SIMATS ENGINEERING

Engineer to Excel

SAVEETHA TRANSDISCIPLINARY ANNUAL RESEARCH SUMMIT — STAR SUMMIT 2024 — SIMATS ENGINEERING

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 Gradient Boosting Machine

INTRODUCTION

- The remarkable capacity of Support Vector Machines (SVMs) to capture complicated feature linkages and set exact decision boundaries makes them very valuable in the field of rice species identification. Because of this, SVMs can recognize subtle differences across species with accuracy.
- SVMs have a strong emphasis on maximizing the margins between classes, which is a tactical method that aids in improving and classifying rice species with sharper distinctions. as well as the model's strong generalization to new data. Conversely, gradient boosting algorithms usually produce borders that lack subtlety.
- Furthermore, SVMs show a notable ability to effectively handle overfitting issues that are commonly observed in high-dimensional datasets related to tasks involving the identification of rice species, ensuring consistent and robust performance even in complex data environments.

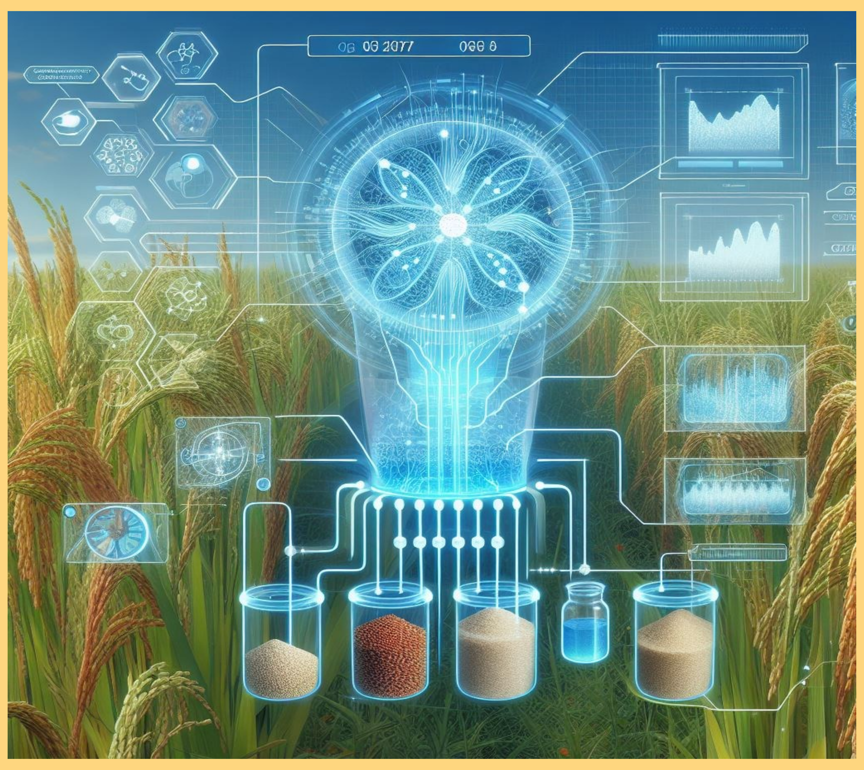

Fig 1 Rice species identification

MATERIALS AND METHODS

Data Collection
Describe the source of the dataset containing images or features of rice species.
Detail the characteristics of the dataset, including the number of samples, classes, and any preprocessing steps applied.

Support vector machine(SVM)model
- Identify relevant features to distinguish rice species.
- Optimize SVM parameters like kernel type and regularization.
- Scale features to ensure SVM's robust performance.

Model Comparison
Present the results of experiments comparing the performance of SVM and GBM models.
Include tables or graphs illustrating classification accuracy and other evaluation metrics.

Data Preprocessing
- Clean, transform, and scale data.
- Handle class imbalance.
- Split data and implement cross-validation.

Gradient boosting machine
GBM algorithms are less sensitive to feature scaling but may benefit from categorical feature encoding.
Ensure that categorical variables are appropriately encoded before training the GBM model.

Model EvaluationDefine the evaluation metrics used to assess the performance of the models, such as accuracy, precision, recall, F1-score, and area under the receiver operating characteristic curve
Justify the choice of evaluation metrics based on the problem context.

Model Integration
- Train SVM and GBM models.
- Compare performance metrics.
- Analyze feature importance.
- Address computational resources and ethical considerations.

RESULTS

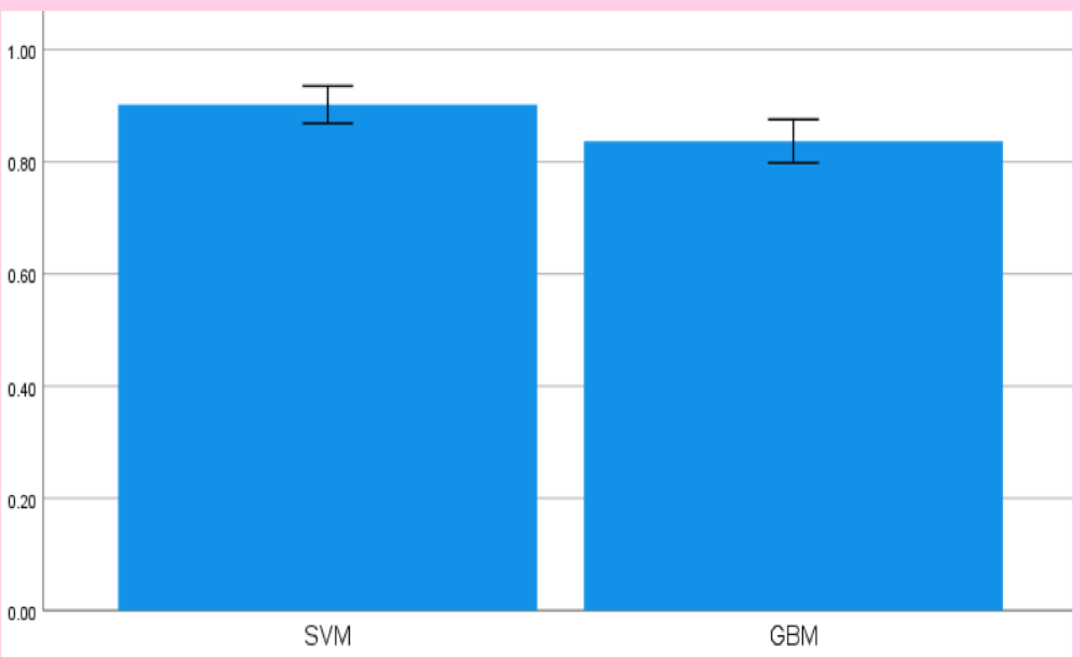

Fig 2: SVM and gradient boosting machine

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

Table 1. Comparison of the Accuracy values of SVM and GBM Algorithms with a Test size of 10 Samples

S.No	Test Size	ACCURACY RATE	
		Support vector Machine Algorithm	Gradient boosting Machine Algorithm
1	Test 1	92.55	82.36%
2	Test 2	91.23	85.19%
3	Test 3	93.78	81.58%
4	est 4	89.67	87.91%
5	Test 5	94.32	79.43%
6	Test 6	90.88	84.65%
7	Test 7	93.45	80.27%
8	Test 8	88.76	86.12%
9	Test 9	91.99	88.76%
10	Test 10	91.99	83.09%
Average Test Results		92.91	83.72%

Table 2. Mean, Standard Deviation and Standard error mean with accuracy rate comparison of Support vector machine over Gradient boosting machine algorithm

	Group	N	Mean	Std. Deviation	Std. Mean
Accuracy	SVM Algorithm	10	.9020	.04662	.01474
	GBM algorithm	10	.8370	.05417	.01713

DISCUSSION AND CONCLUSION

- Accuracy for rice species identification: 0.93, with precision and recall for "Cammeo" at 0.92 and 0.93 respectively, and for "Osmancik" at 0.94 and 0.93 respectively, supported by 762 samples.
- Noisy GBM model achieved an accuracy of 0.84, with precision-recall pairs for "Cammeo" at 0.85-0.78 and for "Osmancik" at 0.83-0.89, supported by 762 samples.
- Future scope involves exploring deep learning and integrating spectral imaging for enhanced accuracy in rice species identification, considering factors like data quality, environmental variability, and limitations in real-world adaptability and computational scalability.
- The robust classification performance, marked by high accuracy and precision metrics, highlights the model's effectiveness in rice species identification, vital for agricultural decisions and biodiversity preservation.

BIBLIOGRAPHY

- Bi, Y., Yassue, R.M., Paul, P., Dhatt, B.K., Sandhu, J., Do, P.T., Walia, H., Obata, T. and Morota, G., 2023. Evaluating metabolic and genomic data for predicting grain traits under high night temperature stress in rice. G3: Genes, Genomes, Genetics, 13(5), p.jkad052.
- Chang, T.T. and Bardenas, E.A., 1965. The morphology and varietal characteristics of the rice plant (Vol. 4). Int. Rice Res. Inst.
- Virmani, S.S., Mao, C.X. and Hardy, B. eds., 2003. Hybrid rice for food security, poverty alleviation, and environmental protection. Int. Rice Res. Inst.
- Kumar, N., Shahnaz, C., Kumar, K., Mohammed, M. and Raw, R., 2020. Advance concepts of image processing and pattern recognition. Pattern Recognition, 4(2), pp.1-10.
- Getachew, B., 2024. plant species classification using deep learning (Doctoral dissertation, St. Mary's University).