# AN EFFICIENT QUALITY ANALYSIS OF RICE GRAINS USING NOVEL SUPPORT VECTOR MACHINE OVER K-NEAREST NEIGHBOR 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 efficient quality analysis of rice grains using a Novel Support Vector Machine over an K-Nearest Neighbor with improved accuracy.
**Materials and Methods:** Novel Support Vector Machine algorithm (N=10) and K-Nearest Neighbor (N=10) was iterated 20 times and analyzed rice grains. The sample size was calculated using G-power of 80% for two groups.
**Results:** Novel Support Vector Machine has significantly better accuracy of 67.41% compared to the K-Nearest Neighbor of 85.92%. The statistical significance of 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 K-Nearest Neighbor appears to be better than the Novel Support Vector Machine algorithm .

**Keywords:** K-Nearest Neighbor, Land, Machine Learning, Population, Rice grains, Novel Support Vector Machine.

[1]Research Scholar, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamilnadu, India, Pincode: 602105.
[2*]Project Guide, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamilnadu, India, Pincode: 602105.

## 1. Introduction

India is the world's second-largest rice producer, with 44 million hectares under cultivation. Rice is consumed by 65 percent of the Indian population, and it accounts for 40 percent of the country's rice production. Rice agriculture is the primary source of income and employment for around 50 million households. In India, there are about 6000 rice types being grown (Singh et al. 2011). Polished white rice (ponni), brown rice, palakkadan matta rice, aromatic biryani rice, and Kavuni (black) rice are the most common rice varieties grown in south India (Kamath 1956). Moisture content, land, grain purity, cracks, presence of immature grains, grain dimensions, whiteness, milling degree, and chalkiness all affect rice quality. Grain purity, size, and weight are among these variables.The major characteristics that determine the rice's grade and price are cracks. As a result, determining the quality of rice (Unnevehr, Duff, and Juliano 1992) is crucial. 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 (Oberoi 2005; Poutanen et al. 2021). 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. 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 multifaceted feature that reflects producers', processors', sellers', and consumers' perspectives on the grain's production, processing, marketing, and consumption. Despite their widespread use in a variety of applications, rice grains are still shipped to other countries such as South America and Australia, and this might help boost the Indian economy (Poutanen et al. 2021)

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 amount of panicles per plant determines rice grain output.grain weight and number of grains per panicle 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.(Bhavikatti et al. 2021; Karobari et al. 2021; Shanmugam et al. 2021; Sawant et al. 2021; Muthukrishnan 2021; Preethi et al. 2021; Karthigadevi et al. 2021; Bhanu Teja et al. 2021; Veerasimman et al. 2021; Baskar et al. 2021)

In the current system, there is a research deficit. Because determining the quality of rice grains can be challenging. Machine Learning techniques for rice quality have been the subject of some research. It is, therefore, critical to examine and evaluate the various categorization algorithms that provide higher accuracy. As a result, the goal of this research is to examine the accuracy of Novel Support Vector Machines and K-Nearest Neighbor algorithms for rice grain quality analysis.

## 2. Materials and Methods

The research study was done in the 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 K-Nearest Neighbor 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 K-Nearest Neighbor algorithm (N=10) were calculated.

**Support Vector Machine**

In sample group preparation 1, SVM 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 SVM. The objective of the SVM 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.

**K-Nearest Neighbor**
K-Nearest Neighbor is a simplest machine learning algorithm based on supervised learning techniques which stores all the available cases and classifies the new data or case based on a similarity measure. It is mostly used to classify a data point based on how its neighbors are classified. The K-Nearest Neighbor algorithm is a type of supervised learning technique that is used for classification and regression. It's a flexible approach that may also be used to fill in missing values and resample datasets. K-Nearest Neighbor considers K-Nearest Neighbors (Data points) to predict the class or continuous value for a new Datapoint, as the name suggests. In statistics. The K-Nearest Neighbors algorithm is a non-parametric supervised learning method first developed by Evelyn Fix and Joseph Hodges in 1951, and later expanded by Thomas Cover. It is used for classification and regression. In both cases, the input consists of the k closest training examples in a data set.

**Pseudo Code**
Input - rice grain assignment dataset 7
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. (McKinney 2017) 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 is used to compare the Novel Support Vector Machine algorithm and K-Nearest Neighbor algorithm 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 K-Nearest Neighbor 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 K-Nearest Neighbor algorithm has an accuracy Mean of 85.92%, Standard Deviation of 20 for the sample size of N=10. Based on the above results the statistical significance of the K-Nearest Neighbor 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 K-Nearest Neighbor 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 KNN algorithm. The mean accuracy of SVM and KNN are 67.41 and 85.9 respectively.

Figure 1 shows the bar graph comparison of the mean accuracy of K-Nearest Neighbor algorithm quality analysis of rice grains. The mean and standard deviation of KNN is better than SVM and this shows that the performance of KNN performs well in the analysis of rice grains effectively.

## 4. Discussion

The accuracy of K-Nearest Neighbor 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, an automated rice grain quality assessment system based on KNN 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 classified the rice grains based on their morphological (Szymanek 2012) and geometrical features (Polster 2012). The experimental analysis showed that the proposed KNN classifier (Sheela et al. 2016) has an overall accuracy of 85.92%, 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 used Novel Support Vector Machine and K-Nearest Neighbor classification algorithms to identify and classify the quality analysis of rice grains. Finally, the experimental findings demonstrate that the Novel Support Vector Machine has a 67.41 percent accuracy and the K-Nearest Neighbor has an 85.92 percent accuracy.

**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

Baskar, M., R. Renuka Devi, J. Ramkumar, P. Kalyanasundaram, M. Suchithra, and B. Amutha. 2021. “Region Centric Minutiae Propagation Measure Orient Forgery Detection with Finger Print Analysis in Health Care Systems.” Neural Processing Letters, January. https://doi.org/10.1007/s11063-020-10407-4.

Bhanu Teja, N., Yuvarajan Devarajan, Ruby Mishra, S. Sivasaravanan, and D. Thanikaivel Murugan. 2021. “Detailed Analysis on Sterculia Foetida Kernel Oil as Renewable Fuel in Compression Ignition Engine.” Biomass Conversion and Biorefinery, February. https://doi.org/10.1007/s13399-021-01328-w.

Bhavikatti, Shaeesta Khaleelahmed, Mohmed Isaqali Karobari, Siti Lailatul Akmar Zainuddin, Anand Marya, Sameer J. Nadaf, Vijay J. Sawant, Sandeep B. Patil, Adith Venugopal, Pietro Messina, and Giuseppe Alessandro Scardina. 2021. “Investigating the Antioxidant and Cytocompatibility of Mimusops Elengi Linn Extract over Human Gingival Fibroblast Cells.” International Journal of Environmental Research and Public Health 18 (13). https://doi.org/10.3390/ijerph18137162.

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

Kamath, M. G. 1956. Rice in India.

Karobari, Mohmed Isaqali, Syed Nahid Basheer, Fazlur Rahman Sayed, Sufiyan Shaikh, Muhammad Atif Saleem Agwan, Anand Marya, Pietro Messina, and Giuseppe Alessandro Scardina. 2021. “An In Vitro Stereomicroscopic Evaluation of Bioactivity between Neo MTA Plus, Pro Root MTA, BIODENTINE & Glass Ionomer Cement Using Dye Penetration Method." Materials 14 (12). https://doi.org/10.3390/ma14123159.

Karthigadevi, Guruviah, Sivasubramanian Manikandan, Natchimuthu Karmegam, Ramasamy Subbaiya, Sivasankaran Chozhavendhan, Balasubramani Ravindran, Soon Woong Chang, and Mukesh Kumar Awasthi. 2021. "Chemico-Nanotreatment Methods for the Removal of Persistent Organic Pollutants and Xenobiotics in Water - A Review." Bioresource Technology 324 (March): 124678.

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.

McKinney, Wes. 2017. Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython. "O'Reilly Media, Inc."

Muthukrishnan, Lakshmipathy. 2021. "Nanotechnology for Cleaner Leather Production: A Review." Environmental Chemistry Letters 19 (3): 2527–49.

Oberoi, Harinder Singh. 2005. Optimization of Bioethanol Production Parameters Using Damaged Rice and Cauliflower Waste.

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.

Preethi, K. Auxzilia, K. Auxzilia Preethi, Ganesh Lakshmanan, and Durairaj Sekar. 2021. "Antagomir Technology in the Treatment of Different Types of Cancer." Epigenomics. https://doi.org/10.2217/epi-2020-0439.

Sawant, Kashmira, Ajinkya M. Pawar, Kulvinder Singh Banga, Ricardo Machado, Mohmed Isaqali Karobari, Anand Marya, Pietro Messina, and Giuseppe Alessandro Scardina. 2021. "Dentinal Microcracks after Root Canal Instrumentation Using Instruments Manufactured with Different NiTi Alloys and the SAF System: A Systematic Review." NATO Advanced Science Institutes Series E: Applied Sciences 11 (11): 4984.

Shanmugam, Vigneshwaran, Rhoda Afriyie Mensah, Michael Försth, Gabriel Sas, Ágoston Restás, Cyrus Addy, Qiang Xu, et al. 2021. "Circular Economy in Biocomposite Development: State-of-the-Art, Challenges and Emerging Trends." Composites Part C: Open Access 5 (July): 100138.

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.

Szymanek, Bogdan. 2012. Introduction to Morphological Analysis.

Unnevehr, Laurian, Bart Duff, and Bienvenido O. Juliano. 1992. Consumer Demand for Rice Grain Quality: Terminal Report of IDRC Projects, National Grain Quality (Asia), and International Grain Quality Economics (Asia). Int. Rice Res. Inst.

Veerasimman, Arumugaprabu, Vigneshwaran Shanmugam, Sundarakannan Rajendran, Deepak Joel Johnson, Ajith Subbiah, John Koilpichai, and Uthayakumar Marimuthu. 2021. "Thermal Properties of Natural Fiber Sisal Based Hybrid Composites – A Brief Review." Journal of Natural Fibers, January, 1–11.

**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 KNN algorithm using deep learning

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

| | | | | | |
|---|---|---|---|---|---|
| **Accuracy** | **SVM** | 10 | 67.41 | 1.73309 | 0.54805 |
| | **KNN** | 10 | 85.92 | 5.77531 | 1.82631 |

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

| **Accuracy** | **Levene's Test for Equality of Variances** | | **T-test for Equality of Means** | | | | | |
|---|---|---|---|---|---|---|---|---|
| | **F** | **Sig.** | **t** | **df** | **Mean Difference** | **Std. Error Difference** | **95% Conf. Interval Lower** | **95% Conf. Interval Upper** |
| Equal Variances assumed | 21.76 | 0.001 | 9.70 | 18 | -18.508 | 1.90677 | -22.51398 | -14.50202 |
| Equal Variance not assumed | 21.76 | 0.001 | 9.70 | 10.6 | -18.508 | 1.90677 | -22.72378 | -14.29222 |

Table 3. Accuracy values of SVM algorithm and KNN algorithm ( Mean accuracy of SVM and KNN = 67.41 & 85.92) respectively

| **group_id** | **SVM** | **KNN** |
|---|---|---|
| 1 | 61.00 | 81.01 |
| 1 | 62.05 | 81.18 |
| 1 | 63.10 | 81.22 |
| 1 | 64.19 | 81.25 |
| 1 | 65.14 | 82.32 |
| 1 | 66.71 | 83.55 |
| 1 | 67.51 | 84.45 |
| 1 | 67.61 | 84.51 |
| 1 | 67.89 | 84.71 |

| | | |
|---|---|---|
| 1 | 68.13 | 85.67 |
| 2 | 69.31 | 85.89 |
| 2 | 69.44 | 86.99 |
| 2 | 69.51 | 88.88 |
| 2 | 69.69 | 89.13 |
| 2 | 69.73 | 89.24 |
| 2 | 70.01 | 89.31 |
| 2 | 70.11 | 89.43 |
| 2 | 70.19 | 89.91 |
| 2 | 70.24 | 90.13 |
| 2 | 70.26 | 94.16 |

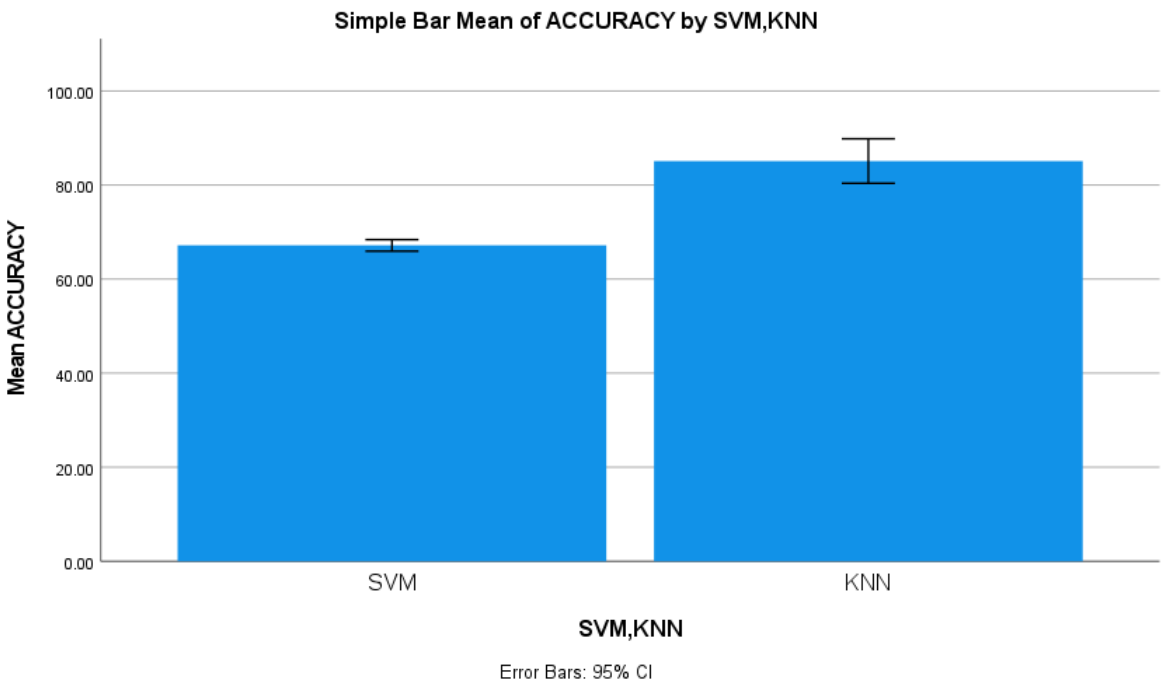


Fig. 1. Simple Bar Graph for Comparison of Accuracy. Mean accuracy of KNN algorithm is better than SVM algorithm and standard deviation of KNN algorithm is better than SVM algorithm. X axis: KNN algorithm vs SVM algorithm, Y axis: Mean accuracy of detection. Error Bars: 95% CI and Error Bars: +/-1 SD.