AN EFFICIENT QUALITY ANALYSIS OF RICE GRAINS USING SUPPORT VECTOR MACHINE OVER RANDOM FOREST WITH IMPROVED ACCURACY

Shaik Imam Rahaman Basha[1], N. Bharatha Devi[2*]

Article History: **Received:** 12.12.2022 **Revised:** 29.01.2023 **Accepted:** 15.03.2023

Abstract

Aim: The aim of the study is to use an eevcvgggfficient quality analysis of ricoverg gg d c c ains using a Novel Support Vector Machine over a Random Forest with improved accuracy.
Materi als and Methods: Novel Support Vector ,Machine algorithm (N=10) and Randov vf RCC z,v (N=of .41% compared to the Random Forest of 80.37%. The statistical significance ovf the analysis of rice grains difference is p=0.001 (p<0.05) and Independent sample T-test value states that the results in the study are significant.
Conclusion: The accuracy performance parameter of the Random Forest appears to be better than the Novel Support Vector Machine algorithm .

Keywords: Land, Machine Learning, Population, Random Forest, Rice grains, Novel Support Vector Machine.

[1]Research Scholar, Department of Computer Science and Engineering, Saveetha School of Engineering, CVv Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamilnadu, India, Pincode: 602105. VC x cvwb g cc. 5,, c,
[2*]Project Guide, Department of Computer Science and Engineering,d d,urx cd grrg':'"__VV.' x z Saveetha School of Engineering,c vf22vg cD c3c vr,cf Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamilnad u, India, Pincode: 602105.
Tc

1. Introduction

India is the world's second-largest rice provider, with 44 million hectares under cultivation. Rice is taken by 65 percent of the Indian population, and it accounts for 40 percent of the country's rice production (Y. Li et al. 2018). Polished white rice (ponni), brown rice, and Kavuni (black) rice are the most common rice varieties grown in south India (Singh et al. 2011). Moisture content, land, grain purity, cracks, presence of immature grains, grain dimensions, whiteness, milling degree, and chalkiness all impact rice quality. Of these parameters, grain purity, dimensions and cracks are the main features that decide the grade of the rice and cost. So, the correct assessment of the quality of rice is an important task. The analysis of grain type, grading, and evaluating quality features are all done by hand. The manual inspection method is more difficult. It is dependent on the land, operating circumstances, human variables, cleaning rate, and salvage rate. Other factors that affect rice quality include the presence of contaminants such as stones, damaged seeds, and fractured granules (Singh et al. 2011). Because manual examination by human inspectors is less accurate, there is a greater chance of contaminants being mixed together. Rice quality suffers as a result of this (de Oliveira, Pegoraro, and Viana 2020). Manual checking is also inconvenient because it requires operator concentration and is time demanding.The use of a sample testing method increases the cost of testing. To solve these challenges, automated classifier systems based on image processing are being developed to test the quality of rice grains. Rice grain quality is a complicated feature that reflects producers', processors', sellers', and consumers' perspectives on the grain's production, processing and marketing . Milling quality, cooking quality, aesthetic and sensory quality, and nutrition quality are the broad categories. The amount of recovery of grain products in order of value is determined by milling quality; appearance quality is determined by the ability to attract purchasers. The edible properties are related to sensory and nutritional quality. Despite their widespread use in a variety of applications, rice grains are still exported to other nations such as South America, Australia, and the United States, and this may be beneficial to the Indian economy(Song et al. 2022) .

Several research publications in this field can be found in IEEE and Science Direct. A total of 124 publications have been published in Science Direct, 42 papers have been published in IEEE, and 426 papers have been published in Google Scholar. Rice is consumed by more than half of the world's population. The most cited articles (de Oliveira, Pegoraro, and Viana 2020) have a total of eight citations. However, a growing global population, changing climate conditions, and decreases in arable land area and irrigation water supply are all putting pressure on the industry's long-term viability (Y. H. Li and Tuong 2001). The number of panicles per plant, the number of grains per panicle, and the grain weight all influence rice grain yield. Grain size, as the primary determinant of grain weight, is one of the most important yield-related features in cereal plants, affecting both yield potential and grain quality, as well as the grain's economic worth.Our team has extensive knowledge and research experience that has translated into high quality publications(K. Mohan et al. 2022; Vivek et al. 2022; Sathish et al. 2022; Kotteeswaran et al. 2022; Yaashikaa, Keerthana Devi, and Senthil Kumar 2022; Yaashikaa, Senthil Kumar, and Karishma 2022; Saravanan et al. 2022; Jayabal et al. 2022; Krishnan et al. 2022; Jayakodi et al. 2022; H. Mohan et al. 2022)

There is a research gap in the already existing system. Since it may be difficult to analyze the quality of rice grains. Some research has been carried out in Machine Learning algorithms for the quality of rice. However, it is important to analyze and compare the various classification algorithms that provide better accuracy. Hence, the work aims to compare the accuracy of Novel Support Vector Machines and Random Forest algorithms for quality analysis of rice grains.

2. Materials and Methods

The research study was done in an Internet Programming Lab, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences. The number of groups identified for the study are two. Group-1 is a Novel Support Vector Machine algorithm and Group-2 is the Random Forest algorithm. Sample size for each group was calculated by using previous study results in clinical.com by keeping g power as 85%, the alpha value of 0.05 and confidence interval as 95% (Edmonston 2011). According to that, the sample size of the Novel Support Vector Machine (N=10) and Random Forest algorithm (N=10) were calculated.

Support Vector Machine

Support Vector Machine is a supervised Machine Learning model that is able to solve the regression and classification problems. It is, however, mostly used to solve classification problems. Every data point placed on N-dimensional space represents space. Every point that is pointed in the dimensional space represents the feature that is the value of a specific coordinate in the Support Vector

Machine. The objective of the Support Vector Machine algorithm is to find a hyperplane in an N-dimensional space that distinctly classifies the data points. The dimension of the hyperplane depends upon the number of features. If the number of input features is three, then the hyperplane becomes a 2-D plane. It becomes difficult to imagine when the number of features exceeds three.

Pseudo Code

Input - rice grain assignment dataset
Output - Accuracy
1. Import the packages and import the dataset into programs using anaconda.
2. Load the dataset bold text and read the dataset with 205 rows and 26 columns.
3. Removing the irrelevant features and checking for null values.
4. Converting categorical values to numeric values.
5. Split the data into training and testing sets and model building and training.
6. Check the Accuracy.

Random Forest

Random Forest is a popular machine learning algorithm that belongs to supervised learning technique. It can be used for both classification and regression problems in machine learning. It is based on the concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model. Random Forest which creates decision trees on data samples and then gets the prediction from each of them and finally selects the best solution by means of voting. It is an ensemble method which is better than a single decision tree because it reduces the over-fitting by averaging the result.

Pseudo Code

Input - rice grain assignment dataset
Output - Accuracy
1. Import the packages and import the dataset into programs using anaconda.
2. Load the dataset bold text and read the dataset with 205 rows and 26 columns.
3. Split the data into sets (x_test,y_test) and (y_train,x_train).
4. Classify the fitting into two sets x,y train and x,y test.
5. Predict and Check the Accuracy.

The research work was experiment in google colab, the Hardware and Software requirements for experimenting the work includes i5 processor, 500GB SSD, 16GB RAM, Windows OS, python: colab/jupiter. The dataset was divided into two parts: Training and testing sets. Then the algorithm is experimented on and the training sets are varied 10 times based on the set size. The dataset contains the information about the rice grain and full data about the quality of rice grains. Based on the different rice image data, it is said that the quality of rice grain is used to predict the rice grain by applying the algorithms. The Dataset was collected from kaggle (https://www.kaggle.com/ntnu-testimon/paysim1).

Statistical Analysis

In this research study the quality analysis of rice grains was done using Anaconda Navigator – Spyder and IBM SPSS. It is a statistical software tool used for data analysis. For both proposed and existing algorithms 10 iterations were done with a maximum of 80-90 samples and for each iteration Dependent variables are batch size. The predicted accuracy was noted for analyzing accuracy.and for every change in the input. Independent T-Test ("T-Test II: Independent Samples T-Test" 2014) which is used to compare Novel Support Vector Machine algorithms and Random Forest algorithms to analyze the quality of rice grains.

3. Results

Table 1 shows the quality analysis of rice grains with respect to its Mean and Standard Deviation of Novel Support Vector Machine algorithm and Random Forest algorithm, which describes that Novel Support Vector Machine has an accuracy mean of 67.41%, Standard Deviation of 20 for the sample size of N=10, where the Random Forest algorithm has an accuracy Mean of 80.3%, Standard Deviation of 20 for the sample size of N=10. Based on the above results the statistical significance of the Random Forest algorithm is high. Table 2 shows the quality analysis of rice grains with respect to its Mean Difference, Standard Error Difference and Significant Difference of Novel Support Vector Machine algorithm and Random Forest algorithm, which describes that there is a significant difference between the two groups with $p=0.01$ ($p<0.05$). Table 3 shows the accuracy values of SVM algorithm and RF algorithm. The mean accuracy of SVM and RF are 67.41 and 80.3 respectively.

Figure 1 shows the bar graph comparison of the mean accuracy of Random Forest algorithm quality analysis of rice grains. The mean and standard deviation of RF is better than SVM and this shows that the performance of RF performs well in the analysis of rice grains effectively.

4. Discussion

The accuracy of Random Forest algorithm quality analysis of rice grains appears to be better than the

Novel Support Vector Machine algorithm (Parhami 2006; Y. Li et al. 2018) quality analysis of rice grains accuracy. The statistical significant difference between the two groups obtained is p=0.001 (p<0.05). Consumers of today are very conscious about the quality of food grains (Poutanen et al. 2021). In order to ensure the quality of rice grains, a quality assessment system based on RF and SVM classifiers has been addressed in this research work. Two types of rice grains viz. Ponni and Matta are taken up for study. The proposed system identified and classified the rice grains based on their morphological (Szymanek 2012) and geometrical features (Polster 2012). The experimental analysis showed that the proposed RF classifier (Sheela et al. 2016) has an overall accuracy of 80.37%, whereas SVM resulted in 67.41%. On the other hand, when the other parameters are considered, such as computational time and parallel processing (Parhami 2006)(classifying more than one sample at a time). The limitation performance of SVM is analyzed in rice quality from the dataset obtained from kaggle. The work has independent values which are present in the data set which makes the process slow and the performance of SVM is far lesser than its counterpart. The future study should analyze the quality of rice and the accuracy of the rice with more accuracy rate and with less computational time when compared to the present study and these are the reasons that may be due to an imbalanced training ratio. It is concluded that the proposed system can effectively classify the type of rice grains.

5. Conclusion

Today's consumers are increasingly concerned about the quality of their food grains. An automated rice grain quality evaluation system based on Machine Learning classifiers has been addressed in this study effort in order to ensure the quality of rice grains. The proposed system identified and classified the quality analysis of rice grains based on Novel Support Vector Machine and Random Forest algorithms. Finally, the experimental results show that the Novel Support Vector Machine accuracy of 67.41% and Random Forest accuracy of 80.37%.

Declarations

Conflicts of interest

No conflicts of interests in this manuscript.

Authors Contribution

Author SIRB was involved in data collection, data analysis, manuscript writing. Author NBD was involved in conceptualisation, data validation and critical review of manuscript.

Acknowledgements

The authors would like to express their gratitude towards Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences (Formerly known as Saveetha University) for providing the necessary infrastructure to carry out this work successfully.

Funding

We thank the following organizations for providing financial support that enabled us to complete the study.

1. BrainoVision Solutions Pvt. Ltd.
2. Saveetha University.
3. Saveetha Institute of Medical and Technical Sciences.
4. Saveetha School of Engineering.

6. References

Edmonston, Phil. 2011. Lemon-Aid Used Cars and Trucks 2011-2012. Dundurn.

Jayabal, Ravikumar, Sekar Subramani, Damodharan Dillikannan, Yuvarajan Devarajan, Lakshmanan Thangavelu, Mukilarasan Nedunchezhiyan, Gopal Kaliyaperumal, and Melvin Victor De Poures. 2022. "Multi-Objective Optimization of Performance and Emission Characteristics of a CRDI Diesel Engine Fueled with Sapota Methyl Ester/diesel Blends." Energy. https://doi.org/10.1016/j.energy.2022.123709.

Jayakodi, Santhoshkumar, Rajeshkumar Shanmugam, Bader O. Almutairi, Mikhlid H. Almutairi, Shahid Mahboob, M. R. Kavipriya, Ramesh Gandusekar, Marcello Nicoletti, and Marimuthu Govindarajan. 2022. "Azadirachta Indica-Wrapped Copper Oxide Nanoparticles as a Novel Functional Material in Cardiomyocyte Cells: An Ecotoxicity Assessment on the Embryonic Development of Danio Rerio." Environmental Research 212 (Pt A): 113153.

Kotteeswaran, C., Indrajit Patra, Regonda Nagaraju, D. Sungeetha, Bapayya Naidu Kommula, Yousef Methkal Abd Algani, S. Murugavalli, and B. Kiran Bala. 2022. "Autonomous Detection of Malevolent Nodes Using Secure Heterogeneous Cluster Protocol." Computers and Electrical Engineering. https://doi.org/10.1016/j.compeleceng.2022.107902.

Krishnan, Anbarasu, Duraisami Dhamodharan, Thanigaivel Sundaram, Vickram Sundaram, and Hun-Soo Byun. 2022. "Computational Discovery of Novel Human LMTK3 Inhibitors by High Throughput Virtual Screening Using

NCI Database." Korean Journal of Chemical Engineering. https://doi.org/10.1007/s11814-022-1120-5.

Li, Y. H., and T. P. Tuong. 2001. Water-Saving Irrigation for Rice: Proceedings of an International Workshop Held in Wuhan, China, 23-25 March 2001. IWMI.

Li, Yunfeng, Peijun Du, Zhaohui Xue, Le Gan, Xin Wang, and Hao Liang. 2018. "A Method of Rice Information Extraction Based on Particle Swarm Optimization SVM Algorithm." Proceedings of the 2018 10th International Conference on Machine Learning and Computing. https://doi.org/10.1145/3195106.3195155.

Mohan, Harshavardhan, Sethumathavan Vadivel, Se-Won Lee, Jeong-Muk Lim, Nanh Lovanh, Yool-Jin Park, Taeho Shin, Kamala-Kannan Seralathan, and Byung-Taek Oh. 2022. "Improved Visible-Light-Driven Photocatalytic Removal of Bisphenol A Using V2O5/WO3 Decorated over Zeolite: Degradation Mechanism and Toxicity." Environmental Research. https://doi.org/10.1016/j.envres.2022.113136.

Mohan, Kannan, Abirami Ramu Ganesan, P. N. Ezhilarasi, Kiran Kumar Kondamareddy, Durairaj Karthick Rajan, Palanivel Sathishkumar, Jayakumar Rajarajeswaran, and Lorenza Conterno. 2022. "Green and Eco-Friendly Approaches for the Extraction of Chitin and Chitosan: A Review." Carbohydrate Polymers 287 (July): 119349.

Oliveira, Antonio Costa de, Camila Pegoraro, and Vívian Ebeling Viana. 2020. The Future of Rice Demand: Quality Beyond Productivity. Springer Nature.

Parhami, Behrooz. 2006. Introduction to Parallel Processing: Algorithms and Architectures. Springer Science & Business Media.

Polster, Burkard. 2012. A Geometrical Picture Book. Springer Science & Business Media.

Poutanen, Kaisa S., Anna O. Kårlund, Carlos Gómez-Gallego, Daniel P. Johansson, Nathalie M. Scheers, Ingela M. Marklinder, Anne K. Eriksen, et al. 2021. "Grains - a Major Source of Sustainable Protein for Health." Nutrition Reviews, November. https://doi.org/10.1093/nutrit/nuab084.

Saravanan, A., P. Senthil Kumar, B. Ramesh, and S. Srinivasan. 2022. "Removal of Toxic Heavy Metals Using Genetically Engineered Microbes: Molecular Tools, Risk Assessment and Management Strategies." Chemosphere 298 (July): 134341.

Sathish, T., R. Saravanan, V. Vijayan, and S. Dinesh Kumar. 2022. "Investigations on Influences of MWCNT Composite Membranes in Oil Refineries Waste Water Treatment with Taguchi Route." Chemosphere 298 (July): 134265.

Sheela, J., Department of Computer Application, National Institute of Technology, Tiruchirappalli, Tamil Nadu, India-, and A. Vadivel. 2016. "Criminal Event Detection and Classification in Web Documents Using ANN Classifier." International Journal of Signal Processing Systems. https://doi.org/10.18178/ijsps.4.5.382-388.

Singh, R. K., C. V. Singh, N. P. Mandal, and M. S. Anantha. 2011. Upland Rice In India. Scientific Publishers.

Song, Shaojie, Haiyang Lin, Peter Sherman, Xi Yang, Shi Chen, Xi Lu, Tianguang Lu, Xinyu Chen, and Michael B. McElroy. 2022. "Deep Decarbonization of the Indian Economy: 2050 Prospects for Wind, Solar, and Green Hydrogen." iScience 25 (6): 104399.

Szymanek, Bogdan. 2012. Introduction to Morphological Analysis.

"T-Test II: Independent Samples T-Test." 2014. https://doi.org/10.4135/9781473997547.

Vivek, J., T. Maridurai, K. Anton Savio Lewise, R. Pandiyarajan, and K. Chandrasekaran. 2022. "Recast Layer Thickness and Residual Stress Analysis for EDD AA8011/h-BN/B4C Composites Using Cryogenically Treated SiC and CFRP Powder-Added Kerosene." Arabian Journal for Science and Engineering. https://doi.org/10.1007/s13369-022-06636-5.

Yaashikaa, P. R., M. Keerthana Devi, and P. Senthil Kumar. 2022. "Algal Biofuels: Technological Perspective on Cultivation, Fuel Extraction and Engineering Genetic Pathway for Enhancing Productivity." Fuel. https://doi.org/10.1016/j.fuel.2022.123814.

Yaashikaa, P. R., P. Senthil Kumar, and S. Karishma. 2022. "Review on Biopolymers and Composites – Evolving Material as Adsorbents in Removal of Environmental Pollutants." Environmental Research. https://doi.org/10.1016/j.envres.2022.113114.

Tables and Figures

Table 1. The above table represents or shows the group statistical which contains Mean, Std.Deviation and Std.Error Mean of SVM and RF algorithm using deep learning

	Groups	**N**	**Mean**	**Std.Deviation**	**Std.Error Mean**

Accuracy	**SVM**	10	67.41	1.73309	0.54805
	RF	10	80.37	3.29731	1.04270

Table 2. The below table shows the independent sample test which consists of Levene's test for equality of variables and t-test equality of means

	Levene's Test for Equality of Variances		**T-test for Equality of Means**					
Accuracy	**F**	**Sig.**	**t**	**df**	**Sig**	**Std. Error Difference**	**95% Conf. Interval Lower**	**95% Conf. Interval Upper**
Equal Variances assumed	5.55	0.001	-11.001	18	<.001	1.1779	-15.4338	-10.4842
Equal Variances not assumed	5.55	0.001	-11.001	183.6	<.001	1.1779	-15.4921	-10.4259

Table 3. Accuracy values of SVM algorithm and RF algorithm (Mean accuracy of SVM and RNN = 67.41&80.3) respectively.

group_id	**SVM**	**RF**
1	61.00	71.01
1	62.05	71.18
1	63.10	71.22
1	64.19	71.25
1	65.14	72.32
1	66.71	73.55
1	67.51	74.45
1	67.61	74.51
1	67.89	74.71

1	68.13	75.67
2	69.31	75.89
2	69.44	78.99
2	69.51	79.88
2	69.69	80.13
2	69.73	80.24
2	70.01	80.31
2	70.11	80.43
2	70.19	80.91
2	70.24	82.13
2	70.26	84.16

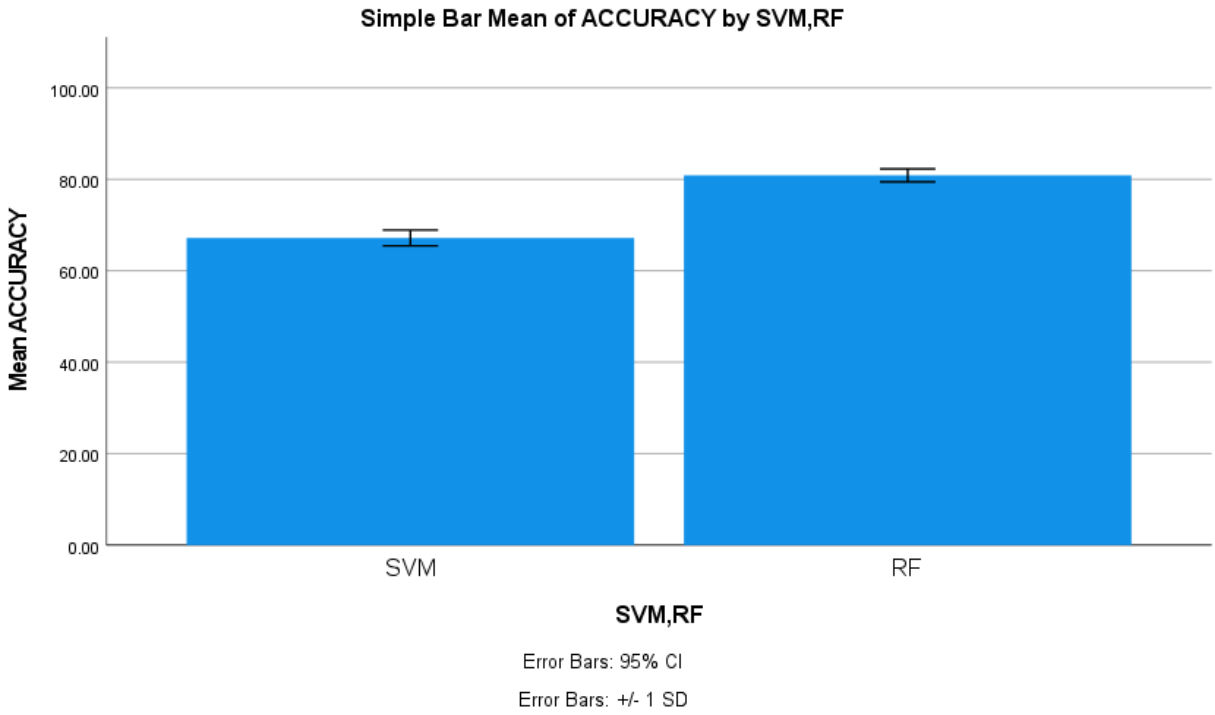

Fig. 1. Simple Bar Graph for Comparison of Accuracy. Mean accuracy of RF algorithm is better than SVM algorithm and standard deviation of RF algorithm is better than SVM algorithm. X axis: RF algorithm vs SVM algorithm, Y axis: Mean accuracy of detection.Error Bars: 95% CI and Error Bars: +/-1 SD.