SIMATS ENGINEERING
Engineer to Excel
TECH STAR SUMMIT 2024
Name: N. Ashok Naidu
Register Number: 192110591
Guided by:Dr.D.Beulah David

Enhancing the Accuracy in Identifying Rice Species Using Support Vector Machine in Comparison With Random Forest

INTRODUCTION

- To improve the accuracy of rice species identification, this study compares two machine learning algorithms: Random Forest and Support Vector Machine (SVM).
- SVM looks for the ideal border between several rice species by employing hyperplane separation as a classification method.
- Random Forest, on the other hand, makes use of an ensemble method, combining several decision trees to provide a collective prediction that improves accuracy through variety.
- We explore each algorithm's versatility, looking at how well it can adjust to various data structures and the difficulties that come with identifying diverse species of rice.
- The project also carefully examines the algorithms' capacity to manage unbalanced datasets, which is crucial for guaranteeing accurate and impartial categorization outcomes.

Fig1:Rice Species Identification

MATERIALS AND METHODS

Data Collection
- Collection of rice species samples including grain images, morphological features, and genetic markers.

Data Preprocessing
- Cleaning and preprocessing the data to remove outliers, normalize features, and handle missing values.

Support vector machine (SVM).
- Testing various kernels like linear, polynomial, and radial basis function (RBF).Optimizing regularization parameter (C) and kernel-specific parameters.

Random Forest
- Tree Dept setting the maximum depth of trees to avoid overfitting.
- Number of Trees determining the optimal number of trees for the ensemble.

Model Evaluation
- Comparing SVM and Random Forest based on accuracy, computational efficiency, and generalization ability.

Model Comparison
- Measuring the training and prediction times for both models. Assessing the models' performance on unseen data to evaluate their generalization capabilities.

Cross-Validation
- Using k-fold cross-validation to assess model performance and prevent overfitting.

RESULTS

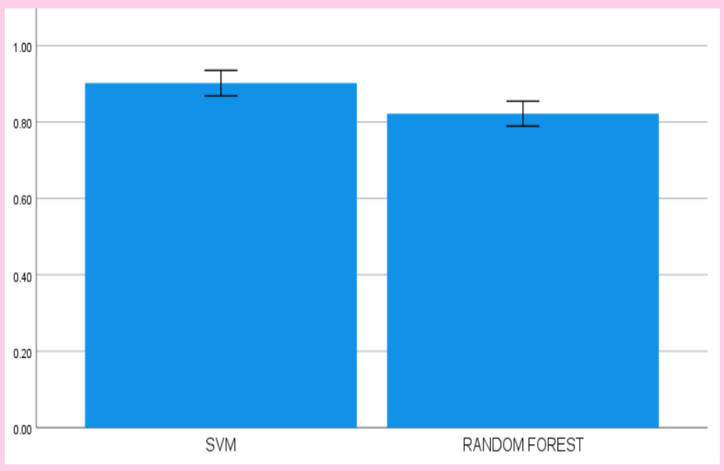

Fig 2. Support vector machine and Random Forest

- The graph represents a visual comparison of SVM and random forest models, highlighting their respective accuracy scores for evaluation.

Table 1. Comparison of the Accuracy values of RNN and LSTM Algorithms with a test size of 10

S.No	Test Size	ACCURACY RATE	
		Support vector machine Algorithm	Random forest Algorithm
1	Test 1	98.41	88.96
2	Test 2	88.85	87.83
3	Test 3	89.50	89.92
4	Test 4	89.90	89.61
5	Test 5	91.09	81.91
6	Test 6	95.76	82.93
7	Test 7	91.53	81.07
8	Test 8	91.51	81.73
9	Test 9	80.38	80.86
10	Test 10	70.47	70.91
Average Test Results		92.92	82.28

Table 2. Mean, Standard Deviation and Standard error mean with Accuracy rate Comparison of LSTM over RNN Algorithm

	Group	N	Mean	Std. Deviation	Std. Mean
Accuracy	SVM Algorithm	10	.0920	.04662	.01474
	Random forest Algorithm	10	.8220	.04566	.01444

DISCUSSION AND CONCLUSION

- The Support Vector Machine (SVM) exhibits a noteworthy accuracy rate of 92.9%, indicating its exceptional efficacy in precisely classifying various varieties of rice. Random Forest, on the other hand, only manages an accuracy of 82.3%, which is significantly lower than SVM.
- Compared to Random Forest, SVM performs better in classification, as seen by its greater precision and recall scores for both rice species classes. This suggests that SVM obtains more relevant instances of each rice species while simultaneously making fewer misclassifications.
- Even when there are external disturbances, SVM performs essentially undisturbed by noise, retaining its high accuracy. SVM's resistance and flexibility to noisy data are highlighted by the fact that Random Forest loses accuracy when noise is introduced.
- Based on its reliable and consistent performance, the results clearly support the use of SVM as the recommended algorithm for jobs involving rice species identification. For agricultural applications needing accurate classification, SVM is a dependable option due to its better accuracy and resilience.

BIBLIOGRAPHY

- Bay-Peterson, J. ed., 2021. Classification and management of rice growing soils.
- Clubb, H.S., 2019. Unpolished Rice: The Staple Food of the Orient.
- Bao, F. and Bambil, D., 2021. Applicability of computer vision in seed identification: deep learning, random forest, and support vector machine classification algorithms. Acta Botanica Brasilica, 35, pp.17-21.
- Wang, Y., Zang, S. and Tian, Y., 2020. Mapping paddy rice with the random forest algorithm using MODIS and SMAP time series. Chaos, Solitons & Fractals, 140, p.110116.
- Liu, M., Wang, M., Wang, J. and Li, D., 2013. Comparison of random forest, support vector machine and back propagation neural network for electronic tongue data classification: Application to the recognition of orange beverage and Chinese vinegar. Sensors and Actuators B: Chemical, 177, pp.970-980.