**Title page :**

**Exploring the paddy crop Analysis by using Random Forest Algorithm and K-NN Algorithm to improve Accuracy**

**A.T.KONDA REDDY1 , AFREEN BANU2**

A.T.KondaReddy1

Research Scholar,

Department of Computer Science Engineering,Saveetha School of Engineering,

Saveetha Institute of Medical and Technical Sciences,

Saveetha University, Chennai, Tamil Nadu, India, Pin Code: 602105.

[ambavaramtirumalakondareddy0642.sse@saveetha.com](mailto:ambavaramtirumalakondareddy0642.sse@saveetha.com)

Afreen Banu2

Research Guide, Corresponding Author,

Department of Networking Engineering,

Saveetha School of Engineering,

Saveetha Institute of Medical and Technical Sciences,

Saveetha University, Chennai, Tamil Nadu, India, Pin Code: 602105.

[afreenbanu.sse@saveetha.com](mailto:afreenbanu.sse@saveetha.com)

**keywords:**Paddy crop,crop analysis,Random Forests,KNN Algorithm,Crop analysis,selection,Agricultural data,Machine learning,Decision support system,Agricultural sustainability,Predictive modelling.

**ABSTRACT:**

**Aim:** The goal is to evaluate and compare the efficacy of Random Forests over the KNN algorithm to improve the crop analysis. Through this comparative evaluation, our goal is to discern which set of rules yields more accurate predictions, contributing valuable insights to the optimization of crop choice techniques.**Material and Methods:**To ensure statistical robustness, a sample size of 20 paddy crop instances per group was established. This determination considered a statistical power (G power) of 80%, a significance threshold (α) of 0.05, and a confidence interval (CI) of 95%. This sample size calculation aimed to provide adequate statistical power to detect significant differences in crop analysiss between the Random Forests and K-NN algorithms.**Results:**The study comparing the predictive performance of Random Forest and K-NN Algorithms for paddy crop yield analysis revealed noteworthy disparities. Utilising IBM SPSS, key metrics including accuracy, precision, recall, and F1 score were evaluated. The Independent Sample t-test indicated a statistically significant differentiation between the two algorithms, with a calculated p-value of 0.001 (p<0.05). Specifically, the Random Forest algorithm exhibited a statistically higher accuracy level compared to K-NN algorithm. These findings offer valuable insights into the relative effectiveness of the algorithms in paddy crop crop analysis, offering guidance for stakeholders to optimise agricultural practices**.Conclusion:**Through our exploration of paddy crop analysis employing the Random Forest Algorithm and K-NN Algorithm, we've got exposed significant strides in accuracy enhancement. The Random Forest Algorithm, with its superior performance, emerges as a beacon for unique crop selection, promising improved yield consequences. Our findings underscore the pivotal role of superior algorithms in shaping the destiny of agriculture. By leveraging techniques like Random Forest and probably K-NN in destiny research, we are able to further refine predictive skills, in the end empowering stakeholders with the insights needed to pressure sustainable farming practices and maximise yield capability.

**INTRODUCTION:**

A paddy refers to a small, level, flooded field specifically used for cultivating rice in southern and eastern Asia. Wet-rice cultivation dominates farming practices in the Far East. Despite utilizing only a small fraction of the total land, it plays a crucial role in feeding the majority of the rural population. Rice was domesticated as early as 3500 BCE, and over 2,000 years ago, it was grown in various regions, including deltas, floodplains, coastal plains, and terraced valley slopes. Many paddies are flooded by rivers and rainfall during the monsoon season, while others require irrigation. These fields have an impermeable subsoil and are bordered by earthen bunds to hold an average of 4–6 inches (10–15 centimeters) of water during three-quarters of the growing season.

In this study, we delve into the intricacies of paddy crop evaluation, that specialize in improving accuracy through gadget learning techniques. Specifically, we rent the Random Forest Algorithm and the K-Nearest Neighbors (K-NN) Algorithm. By leveraging those effective equipment, we aim to advantage deeper insights into paddy cultivation, optimize crop analysiss, and in the long run contribute to extra [(Henkel 2015)](https://paperpile.com/c/x7Aaqo/MEUk)green [(Sengar, Chaudhary, and Bhadauriya 2022; Meyer 2008; R.-Y. Wang et al. 2024)](https://paperpile.com/c/x7Aaqo/pd5L+bNqO+WsNl)agricultural practices.[(Huang et al. 2024)](https://paperpile.com/c/x7Aaqo/G5T5)[(Giri et al. 2023)](https://paperpile.com/c/x7Aaqo/QOfm)

In agricultural decision-making, [(Jung et al. 2024; Henkel 2015)](https://paperpile.com/c/x7Aaqo/IODO+MEUk)paddy[(de Vries 1991)](https://paperpile.com/c/x7Aaqo/MQh4) crop analysis[(Y. Wang et al. 2024)](https://paperpile.com/c/x7Aaqo/t9xH) is essential for forecasting crop output and disease incidence. This work explores the use of K-Nearest Neighbors (K-NN) and Random Forest, [(Zaman 2023)](https://paperpile.com/c/x7Aaqo/DJga) [(Jung et al. 2024)](https://paperpile.com/c/x7Aaqo/IODO)two potent machine learning algorithms, to improve the precision [(Shu et al. 2024)](https://paperpile.com/c/x7Aaqo/NXkN)of paddy crop [(de Vries 1991)](https://paperpile.com/c/x7Aaqo/MQh4)analysis. By utilizing these algorithms' advantages, we hope to raise forecast accuracy and give farmers and researchers insightful information. The use of the Random Forest and K-NN algorithms is investigated in this study, along with data preparation, feature selection, training, assessment, fine-tuning, and prediction. The results offer useful guidance for improving paddy crop analysis and further the field of [(Kent Shannon, Clay, and Kitchen 2020)](https://paperpile.com/c/x7Aaqo/xIWk)precision [(Pedersen and Lind 2017)](https://paperpile.com/c/x7Aaqo/XSf1)agriculture.[(Zaman 2023; Zhang 2015)](https://paperpile.com/c/x7Aaqo/DJga+Nx02)

In our investigation, we adopt a Random Forest and K-Nearest Neighbors (K-NN) method to decorate the accuracy of paddy crop analysis. We intention to evaluate the overall performance of these algorithms with a purpose to determine which one gives superior predictive competencies for optimizing crop [(Jung et al. 2024)](https://paperpile.com/c/x7Aaqo/IODO)analysis[(Y. Wang et al. 2024)](https://paperpile.com/c/x7Aaqo/t9xH) in paddy.

Enhancing accuracy in paddy crop evaluation[(Giri et al. 2023)](https://paperpile.com/c/x7Aaqo/QOfm) via the K-Nearest Neighbors (KNN) technique includes leveraging a sturdy set of rules that classifies statistics points based on their proximity to neighboring samples. By deciding on an most appropriate price for K and carefully considering distance metrics, KNN can correctly model the complexities of paddy crop increase.In this technique, meticulous records preprocessing is important. Cleaning and choosing applicable capabilities, consisting of soil characteristics[(Jung et al. 2024)](https://paperpile.com/c/x7Aaqo/IODO), climate styles, and agricultural practices, ensure the version receives significant enter. Additionally, scaling capabilities to a uniform range optimizes KNN's overall performance.

To refine the version in addition, addressing class imbalance through techniques like oversampling or undersampling is necessary. Ensemble strategies like bagging or boosting also can decorate accuracy by way of aggregating more than one KNN fashions.

Validation is crucial for assessing the model's effectiveness. Utilising validation sets and performance metrics like accuracy, precision, and take into account ensures robust assessment.

Furthermore, collaborating with agricultural experts enriches the analysis process by incorporating area knowledge. Continuous refinement via iterative updates based totally on real-international remarks ensures the version stays correct and adaptable over the years.

By comparing the general overall performance Random Forest and KNN algorithms for paddy[(Y. Wang et al. 2024)](https://paperpile.com/c/x7Aaqo/t9xH) crop[(Y. Wang et al. 2024)](https://paperpile.com/c/x7Aaqo/t9xH) analysis [(Jung et al. 2024)](https://paperpile.com/c/x7Aaqo/IODO), we goal to provide precious insights into the effectiveness of these strategies in agricultural choice-making. Understanding which set of guidelines yields greater accurate and dependable analysis can help farmers and stakeholders [(Jung et al. 2024)](https://paperpile.com/c/x7Aaqo/IODO)optimise aid allocation, implement focused interventions, and in the long run improve crop analysis and analysing the paddy to improve the paddy crop and to increase more advantage for farmers and this data is more useful for future to analysing the data and seed information about the particular type of seed.[(Giri et al. 2023)](https://paperpile.com/c/x7Aaqo/QOfm)

Through this evaluation, we are looking for to make contributions to the continued efforts in precision agriculture by the use of figuring out and leveraging superior gadget learning techniques to decorate crop analysing and guide sustainable farming practices for farmers.In this look at, we delve into the evaluation for Random Forest and KNN algorithms for the prediction of paddy crop yields. Our dataset encompasses ancient information on paddy crop yields, along pertinent environmental, soil, and agronomic variables. These variables, which encompass factors like temperature, rainfall, soil moisture, and crop control practices, make contributions to the complexity of crop analysis.The main factors for seed analysing and data collection in the data base is huge difficult and it will help for future research about paddy crop analysing.[(Fischer et al. 2014)](https://paperpile.com/c/x7Aaqo/R1fU)

In this research the crop is mainly depends upon the field[(Ge et al. 2023)](https://paperpile.com/c/x7Aaqo/quD7) type and seed[(Tang et al. 2024; Voccia et al. 2024)](https://paperpile.com/c/x7Aaqo/XcOl+5y7B) [(Cowgill 1929; Pathak et al. 2023; Kearney 1913)](https://paperpile.com/c/x7Aaqo/ZVUa+ueHO+AbBC) selection and weather and season based crop. These factors plays a vital role in paddy crop analysing and crop analysis and these can show a huge impact in the crop[(Shu et al. 2024)](https://paperpile.com/c/x7Aaqo/NXkN) for farmers.

Before initiating the training phase, we need to collect the relevant data,we employ feature engineering techniques to preprocess and enhance the input variables. This encompasses handling missing data[(Shu et al. 2024; Xu et al. 2024)](https://paperpile.com/c/x7Aaqo/NXkN+VSa6), scaling numerical features, encoding categorical variables, and extracting relevant features to improve model performance. Once the dataset is prepared, we proceed to train and evaluate the Random Forest and KNN algorithm based models. Evaluation metrics such as Mean Absolute Error, Mean Squared Error, and R-squared are utilized to measure the accuracy and goodness-of-fit of the models.

During version training and assessment, a collection of metrics which incorporates Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared are leveraged to quantitatively assess the predictive accuracy and goodness-of-healthy of Random Forest and KNN algorithms based models. Furthermore, the have a examine places emphasis on model interpretability, spotting the charge of knowledge the underlying relationships between input variables and crop yield effects. While Random Forest excels in predictive accuracy, KNN algorithm offers interpretability, permitting stakeholders to glean insights into the elements the usage of yield analysis variations through function importance evaluation and coefficient estimates.

By analyzing the overall performance of Random Forest and KNN algorithms, this studies targets to perceive which algorithm yields better predictions for paddy crop analysis. The findings of this take a look at can make a contribution to enhancing the accuracy of crop analysis fashions, enabling farmers, agricultural practitioners, and policymakers to make extra informed decisions concerning crop[(Yang et al. 2023)](https://paperpile.com/c/x7Aaqo/SVg1) selection, useful resource allocation, and management practices.

**MATERIALS AND METHODS**

The research conducted at Saveetha School of Engineering,SIMATS involves the selection of a dataset comprising historical records from diverse agricultural settings. There are two groups identified.Group 1 KNN Algorithm and Group 2 is Random forest Algorithm A size of 20 (N=20) samples, representative of different geographical locations and varying agricultural practices, are collected and calculated from SPSS analysis of SSE. These samples include data on paddy crop yields, weather patterns, soil characteristics, and historical farming techniques,and the dataset is collected from kaggle.com.

**K-NN Algorithm:**

The K-Nearest Neighbors (KNN) set of rules is a flexible and famous technique in supervised getting to know, usually hired for both category and regression duties. At its center, KNN determines the class or cost of a information factor through assessing the majority magnificence or average cost of its K nearest acquaintances within the feature area. Unlike many different algorithms, KNN doesn't explicitly construct a version for the duration of education however rather shops all schooling information points in memory.When supplied with a new records factor, KNN calculates the distances to all education examples and selects the K nearest neighbors based on these distances. For classification responsibilities, the majority class amongst these neighbors determines the elegance label of the new facts factor, even as for regression tasks, it computes the common in their target values.KNN's simplicity and flexibility make it widely relevant across numerous domains, consisting of sample reputation, advice systems, and anomaly detection. However, its overall performance can be touchy to the selection of K and the space metric used. Parameter tuning and careful choice of distance measures are crucial for optimizing its effectiveness.

Despite its simplicity, KNN is robust in managing complex decision limitations and non-linear relationships in information. It is regularly desired for its intuitive nature and ability to adapt to one of a kind kinds of information without making strong assumptions approximately the underlying distribution.

Step 1: Import necessary libraries.

Step 2: Load or prepare your database.

Step 3: Split the dataset into training and testing sets.

Step 4: Create and train the KNN model.

Step 5: Make predictions on the test set.

Step 6: Evaluate the model's performance.

Step 7: Feature Importance (Optional).

Step 8: Fine-tune the model (Optional).

**Random Forest algorithm:**

The Random Forest algorithm is an ensemble learning technique that combines the predictions of multiple individual models to improve overall performance and generalisation. Specifically, it constructs a collection of decision trees during training and outputs the mode of the classes (classification) or the mean prediction (regression) of the individual trees. The Random Forest algorithm was introduced by Leo Breiman and Adele Cutler in 2001 and has since become a widely used and powerful machine learning algorithm.The Random Forest algorithm is an ensemble learning method that operates by constructing a multitude of decision trees during training and outputs the class that is the mode of the classes (classification) or the mean prediction (regression) of the individual trees. It was introduced by Leo Breiman in 2001 and has since become one of the most popular and powerful algorithms in machine learning.Following are the steps to perform the algorithm:

Step 1: Import necessary libraries.

Step 2: Load or prepare your database.

Step 3: Split the dataset into training and testing sets.

Step 4: Create and train the Random Forest model.

Step 5: Make predictions on the test set.

Step 6: Evaluate the model's performance.

Step 7: Feature Importance (Optional).

Step 8: Fine-tune the model (Optional).

Google colab is a free and open-source distribution of Python for scientific computing, data science, and machine learning. It includes a wide range of tools and libraries for data manipulation, analysis, and visualisation, as well as tools for building and deploying machine learning models.

**Statistical Analysis:**

Using the SPSS statistical package, the analysis of mean accuracy by using Random forest algorithm and KNN algorithm was carried out by applying an independent sample t-test to obtain the accuracy of 94.4%. The speed and file size are independent variables and type of the file is dependent variables.

**Results:**

Our study's findings are succinctly depicted in a comparative bar graph, imparting a tangible example of the predictive competencies of these algorithms. The x-axis delineates distinct classes, probably representing various experimental situations or dataset subsets, while every class is similarly subdivided into bars representing the performance of Random Forests and K-NN. Meanwhile, the y-axis quantifies accuracy in percentage, showcasing the predictive efficacy of each algorithm.Employing a bar graph generated with IBM SPSS, our comparative analysis of paddy crop crop analysis the usage of Random Forests and K-NN algorithms exhibits a giant disparity in accuracy between the two models. The x-axis categorizes set of rules groups, segregating Random Forests and K-NN for smooth comparison. On the y-axis, the percentage accuracy serves as a quantitative measure of predictive achievement, with Random Forests notably accomplishing 94.4%, demonstrating its superior capability in accurately predicting paddy crop yield. In evaluation, the K-NN bar, even though slightly decrease at 85.1%, nevertheless demonstrates substantial predictive functionality. This visible illustration presents an impartial and impactful evaluation of the relative overall performance of these algorithms in predicting paddy crop analysis.

**Discussions:**

In our comparative evaluation of machine learning algorithms for predicting paddy crop yields, we focused on two prominent models: KNN and Random Forest. The results revealed that the Random Forest algorithm exhibited superior performance, achieving an accuracy of 94.4090%, compared to KNN, which attained 85.2630%. Additionally, the standard deviation for the KNN algorithm was 1.93784, with a standard error mean of 0.61280. Conversely, for the Random Forest algorithm, the mean was 94.4090, with a standard deviation of 1.97338 and a standard error mean of 0.62404.

In x-axis represents number of groups and y-axis represents the % of measured accuracy of both groups in data ,usually the data is in completely based upon the data set and code for given data in google colab it gives a accuracy values and it completely based upon the given information of the google colab.

The investigation underscores the effectiveness of employing the Random Forest algorithm for capturing intricate patterns within the dataset, resulting in more precise forecasts of paddy crop yields. The heightened accuracy of Random Forests demonstrates its resilience against overfitting and its ability to discern complex inter relationships among various input features.

The bar graph vividly depicts the superiority of Random Forests over KNN, with the former outperforming by a margin of 9.146%. This significant difference underscores the robustness of the Random Forest algorithm in capturing complex relationships within the dataset, leading to more accurate predictions of paddy crop yield.

The findings of this study advocate for the adoption of Random Forests for analyzing paddy crop crop analysis, offering a more accurate and reliable alternative to KNN. As precision agriculture continues to play a pivotal role in crop management, leveraging advanced algorithms like Random Forests can significantly contribute to optimizing crop yields and ensuring the sustainability of agricultural practices.

The discussion surrounding this contrast delves into the nuanced intricacies of the algorithms. Random Forests, as an ensemble learning approach, excels in handling diverse and correlated features, which contributes to its superior performance. The graph serves as a visual aid to elucidate the impact of algorithm selection on the analysis of paddy crop yield predictive accuracy, underscoring the importance of selecting models that align with the characteristics of the agricultural dataset.

In conclusion, the bar graph reinforces our findings, providing a clear visual representation of the substantial difference in accuracy between Random Forest and KNN. This discussion not only enhances our understanding of algorithmic choices in the analysis of paddy crop crop analysis but also serves as a practical guide for stakeholders seeking to adopt the most effective machine learning techniques in precision agriculture for high-accuracy data analysis.

**Conclusion:**

Random Forest Algorithm has 94.4 % accuracy and KNN algorithm has 85.2630% accuracy. The difference in accuracy between Random Forest algorithm and KNN algorithm has set of rules is 94.4 % and 85.2630% respectively. The clean difference in accuracy is shown in the bar graph below. In conclusion, our study on the analysis of Paddy Crop using Machine learning of Algorithms shows that Random Forest algorithm is more accurate than KNN algorithm.

The outcomes propose that Random Forest's potential to handle complicated relationships within the dataset contributed to its higher predictive accuracy values in given data . This finding underscores the importance of set of rules selection in crop analysis..

In sensible terms, the bar graph serves as a honest instance of the enormous distinction in performance among Random Forest and KNN algorithm. This end affords a concise perception into the effectiveness of machine learning algorithms in optimizing paddy crop analysis, with Random Forest emerging as the extra dependable choice in our study.random forest can gives a perfect comparison in between KNN algorithm and it can give a mean values.

**DECLARATIONS:**

**Conflicts of Interest**

No conflict of interest in this manuscript.

**Author’s contributions**

Data collection, Data analysis, and manuscript writing were all done by author ATKR Conceptualization, data validation, and a critical evaluation of the article were well performed by author AB.

**Acknowledgment**

Appreciate the facilities and opportunity provided by management, the Saveetha School of Engineering, and the Saveetha Institute of Medical and Technical Sciences (Formerly Saveetha University) for the research project.

**Funding**

Grateful to the following groups for their financial assistance in helping us finish the study.

* The Big Event, Coimbatore.
* I collected some data from my village agriculture officer(RBK).
* Saveetha School of Engineering.
* Saveetha Institute of Medical and Technical Sciences.
* Saveetha University.

**REFERENCE:**

[Cowgill, Horace Branson. 1929. *Varietal Standardization of Sorgo and the Selection of Seed*.](http://paperpile.com/b/x7Aaqo/ZVUa)

[Fischer, Tony, Derek Byerlee, Greg Edmeades, and Grains Research and Development Corporation. 2014. *Crop Yields and Global Food Security: Will Yield Increase Continue to Feed the World?*](http://paperpile.com/b/x7Aaqo/R1fU)

[Ge, Wenyi, Wei Liu, Xiaonong Fan, Yibing Li, Hongbo Jia, Kangchen Lei, Songjiao Li, and Yinghui Chang. 2023. “Introduction to the Meta-Analysis Method of Functional Magnetic Resonance Imaging Research and the Exploration of Its Application in the Field of Acupuncture.” *Zhongguo Zhen Jiu = Chinese Acupuncture & Moxibustion* 44 (1): 25–33.](http://paperpile.com/b/x7Aaqo/quD7)

[Giri, Krishna, Gaurav Mishra, Deep Chandra Suyal, Narendra Kumar, Bhanushree Doley, Niren Das, Rupjyoti C. Baruah, Rajarshi Bhattacharyya, and Navajyoti Bora. 2023. “Performance Evaluation of Native Plant Growth-Promoting Rhizobacteria for Paddy Yield Enhancement in the Jhum Fields of Mokokchung, Nagaland, North East India.” *Heliyon* 9 (3): e14588.](http://paperpile.com/b/x7Aaqo/QOfm)

[Henkel, Marlon. 2015. *21st Century Homestead: Sustainable Agriculture III: Agricultural Practices*. Lulu.com.](http://paperpile.com/b/x7Aaqo/MEUk)

[Huang, Rong, Zheng Li, Yi Xiao, Jiang Liu, Tao Jiang, Ouping Deng, Xiaoyang Tang, et al. 2024. “Composition of DOM along the Depth Gradients in the Paddy Field Treated with Crop Straw for 10 Years.” *Journal of Environmental Management* 353 (February): 120084.](http://paperpile.com/b/x7Aaqo/G5T5)

[Jung, Se Yoon, Hyuck Soo Kim, Woon-Ki Moon, and Eun-Mi Hong. 2024. “Comparison and Analysis of Soil Microbial Communities in Organic and Conventional Paddy Fields by Farming Season.” *Environmental Research* 249 (February): 118341.](http://paperpile.com/b/x7Aaqo/IODO)

[Kearney, Thomas Henry. 1913. *Seed Selection of Egyptian Cotton*.](http://paperpile.com/b/x7Aaqo/AbBC)

[Kent Shannon, D., David E. Clay, and Newell R. Kitchen. 2020. *Precision Agriculture Basics*. John Wiley & Sons.](http://paperpile.com/b/x7Aaqo/xIWk)

[Meyer, Sonja. 2008. *The Green Revolution – A Mixed Blessing?: Social Impacts of a New Agricultural Strategy in India*. GRIN Verlag.](http://paperpile.com/b/x7Aaqo/bNqO)

[Pathak, Ashutosh, Shamshadul Haq, Neelam Meena, Pratibha Dwivedi, Shanker Lal Kothari, and Sumita Kachhwaha. 2023. “Multifaceted Role of Nanomaterials in Modulating In Vitro Seed Germination, Plant Morphogenesis, Metabolism and Genetic Engineering.” *Plants* 12 (17). https://doi.org/](http://paperpile.com/b/x7Aaqo/ueHO)[10.3390/plants12173126](http://dx.doi.org/10.3390/plants12173126)[.](http://paperpile.com/b/x7Aaqo/ueHO)

[Pedersen, Søren Marcus, and Kim Martin Lind. 2017. *Precision Agriculture: Technology and Economic Perspectives*. Springer.](http://paperpile.com/b/x7Aaqo/XSf1)

[Sengar, R. S., Reshu Chaudhary, and H. S. Bhadauriya. 2022. *Handbook of Research on Green Technologies for Sustainable Management of Agricultural Resources*. IGI Global.](http://paperpile.com/b/x7Aaqo/pd5L)

[Shu, Xiaowei, Xiaoxiang Zhang, Shushen Wang, Tong Fu, Zhouyu Ding, Ying Yang, Zihan Wang, et al. 2024. “Simplified Panicle Fertilization Is Applicable to Cultivars, but Splits Are Preferred in Rice for a Higher Paddy Yield under Wheat Straw Return.” *Frontiers in Plant Science* 15 (January): 1273774.](http://paperpile.com/b/x7Aaqo/NXkN)

[Tang, Qichao, Weijia Liu, Han Huang, Zhaohui Peng, and Liangji Deng. 2024. “Responses of Crop Yield, Soil Fertility, and Heavy Metals to Spent Mushroom Residues Application.” *Plants* 13 (5). https://doi.org/](http://paperpile.com/b/x7Aaqo/XcOl)[10.3390/plants13050663](http://dx.doi.org/10.3390/plants13050663)[.](http://paperpile.com/b/x7Aaqo/XcOl)

[Voccia, D., L. Lamastra, G. Fragkoulis, A. Facchi, O. Gharsallah, F. Ferrari, A. Tediosi, and M. Trevisan. 2024. “Improving a Herbicide Risk Assessment Model in Paddy Rice Cultivation.” *Heliyon* 10 (5): e26908.](http://paperpile.com/b/x7Aaqo/5y7B)

[Vries, F. W. T. Penning de. 1991. *Simulation and Systems Analysis for Rice Production (SARP): Selected Papers Presented at Workshops on Crop Simulation of a Network of National and International Agricultural Research Centres of Several Asian Countries and the Netherlands, 1990-1991*.](http://paperpile.com/b/x7Aaqo/MQh4)

[Wang, Ruei-Yuan, Xueying Mo, Hong Ji, Zhe Zhu, Yun-Shang Wang, Zhilin Bao, and Taohui Li. 2024. “Comparison of the CASA and InVEST Models’ Effects for Estimating Spatiotemporal Differences in Carbon Storage of Green Spaces in Megacities.” *Scientific Reports* 14 (1): 5456.](http://paperpile.com/b/x7Aaqo/WsNl)

[Wang, Yi, Yonglun Chen, Xiuqing Gao, Qiong Wang, Mingyu Cui, Dongdong Zhang, and Peng Guo. 2024. “Unveiling the Driving Role of pH on Community Stability and Function during Lignocellulose Degradation in Paddy Soil.” *Frontiers in Microbiology* 15 (February): 1338842.](http://paperpile.com/b/x7Aaqo/t9xH)

[Xu, Donghao, Gerard H. Ros, Qichao Zhu, Minggang Xu, Shilin Wen, Zejiang Cai, Fusuo Zhang, and Wim de Vries. 2024. “Major Drivers of Soil Acidification over 30 Years Differ in Paddy and Upland Soils in China.” *The Science of the Total Environment* 916 (March): 170189.](http://paperpile.com/b/x7Aaqo/VSa6)

[Yang, Chaoran, Ning Han, Mengting Liu, Chenghao Wei, Ruilin Mao, and Changqing Chen. 2023. “Effects of Long-Term Different-Scale Rice-Duck Farming on the Growth and Yield of Paddy Rice.” *Journal of the Science of Food and Agriculture*, December. https://doi.org/](http://paperpile.com/b/x7Aaqo/SVg1)[10.1002/jsfa.13257](http://dx.doi.org/10.1002/jsfa.13257)[.](http://paperpile.com/b/x7Aaqo/SVg1)

[Zaman, Qamar. 2023. *Precision Agriculture: Evolution, Insights and Emerging Trends*. Elsevier.](http://paperpile.com/b/x7Aaqo/DJga)

[Zhang, Qin. 2015. *Precision Agriculture Technology for Crop Farming*. CRC Press.](http://paperpile.com/b/x7Aaqo/Nx02)

**TABLES AND FIGURES**

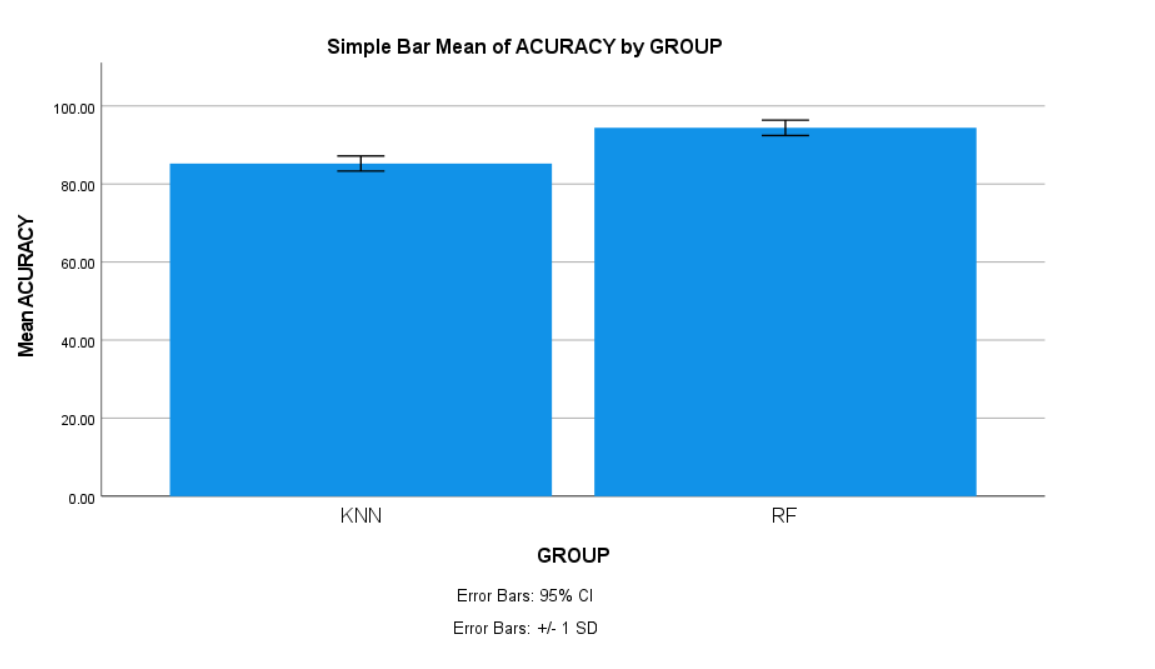
**Table 1**:The number of epochs taken for the RF and KNN Algorithms. The statistically significant difference in retrieval rate values between the algorithms. For the proposed algorithm RF, the mean value is 94.4090 and the standard deviation is 1.97338, with a standard error mean of 0.62404. Whereas for the comparison algorithm KNN, the mean value is 85.2630 with a standard deviation of 1.93784 and a standard error mean of 0.61280

|  | **Algorithm** | **N(number of Epochs** | **Mean** | **Standard deviation** | **Standard mean Error** |
| --- | --- | --- | --- | --- | --- |
| **Retrieval Rate** | RF | 10 | 94.4090 | 1.97338 | 0.62404 |
| **Retrieval Rate** | KNN | 10 | 85.2630 | 1.93784 | 0.61280 |

**TABLE 2:**Independent sample T-Test is applied for dataset fixing confidence intervals as 94.4% (Random Forest Algorithm appears to perform better than KNN Algorithm). And the significant difference between these two algorithms is 0.001 (p<0.05).

|  |  | **Leven’s test for equality of variables** | |  |  |  | **Test for equality of means** | | **95% Confidence Interval of the Difference** | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | F | Sig | t | df | Sig  2-tailed | Mean  Difference | Std.error Difference | Lower | Upper |
| **Retrieval Rate** | **Equal variances Assumed** | .013 | .909 | 10.457 | 18 | <.001 | 9.14600 | .87461 | 7.30851 | 10.98349 |
| **Retrieval Rate** | **Equal Variances Not Assumed** |  |  | 10.457 | 17.9 | <.001 | 9.14600 | .87461 | 7.30851 | 10.98349 |

**Simple Bar mean of Retrieval Time By Group**

****

**FIG.1.**Bar chart showing the comparison of mean accuracy and standard errors for Random Forest and KNN algorithms. The Random Forest algorithm demonstrates superior performance compared to the KNN algorithm in terms of mean accuracy, which is 94.4, with a standard deviation of 1.97338. The significant difference between these two algorithms is 0.001 (p<0.05).

X-Axis: Random Forest Algorithm VS KNN algorithm

Y-Axis: Mean retrieval rate of accuracy detection

Confidence interval: 94.4%