PREDICTION OF QUALITY OF RICE IN RICE MILL USING RANDOM FOREST COMPARED WITH DECISION TREE WITH IMPROVED ACCURACY

Chinnam Koti Pullarao[1], V.Nagaraju[2*]

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

Abstract

Aim: This research article aims to improve the accuracy rate in the Novel prediction of quality of rice in rice mills by using Random Forest (RF) in comparison with Decision Tree (D-Tree) Classifier.
Materials & Methods: The data set in this paper utilizes the publicly available Kaggle data set for Novel prediction of the quality of rice in rice mills. The sample size of Novel prediction of quality of rice in rice mill with improved accuracy rate was sample 80 (Group 1=40 and Group 2 =40), and calculation is performed utilizing G-power 0.8 with alpha and beta qualities are 0.05, 0.2 with a confidence interval at 95%. Novel Prediction of quality of rice in rice mills with improved accuracy rate is performed by Random Forest (RF) whereas several samples (N=10) and Decision Tree (D-Tree) where the number of samples (N=10).
Results: The Random Forest (RF) classifier has 94.0 higher accuracy rates when compared to the accuracy rate of Decision Tree (D-Tree) is 92.7. The study has a statistical significant value of $p<0.05$, i.e., $p=0.0313$.
Conclusion: Random Forest (RF) provides better outcomes in accuracy rate when compared to Decision Tree (D-Tree) for Novel prediction of quality of rice in rice mills.

Keywords: Quality of Rice, Random Forest (RF), Decision Tree (D-Tree), Accuracy rate, Rice Mill, Novel Prediction.

[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

Rice is the most eaten food worldwide, and the rice market is generally extreme(Parrado et al. 2006). In rice-producing businesses, the market request is generally focused on the nature of rice. In assessing rice quality, assessing essential aspects like length, width and thickness assume a significant part (Bhattacharya 2011). One of the serious issues in foreseeing the quality properties of rice is that leading examinations in the food business can be exceptionally costly. Many complex calculations are additionally acquainted with grading the rice quality by utilizing image processing-based robotized classifiers. This paper presents a random forest (RF) algorithm to predict the rice quality and compares the results with the decision tree (D-Tree) algorithm(Wang et al. 2020). The proposed method is executed using Python software and produces better accuracy results(Aquerreta et al. 2007).

Numerous researchers and scientists have endeavored to plan and foster programmed frameworks in light of PC vision and artificial reasoning for quality assessment and evaluating rice. IEEE Xplore distributed 77 examination papers, and Google Scholar tracked down 87 articles.(Aulakh and Banga 2012) have proposed image processing techniques for grading of rice samples based on their sizes.(Maheshwari, Jain, and Modi 2012) have proposed image processing procedures for reviewing rice tests because of their sizes.(Siddagangappa and Kulkarni 2014) Have proposed Classification and Quality Analysis of Food Grains. The work was expected to foster a consistent application utilized to recognize and group the food grains and grade the rice grains.(Ajay et al. 2013) have proposed a quality assessment of rice grains using great morphological techniques. The quality evaluation of rice plays an urgent part in rice quality assurance strategies in rice creation and in deciding the ensuing rice cost in the food grain market(Sharma, Sethi, and Bawa 2020). For the quality arranging, investigation and quality assessment of rice, there is an immense commitment to made PC advancements(Parmar, Jain, and Modi 2011).(Kuchekar and Yerigeri 2018) have endeavored to grade the rice grains because of image processing procedures. Nature of rice is resolved because of the physical and synthetic qualities like region, length, width, dampness content, whiteness, processing degree and so on(Avudaiappan et al. 2019)zeroed in on a review to examine visual elements, for example, shape, shading, the surface of rice seed pictures. Here image processing method was joined with grouping strategies like MLP, SVM, Decision tree and Bayesian organization to analyze and order rice seeds from arranged examples in a non-contact mode.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)

The drawback of the conventional decision tree strategy is tedious, conflicting, and unfortunate exact results. The focal point of the recommended work is to acquaint a mechanical structure with recognizing and arranging rice grains as per the market assumptions. To measure up to this assumption and accomplish a predictable standard quality, a portion of rice's physical and synthetic attributes must be broken down. Here the random forest (RF) algorithm predicts the rice grains. The proposed RF technique distinguishes and arranges the rice grains with good precision.

2. Materials and Methods

This work was carried out at Machine Learning Laboratory in Saveetha School of Engineering, SIMATS, Tamil Nadu, and India. In this study, the rice data set was collected from various rice mills. 13 different types of rice samples are taken. The Sample size was analyzed utilizing earlier works(Du et al. 2019). Group 1 was a Decision Tree (D-Tree) algorithm and Group 2 was a Random Forest (RF) algorithm. In this work two groups are taken and 10 samples for each group, total samples considered are 20. The calculation is performed utilizing G-power 0.8 with alpha and beta qualities 0.05, 0.2 with a confidence interval at 95%.

Decision tree algorithm

A decision tree classifier is a proactive model having applications in various regions. It comprises a node graph that addresses attributes and clarifies some pressing issues. Then again, edges in a chart display the solutions to the inquiries posed before. Ultimately, the leaves signify the actual outcome acquired after following away down the tree. This stream is recursive, and a similar interaction is rehashed for individual subtrees. All data input training is taken as root when using a decision tree algorithm. An in-depth analysis of the drug decision algorithm is also considered when it is used for isolation problems. If the depth of the algorithm is high, then the issue of overheating will occur. Decision Trees is a systematic machine learning method in which data is subdivided according to a specific parameter. It has two businesses like decision node and leaves. The branches and leaves are the building blocks of the tree. The size represents the maximum number of nodes present in the tree. Since each node of the

tree algorithm decision is used to do a binary split, the size should be as large as 2d + 1-1, where d indicates the depth of the tree. And this proper depth can be applied to any binary split problem. At higher depths, the issue of overdose occurs. And in the case of multiple classification problems, depth can vary.

Step 1: Separate the images based on the labels.
Step 2: For training, 75% of the data is used to build the model.
Step 3: Remaining 25% of the data is used for testing the model.
Step 4: Remove all the outliers from the dataset, So the model does not over fit.
Step 5: Deep Decision tree relates to the data given.
Step 6: Predicts the nutrition analysis and calorie count significantly.

Random forest (RF) algorithm

The sample preparation group 2 is the innovative random forest (RF) algorithm, a supervised machine learning algorithm to anticipate rice quality. The trial results show that the proposed HOG strategy has improved precision. Random forest is an ensemble classifier and works effectively on a colossal dataset. It can anticipate the missing information precisely, even when vast parts of information are missing without pre-handling. It consolidates packing and irregular element determination. Random Forest contains decision trees that are combined with individual learners. Each decision tree is built by utilizing a distinctive subset of the preparation information, determined to defeat over-fitting; for example, the calculation clarifies the info instead of observing standards that sum up the issue of individual choice trees. Random forest is one of the most widely used machine learning algorithms for classification. It can likewise be utilized for regression models, i.e., continuous values; however, it, for the most part, performs well on arrangement models of classifying values. Random forest reduces overfitting as the classification is made by voting the Novel prediction from different decision trees on sampled data. The accuracy from random forests is also good. The features are then subjected to a Random Forest classifier, which classifies the rice grain into either grade1 or grade2.

Step 1: Separate the images based on the labels.
Step 2: For training, 75% of the data is used to build the model.
Step 3: Remaining 25% of the data is used for testing the model.
Step 4: Remove all the outliers from the dataset, So the model does not overfit.
Step 5: Deep Random forest (RF) relates to the data given.
Step 6: Predicts the nutrition analysis and calorie count significantly.

Statistical Analysis

Statistical analysis is done by the software named IBM SPSS Statistics 26. Google co-lab is used for simulation & verification(J. P. Verma 2012). The independent sample t test was performed to find the mean, standard deviation and the standard error mean statistical significance between the groups, and then comparison of the two groups with the SPSS software will give the accurate values. The Dependent variable used is accuracy and the independent variable is image size.

3. Results

Figure 1 shows the simple bar graph for Decision Tree (D-Tree) Classifier accuracy rate is compared with Random Forest (RF) Classifier. The Random Forest (RF) Classifier is higher in terms of accuracy rate 94.0 when compared with Decision Tree (D-Tree) Classifier 92.7.

Table 1 shows the Evaluation Metrics of Comparison of Decision Tree (D-Tree) and Random Forest (RF) Classifier. The accuracy rate of Decision Tree (D-Tree) is 92.7 and Random Forest (RF) has 94.0. Random Forest (RF) provides better performance compared with the Decision Tree (D-Tree) of Novel prediction of quality of rice in rice mills with improved accuracy rate.

Table 2 shows the statistical calculation such as Mean, standard deviation and standard error Mean for Decision Tree (D-Tree) and Random Forest (RF). The mean accuracy rate of Decision Tree (D-Tree) is 92.7 and Random Forest (RF) is 94.0. The Standard Deviation of Decision Tree (D-Tree) is 1.92348 and Random Forest (RF) is 0.25833. The Standard Error Mean of Decision Tree (D-Tree) is 0.97453 and Random Forest (RF) is 0.19345.

Table 3 displays the statistical calculations for independent samples test between Decision Tree (D-Tree) and Random Forest (RF). The significance level for signal to noise ratio is 0.0313.

4. Discussion

The above results show the comparison of accuracy rate for Decision Tree (D-Tree) Classifier and Random Forest (RF) Classifier. The Random Forest (RF) Classifier is higher in terms of accuracy rate 94.0 when compared with Decision Tree (D-Tree) Classifier 92.7.

Many studies have been published in the literature for the past few years. (Guzman, Peralta, and Others 2008) proposed improving a structure of machine vision innovation and multi-facet fake brain frameworks for distinguishing proof of the rice sizes, rice shapes, and rice assortments of five varietal bunches consequently. Multi-facet ANNs

pick 13 morphological highlights from each rice pictures. The ANN classifiers yielded identical sizes and individually shaped 98.76% and 96.67%. Verma(B. Verma 2010) used a faster CV framework for arranging and examination of rice portions. On three Indian rice assortments, Hafed Basmati, Markfed Supreme, and Golden Smoothing, binarization techniques for image processing were applied to assess the extricated boundaries region, border, greatest width, most extreme length, depreciation, and stretching. An order precision between 90-95% was achieved in light of contrasts in pastiness, shape, broke, harmed, and broken rice portions of three assortments. Kambo and Yerpude(Kambo and Yerpude 2014b) proposed another PCA approach utilizing morphological highlights with the assistance of a KNN classifier to order three kinds of Basmati rice and accomplished generally speaking exactness of 79%. Kambo and Yerpude(Kambo and Yerpude 2014a) characterized three types of basmati rice in their proposed framework utilizing investigation with the assistance of KNN classifiers that created a general exactness of 73%. Tahir et al.(Tahir et al. 2015) examined the different strategies used for quality examination and reviewing rice with the assistance of picture processing. There is no contradicting observation connected with this work. They proposed utilizing one more technique with the exactness of simply 46.60% utilizing image processing for the quality investigation and evaluating rice in light of shading, shape, and length highlights.

The limitations of the proposed framework are not fit for getting grain synthesis, inward bug pervasion, and organoleptic properties, factors that are primary deciders of food grain quality. In the future, new frameworks with higher precision can be created with the assistance of genuinely progressing recent advancements of AI, deep learning, and image processing utilizing a blend of various consolidated highlight sets with a more extensive dataset of unmistakable rice assortments.

5. Conclusion

The proposed model exhibits the Decision Tree (D-Tree) and Random Forest (RF), in which the Random Forest (RF) has the highest values. The accuracy Rate of Random Forest (RF) is 94.0 is higher compared with Decision Tree (D-Tree) that has an accuracy rate of 92.7 in analysis of Novel prediction of quality of rice in rice mills with improved accuracy rate and with a 2 - tailed significant value of 0.001.

Declarations

Conflicts of Interest

No conflict of interest in this manuscript

Authors Contributions

Author KPR was involved in data collection, data analysis & manuscript writing. Author VN was involved in conceptualization, data validation, and critical review of manuscripts.

Acknowledgments

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 successfully carrying out this work.

Funding

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

1. VEE EEE Technologies
2. Saveetha University
3. Saveetha Institute of Medical And Technical Sciences
4. Saveetha School of Engineering

7. References

Ajay, G., M. Suneel, K. Kiran Kumar, and P. Siva Prasad. 2013. "Quality Evaluation of Rice Grains Using Morphological Methods." *International Journal of Soft Computing and Engineering* 2: 35–37.

Aquerreta, J., A. Iguaz, C. Arroqui, and P. Vírseda. 2007. "Effect of High Temperature Intermittent Drying and Tempering on Rough Rice Quality." *Journal of Food Engineering* 80 (2): 611–18.

Aulakh, Jagdeep Singh, and V. K. Banga. 2012. "Grading for Rice Grains by Image Processing." IJERT.

Avudaiappan, T., S. Sangamithra, A. S. Roselin, S. S. Farhana, and K. M. Visalakshi. 2019. "Analysing Rice Seed Quality Using Machine Learning Algorithms." *SSRG International Journal of Computer Science and Engineering (SSRG—IJCSE)—Special Issue ICRTCRET* 474.

Bhattacharya, Kshirod R. 2011. *Rice Quality: A Guide to Rice Properties and Analysis*. Elsevier.

Du, Hongying, Yinqiang Huo, Huili Liu, Ghulam Mustafa Kamal, Jiaren Yang, Yongchao Zeng, Siming Zhao, and Youming Liu. 2019. "Fast Nutritional Characterization of Different Pigmented Rice Grains Using a Combination of NMR and Decision Tree Analysis." *CyTA - Journal of Food* 17 (1): 128–36.

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.

Kambo, Rubi, and Amit Yerpude. 2014a. "Principal Component Analysis Based Classification Technique for Basmati Rice Grain Analysis." *Dimensions* 1000: 1.

———. 2014b. "Classification of Basmati Rice Grain Variety Using Image Processing and Principal Component Analysis." *arXiv [cs.CV]*. arXiv. http://arxiv.org/abs/1405.7626.

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.

Kuchekar, N. A., and V. V. Yerigeri. 2018. "Rice Grain Quality Grading Using Digital Image Processing Techniques." *IOSR J Electronics Communication Eng* 13 (3): 84–88.

Maheshwari, Chetna Vasudevbhai, Kavindra R. Jain, and Chintan K. Modi. 2012. "Non-Destructive Quality Analysis of Indian Basmati Oryza Sativa SSP Indica (Rice) Using Image Processing." *2012 International Conference on Communication Systems and Network Technologies*. https://doi.org/10.1109/csnt.2012.47.

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.

Parmar, Rohit R., Kavindra R. Jain, and Chintan K. Modi. 2011. "Unified Approach in Food Quality Evaluation Using Machine Vision." In *Advances in Computing and Communications*, 239–48. Communications in Computer and Information Science. Springer, Berlin, Heidelberg.

Parrado, Juan, Esther Miramontes, Maria Jover, Juan Fco Gutierrez, Laura Collantes de Terán, and Juan Bautista. 2006. "Preparation of a Rice Bran Enzymatic Extract with Potential Use as Functional Food." *Food Chemistry* 98 (4): 742–48.

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.

Sharma, Komal, Ganesh Sethi, and Rajesh Bawa. 2020. "State-of-the-Art in Automatic Rice Quality Grading System." https://doi.org/10.2139/ssrn.3564372.

Siddagangappa, Megha R., and A. H. Kulkarni. 2014. "Classification and Quality Analysis of Food Grains." *IOSR Journal of Computer Engineering (IOSR-JCE)* 16 (4): 01–10.

Tahir, Wan Putri Nwm, Norhaida Hussin, Zaw Zaw Htike, and Wai Yan Nyein Naing. 2015. "Rice Grading Using Image Processing." *ARPN Journal of Engineering and Applied Sciences* 10 (21): 10131–37.

Verma, Bhupinder. 2010. "Image Processing Techniques for Grading Amp; Classification of Rice." In *2010 International Conference on Computer and Communication Technology (ICCCT)*, 220–23.

Verma, J. P. 2012. *Data Analysis in Management*

with SPSS Software. Springer Science & Business Media.

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.

Wang, Jinning, Kun Li, Yun Shao, Fengli Zhang, Zhiyong Wang, Xianyu Guo, Yi Qin, and Xiangchen Liu. 2020. "Analysis of Combining SAR and Optical Optimal Parameters to Classify Typhoon-Invasion Lodged Rice: A Case Study Using the Random Forest Method." *Sensors* 20 (24). https://doi.org/10.3390/s20247346.

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. Comparison of Decision Tree (D-Tree) and Random Forest (RF) Classifier for predicting the quality of rice in rice mills with improved accuracy rate. The accuracy rate of Decision Tree (D-Tree) is 92.7 and Random Forest (RF) has 94.0.

S.No.	**Test Size**	**ACCURACY RATE**	
		Decision Tree (D-Tree) Classifier	Random Forest (RF) Classifier
1	Test1	92.004	94.004
2	Test2	92.031	94.031
3	Test3	92.675	94.675
4	Test4	92.192	94.192
5	Test5	91.484	94.484
6	Test6	91.872	92.872
7	Test7	91.483	91.483
8	Test8	92.394	93.394
9	Test9	92.493	93.493
10	Test10	92.393	93.393
Average Test Results		92.7	94.0

Table. 2. The statistical calculation such as Mean, standard deviation and standard error Mean for Decision Tree (D-Tree) and Random Forest (RF). The accuracy rate parameter used in the t-test. The mean accuracy rate of Decision Tree (D-Tree) is 92.7 and Random Forest (RF) is 94.0. The Standard Deviation of Decision Tree (D-Tree) is 1.92348 and Random Forest (RF) is 0.25833. The Standard Error Mean of Decision Tree (D-Tree) is 0.97453 and (RF) is 0.19345.

Group		**N**	**Mean**	**Standard Deviation**	**Standard Error Mean**
ACCURACY	RANDOM FOREST (RF)	10	94.0	0.25833	0.19345
	DECISION TREE (D-TREE)	10	92.7	1.92348	0.97453

Table 3. The statistical calculations for independent samples test between Decision Tree (D-Tree) and Random Forest (RF). The significance level for signal to noise ratio is 0.0313. Independent samples T-test is applied for comparison of Decision Tree (D-Tree) and Random Forest (RF) with the confidence interval as 95% and level of significance as 0.2323.

Group		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval (Lower)	95% Confidence Interval (Upper)
Accuracy	Equal variances assumed	9.293	0.031	18.45	18	.001	12.3455	0.83849	12.7285	15.2943
	Equal variances not assumed			12.34	14.675	.001	12.0456	0.14537	10.2839	13.0239

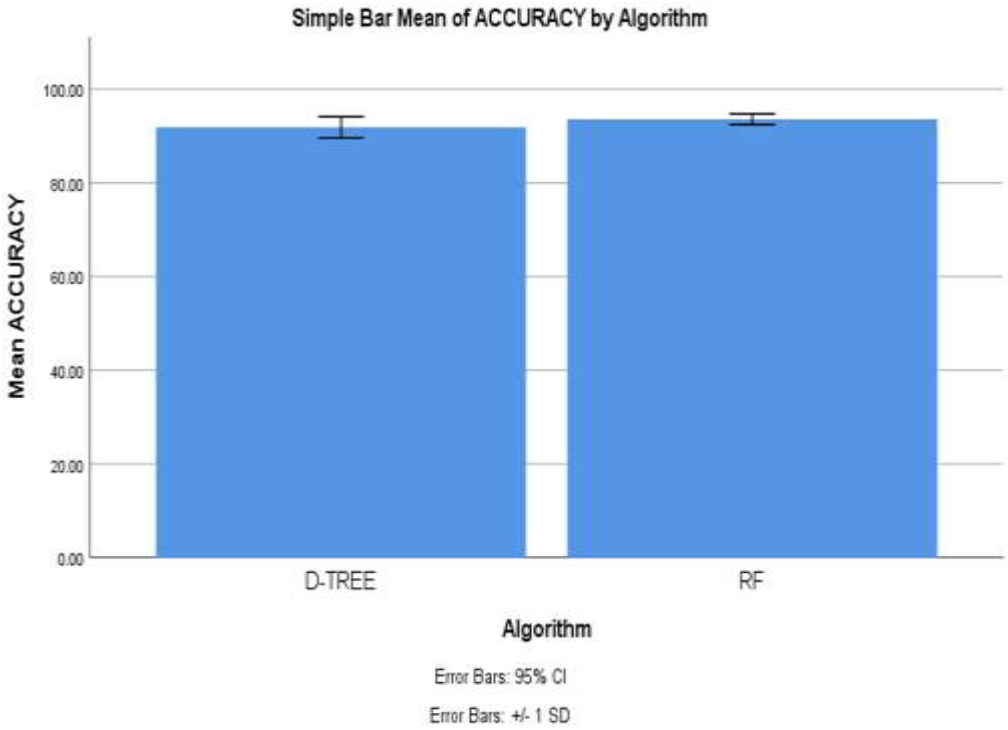

Fig. 1. Simple Bar graph for Decision Tree (D-Tree) Classifier accuracy rate is compared with Random Forest (RF) Classifier. The Decision Tree (D-Tree) Classifier is higher in terms of accuracy rate 92.7 when compared with Random Forest (RF) Classifier 94.0. Variable results with its standard deviation ranging from 80 lower to 90 higher Decision Tree (D-Tree) Classifier where Random Forest (RF) Classifier standard deviation ranging from 90 lower to 100 higher. There is a significant difference between Decision Tree (D-Tree) Classifier and Random Forest (RF) Classifier ($p<0.05$ Independent sample test). X-axis: Random Forest (RF) Classifier accuracy rate vs Decision Tree (D-Tree) Classifier Y-axis: Mean of accuracy rate, for identification of keywords ± 1 SD with 95 % CI.