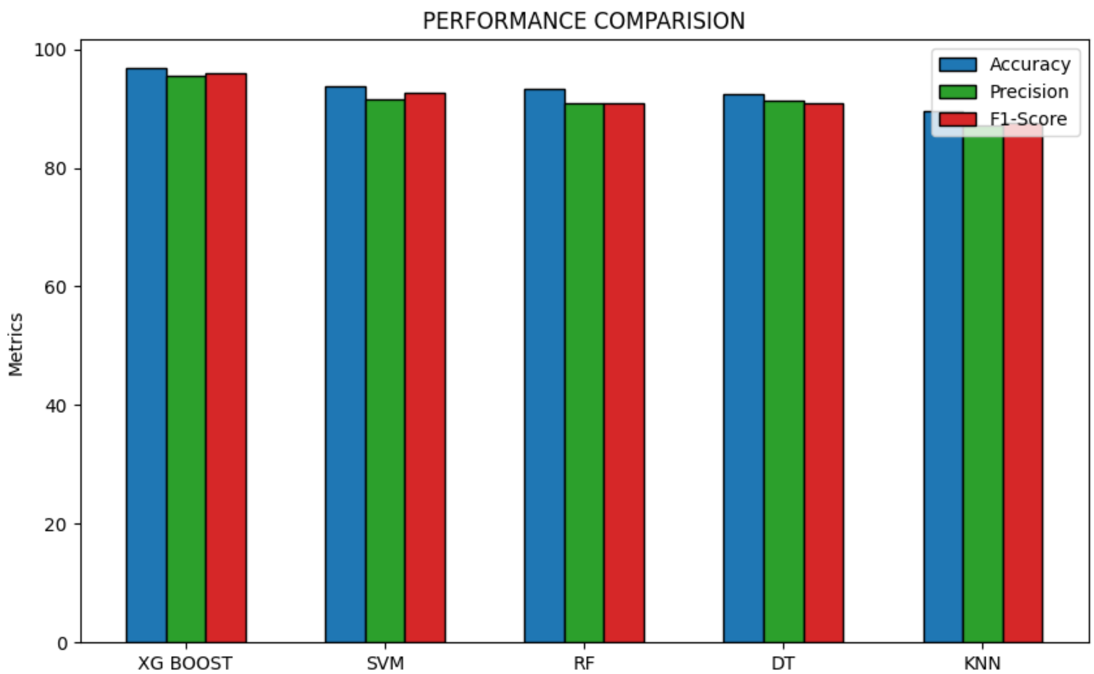

Figure 2:- Performance Comparison of Various Classifiers on Rice Dataset.

As per the result analysis, XG Boost emerges as the standout performer with an impressive accuracy rate of 96.78%, signifying its strong predictive capabilities in correctly categorizing data points. Additionally, XG Boost displays exceptional precision and F1-Score values, underscoring its ability to minimize false positive predictions while maintaining a balanced trade-off between precision and recall. While SVM (Support Vector Machine) closely follows with an accuracy of 93.83%, showcasing its effectiveness in separating and classifying data,. SVM also demonstrates robust precision and F1-Score metrics, further reinforcing its classification prowess.

Random Forest (RF) and Decision Tree (DT) models perform well, achieving high accuracy scores of 93.39% and 92.49%, respectively. While both maintain strong precision levels, their F1 scores suggest a slightly lesser balance between precision and recall when compared to the leading classifiers. Conversely, the K-Nearest Neighbour (KNN) model offers respectable performance but falls short in terms of accuracy, precision, and F1-Score. Although it does not reach the level of the top classifiers, it still delivers credible results.

Conclusion:-

This research paper has investigated the application of machine learning techniques to enhance rice crop classification. The assessment of various machine learning models, including Support Vector Machines (SVM), Random Forests (RF), Decision Trees (DT), and XG Boost, in this study has yielded insightful findings regarding their classification efficacy. Notably, XG Boost stood out as the most proficient model, exhibiting exceptional accuracy, precision, and F1-Score in the classification of rice crops. SVM also performed impressively, closely trailing XG Boost with its robust classification capabilities. While the Random Forest and Decision Tree models also showed commendable results, they were slightly outperformed in terms of F1 scores. On the other hand, the K-Nearest Neighbor model, despite being adequate, did not match the accuracy and precision levels of the leading models. The choice of the most appropriate classifier should be tailored to the specific demands of the task, with XG Boost and SVM emerging as prime candidates for scenarios requiring high accuracy and precision. Continued efforts in model optimization could further refine their practical utility in agricultural applications.

The findings of this study in rice crop classification offer substantial potential for enhancing agricultural management and decision-making, particularly in accurate crop staging, disease detection, and yield prediction. Future research should integrate diverse datasets, employ advanced machine learning methods, and apply these in precision agriculture to transform rice cultivation and bolster global food security.

References:-

1. Z. L. Lee and L. C. Tay, "Rice Grain Classification Using Convolution Neural Network with Small Dataset," 2022 Joint 12th International Conference on Soft Computing and Intelligent Systems and 23rd International Symposium on Advanced Intelligent Systems (SCIS&ISIS), Ise, Japan, 2022, pp. 1-6, doi: 10.1109/SCISISIS55246.2022.10002003.
2. Moussa Diallo, ShengwuXiong, Moussa NianiguéCoulibaly, and PrabhatRanjan Singh. 2020. Synthetic minority oversampling technique in stages for unbalanced climate and rice dataset: the Office Du Niger case study. In Proceedings of the 3rd International Conference on Telecommunications and Communication Engineering (ICTCE '19). Association for Computing Machinery, New York, NY, USA, 69–74. https://doi.org/10.1145/3369555.3369578
3. F. Pherry, Gregorius, J. Kristanto and F. I. Kurniadi, "Rice Plants Disease Classification Using Transfer Learning," 2022 4th International Conference on Cybernetics and Intelligent System (ICORIS), Prapat, Indonesia, 2022, pp. 1-4, doi: 10.1109/ICORIS56080.2022.10031449.
4. S. Nindam, T. -O. Manmai and H. J. Lee, "Multi-Label Classification of Jasmine Rice Germination Using Deep Neural Network," 2022 7th International Conference on Business and Industrial Research (ICBIR), Bangkok, Thailand, 2022, pp. 264-268, doi: 10.1109/ICBIR54589.2022.9786383.
5. Cortes, C., &Vapnik, V. (1995). Support-vector networks. Machine learning, 20(3), 273-297. DOI: 10.1007/BF00994018
6. Quinlan, J. R. (1986). Induction of decision trees. Machine learning, 1(1), 81-106. DOI: 10.1023/A:1022643204877
7. Breiman, L. (2001). Random forests. Machine learning, 45(1), 5-32. DOI: 10.1023/A:1010933404324
8. Chen, T., &Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining (pp. 785-794). DOI: 10.1145/2939672.2939785
9. Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. IEEE transactions on information theory, 13(1), 21-27. DOI: 10.1109/TIT.1967.1053964
10. Kaggle Dataset: Rice Dataset Commeo and Osmancik. (Accessed on 10 July 2023)https://www.kaggle.com/datasets/muratkokludataset/rice-dataset-commeo-and-osmancik/
11. PDD. Osmacik and Baldo Rice. 2019 08 September 2019]; Available from: http://www.pdd.org.tr/osmancik-ve-baldo-pirinc. (Accessed on 10 July 2023).