



## AN INITIATIVE FORECASTING APPROACH FOR CROP-BASED CULTIVATION USING K-NEAREST NEIGHBORS (KNN) ALGORITHM COMPARED USING RANDOM FOREST

Yaragadla Sivaram<sup>1</sup>, Aishwarya B<sup>2\*</sup>

---

**Article History:** Received: 12.12.2022

Revised: 29.01.2023

Accepted: 15.03.2023

---

### Abstract

**Aim:** In this study, K-Nearest Neighbors algorithm (KNN) is used to predict climatic parameters of the crops and its performance is tested by comparing it with the Random Forest algorithm for the crop-based cultivation for better accuracy.

**Materials and Methods:** The Crop-based on Climatic Parameters consists of 3101 of different crops and climatic parameters used for training 3000 (80%) and testing 101 (20%) the predictive model in python and the statistical analysis is done using SPSS software. The sample size is estimated using G power to be 3101 records in each group with 80% of power and a 0.05 Error rate. K-Nearest Neighbors (KNN) algorithm is used and compared with Random Forest algorithm.

**Results:** The K-Nearest Neighbors algorithm's predictive model shows a higher accuracy of 86.61% than the Random Forest based model with an accuracy of 93.54% and with a significance value 0.010 ( $p < 0.05$ ).

**Conclusion:** Within the limits of study confirms that the Random Forest based model provides more promising results in the Crop-based cultivation on Climatic parameters than the K-Nearest Neighbors (KNN) based model.

**Keywords:** Novel Crop-based Cultivation, Forecasting, Initiative Approach, Cultivation, Machine learning, Random Forest, K-Nearest Neighbors algorithm.

---

<sup>1</sup>Research Scholar, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamil Nadu, India, Pincode: 602105.

<sup>2\*</sup>Project Guide, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamil Nadu, India, Pincode: 602105.

## 1. Introduction

A forecast yield is vital to department of agriculture stakeholders, and it can be made possible by machine learning modeling and information from multiple sources (Hodges 1990). Systems based on NDVI (Normalized Difference Vegetation Index) information are typically used for yield forecasting (Mesghinna 1979). The Novel Crop-based K-Nearest Neighbors algorithm Cultivation system provides significant value by excluding the need for high-resolution remote control-perception data and by enabling the farmer to take account of adverse climate influences on Forecasting harvest oscillation (Hodges 1990; Singh et al. 2021). Initiative Approach In our studies, we forecast the soybean and maize takings for various time periods. The applications involved in the forecasting methodologies of the climatic parameters developed and evaluated by the researchers all over India and the world who will help in the field of cultivation and agriculture. Researchers found the error metric unit for soybean and maize yield forecasts to be comparable to schemes that application provide yield forecast information during the first weeks to calendar months of the crop cycle (Pradeep et al. 2019).

Over the past 7 years, nearly 1708 articles published in both google scholar and 520 articles in IEEE Xplore and 360 articles in the ScienceDirect related to crop-based cultivation on climatic parameters using machine learning (Ahuja and Ma 2020). And the Forecasting of the climatic parameters will affect the growth of the crop or the plant in the field (Tomar et al. 2021) Agro-climatic input parameters influence crop production. Agriculture input parameters vary from field to field and farmer to farmer. It is difficult to collect such information on a larger area (Teufelberger 2019). Such data sets can be used for predicting climate change (Wilson et al. 2021; Schrauf, de Los Campos, and Munilla 2021). There are different forecasting methodologies developed and evaluated by researchers all over the world in the agriculture and related sciences fields (Wilson et al. 2021).

The aim of the paper showed that data mining integrated agricultural data including pest scouting, pesticide usage, and meteorological data can help optimize pesticide usage (Doorenbos, Pruitt, and Food and Agriculture Organization of the United Nations 1975). Thematic information that pertains to agriculture and has spatial attributes is presented in one of the studies (Wilson et al. 2021; Schrauf, de Los Campos, and Munilla 2021; Montesinos-López et al. 2021). The gap of this paper is that vector machine algorithms are long training time for large datasets difficult to understand and

interpret the final model, variable weights and individual impact. The application of data mining techniques to the development of association rules for Initiative Approach droughts and floods in India was based on climate inputs. To help the researchers the Indian Meteorological Department tabulates the information collected at every 1 square meter in cities across India (J. Li et al. 2021). Providing Novel Crop-based Cultivation timely advice to predict crop productivity and analyzing crop production is crucial to help farmers to maximize crop production (Stanley et al. 2021). Yield expectation is a K-Nearest Neighbors algorithm significant farming issue. In the past ranchers used to anticipate their yield from earlier year yield encounters (Wilson et al. 2021; Schrauf, de Los Campos, and Munilla 2021; Montesinos-López et al. 2021; Liang et al. 2021).

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) (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) Research gap is to utilize this multitude of calculations and with the help of connections between them, there is a developing reach of uses and the job of Big information investigation procedures in agribusiness (Dhamu et al. 2021). Since the making of new imaginative innovations and procedures the farming field is gradually corrupting (Hodges 1990). Because of Forecasting, bountiful creation individuals are focused on developing fake items that are mixture items where there prompts an undesirable life (IEEE Staff 2020). These days, K-Nearest Neighbors algorithm current individuals don't have any awareness about the development of the harvests at the ideal opportunity and at the right place (Baseggio et al. 2021). Research experience of Novel Crop-based Cultivation developing procedures the occasional climatic conditions are additionally being changed against the crucial resources like soil, water and air which lead to instability of food (Mohanty et al. 2020).

## 2. Materials and Methods

This research study was carried out in the Department of Artificial Intelligence Laboratory

belonging to Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences (SIMATS), Chennai. This study consists of two sample groups i.e Group 1 random forest and group 2 is k-nearest neighbor. The required 3101 samples for this analysis are done using G-power calculation out of the 3101 samples. Minimum Forecasting K-Nearest Neighbors algorithm power of the analysis is fixed as 0.8 and the maximum accepted Initiative Approach error is fixed as 0.5 and the significant value is  $>0.05$ . The dataset is collected from customer based products through Kaggle, an open source data repository.

This dataset was collected from cluster, machine learning and different sample values. This dataset is about classification of work with 780 records and comments from different study sites. This helps in future decision making purposes. Finally, import the dataset into the colab and spyder to identify the crop growth and apply the coding to perform operations. The Climatic conditions dataset is chosen from the particular districts in a state. Dependent values for k-nearest neighbor standard error mean and random forest independent sample t-test is applied for the data set fixing confidence interval. This data set consists of 3101 samples of the different DT algorithm crop samples from different places on the particular districts, temperature, pH, Humidity, including Rainfall fallen on the particular districts and types of crops etc, This dataset was collected from cluster, machine learning and different sample values. This dataset is about classification of work with 3101 records and comments from different study sites. This helps in future decision making purposes. Finally, import the dataset into the colab and spyder to identify the crop growth and yield based on the climatic parameters.

### Random Forest

In statistics, Random forest is the one of machine learning algorithms. Which is used for the best solution for the prediction of the problem. It is a supervised learning method which gives a decision tree for every data sample which is used in the datasets and takes the best decision tree out of the data samples in the dataset using performance analysis.

### Pseudocode for Random Forest

1. Let rf be the new variable
2. Allocate some estimators to it
3. rf.fit be another variable
4. Connect it with X\_train set
5. Initialize the variables
6. Calculate rf\_prediction
7. Using rf.predict(x\_test)
8. Make use of the metrics.classification

9. Give formula to result\_3
10. Print result\_3

### K-Nearest Neighbor

The k-Nearest neighbors (KNN) algorithm is a simple and interesting machine learning prediction algorithm. Which is used for the prediction of complex problems. And it stores all available data and adds the new data to show similarity using the data. It can't handle the large dataset. It also has Initiative approach problems.

### Pseudocode for K-Nearest Neighbor

1. knn.fit(X\_train,y\_train)
2. knn.fit=knn
3. knn\_prediction=knn.predict
4. accuracy = accuracy\_score(y\_test, knn\_prediction)
5. print("Accuracy : ", accuracy)

### Statistical Analysis

The testing set up for the proposed system to implement with the following system is configuration of hardware Desktop with 64-bit, OS, RAM and Software Window 10, colab SPSS tool used for statistical analysis. In SPSS the dataset is prepared using a sample size of 10 data is analyzed with random forest and k-nearest neighbor, Machine learning the Initiative Approach statistical analysis is done on the two groups using the train set and test set. A Comparison Novel Crop-based Cultivation means table for random forest and knn is shown below. The system architecture of the experimental setup First we import and load data. The classify the data into test and train after the classification applies the algorithm and generates the model. After that apply the generated train model to the test. Then generate the accuracy and find the highest accuracy. The dependent variables in the given dataset are names of the crops, location of the crops, temperature, rainfall, pH values and humidity. The dependent variables in the given dataset are temperature, humidity, rainfall and pH and the independent variables are type of crop, pesticides and insecticides for the crop.

### 3. Results

Table 1, The group statistical analysis on the two groups shows that random forest has more accuracy than the other and its standard error mean is slightly different than the k-nearest algorithm. In the independent sample test, the significance of both algorithms when the equal variance is 0.115. In Figure 1, the bar chart of accuracies with standard deviation error is plotted for both the algorithms. The random forest algorithm produces an accuracy of 93.03% and the k-nearest neighbor algorithm has scored 85.82%. Table 2, Novel Crop-

based Cultivation represents the comparison of k-nearest neighbor Initiative Approach algorithm and random forest algorithm in terms of mean accuracy and the Accuracy, Precision and Recall are the methods used for measuring the overall performance of data mining. Table 3, other processes constructed on positive and negative reviews calculating accuracy, precision and recall. The Independent samples test categorizes test for equality of variance and T-test for equality of means for mean difference, standard error difference. The confidence interval and level of significance is set to  $p=0.01$ . The bar chart of accuracies with standard deviation error DT algorithm is plotted for both the algorithms. The support vector machine algorithm produces an accuracy of 83.70% and the decision tree algorithm has scored 80.21%. The variables constructed on positive and negative reviews calculating accuracy, precision and recall. Group statistics for Mean, Std. Deviation, Std. Error Mean with sample size of 10.

#### 4. Discussion

In this research work the random forest algorithm gets higher accuracy though the comparisons with plotting and stabilizing the data in Table 3 and it represents the statistical values obtained by the proposed system random forest algorithm have high accuracy value with 93% and significance value less with 0.02, when compared k-nearest neighbor with accuracy value 85%. The algorithm used in this paper provides the consolidated results. In The Novel predictions of crop growth are done based on the temperature and other elements. The random forest algorithm is showing high accuracy 93.0% positive and negative values compared to k-nearest neighbor (85%). This study works on k-nearest neighbor classifiers for data mining proved with an accuracy of 65% and proposed that decision table is a minimal and most effective algorithm. (D. Li et al. 2021) Although it is not possible to predict the Novel Crop-based Cultivation of different types of the crops on the different locations, with the help of a random forest algorithm applied to a clean dataset we can find the optimization. The future Scope, there is an idea to increase training and testing dataset and to find a variety of accuracy and can deploy as web content for the frameworks. (Nazeri, Crawford, and Tuinstra 2021). Although the results of the study showed a better Predictive model performance can be calculated by the algorithms, there are certain limitations. The limitations of k-nearest neighbor algorithms are long training time for large datasets difficult to understand and interpret and justification of the final model, variable weights and individual impact (Gu et al. 2021). By increasing the number of attributes in the input

dataset the performance will be better and accurate. With advanced preprocessing techniques, overfitting can be avoided in the future.

#### 5. Conclusion

Within the limits of this study the proposed random forest algorithm shows a significant accuracy than k-nearest neighbor algorithm. Random forest algorithms will primarily reduce the effort of physically gathering ready data for arrangement. The accuracy has increased by about 7%. The outcome demonstrates that the characterization precision of the k-nearest neighbor algorithm was moderately low in this examination and random forest algorithms have shown a better significant accuracy. The main aim is to increase the accuracy using natural parameters.

#### Declarations

#### Conflict of Interest

No conflict of interest in this manuscript.

#### Authors Contributions

Author YSR was involved in data collection, data analysis, manuscript writing. Author AB was involved in conceptualization, 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.

#### Fundings

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

1. Ed-Trix Edu Pvt. Ltd.
2. Saveetha University,
3. Saveetha Institute of Medical and Technical Sciences,
4. Saveetha School of Engineering.

#### 6. References

- Ahuja, Lajpat R., and Liwang Ma. 2020. *Methods of Introducing System Models into Agricultural Research*. John Wiley & Sons.
- Baseggio, Matheus, Matthew Murray, Di Wu, Gregory Ziegler, Nicholas Kaczmar, James Chamness, John P. Hamilton, et al. 2021. "Genome-Wide Association Study Suggests an Independent Genetic Basis of Zinc and Cadmium Concentrations in Fresh Sweet Corn Kernels." *G3* 11 (8).



- <https://doi.org/10.1093/g3journal/jkab186>.
- Dhamu, Vikram Narayanan, Suhashine Sukumar, Crisvin Sajee Kadambathil, Sriram Muthukumar, and Shalini Prasad. 2021. "Targeted On-Demand Screening of Pesticide Panel in Soil Runoff." *Frontiers in Chemistry* 9 (November): 782252.
- Doorenbos, J., W. O. Pruitt, and Food and Agriculture Organization of the United Nations. 1975. *Guidelines for Predicting Crop Water Requirements*.
- Gu, Xin, Shanshan Huang, Zhiguo Zhu, Yansong Ma, Xiaohe Yang, Liangliang Yao, Xuedong Gao, et al. 2021. "Genome-Wide Association of Single Nucleotide Polymorphism Loci and Candidate Genes for Frogeye Leaf Spot (*Cercospora Sojina*) Resistance in Soybean." *BMC Plant Biology* 21 (1): 588.
- Hodges, Tom. 1990. *Predicting Crop Phenology*. CRC Press.
- IEEE Staff. 2020. 2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT).
- Jayabal, Ravikumar, Sekar Subramani, Damodharan Dillikannan, Yuvarajan Devarajan, Lakshmanan Thangavelu, Mukilarasan Nedunchezhiyan, Gopal Kaliyaperumal, and Melvin Victor De Pours. 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, Mikhliid 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.
- Kotteswaran, 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>.
- Liang, Zhanmin, Lei Zhang, Haoting Chen, Daiyun Huang, and Bowen Song. 2021. "m6A-Maize: Weakly Supervised Prediction of mA-Carrying Transcripts and mA-Affecting Mutations in Maize (*Zea Mays*)." *Methods*, November. <https://doi.org/10.1016/j.ymeth.2021.11.010>.
- Li, Dongdong, Zhiqiang Zhou, Xiaohuan Lu, Yong Jiang, Guoliang Li, Junhui Li, Haoying Wang, et al. 2021. "Genetic Dissection of Hybrid Performance and Heterosis for Yield-Related Traits in Maize." *Frontiers in Plant Science* 12 (November): 774478.
- Li, Jing, Urminder Singh, Priyanka Bhandary, Jacqueline Campbell, Zebulun Arendsee, Arun S. Seetharam, and Eve Syrkin Wurtele. 2021. "Foster Thy Young: Enhanced Prediction of Orphan Genes in Assembled Genomes." *Nucleic Acids Research*, December. <https://doi.org/10.1093/nar/gkab1238>.
- Mesghinna, Woldezion. 1979. *Crop Yield Prediction Under Conditions of Limited Data*.
- 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.
- Mohanty, Sachi Nandan, Jyotir Moy Chatterjee, Sarika Jain, Ahmed A. Elngar, and Priya Gupta. 2020. *Recommender System with Machine Learning and Artificial Intelligence: Practical Tools and Applications in Medical, Agricultural and Other Industries*. John Wiley & Sons.
- Montesinos-López, Osvál Antonio, José Cricelio Montesinos-López, Abelardo Montesinos-Lopez, Juan Manuel Ramírez-Alcaraz, Jesse Poland, Ravi Singh, Susanne Dreisigacker, et al. 2021. "Bayesian Multi-Trait Kernel Methods Improve Multi-Environment Genome Based Prediction." *G3*, November. <https://doi.org/10.1093/g3journal/jkab406>.
- Nazeri, Behrokh, Melba M. Crawford, and Mitchell R. Tuinstra. 2021. "Estimating Leaf Area

- Index in Row Crops Using Wheel-Based and Airborne Discrete Return Light Detection and Ranging Data.” *Frontiers in Plant Science* 12 (November): 740322.
- Pradeep, N., Sandeep Kautish, C. R. Nirmala, Vishal Goyal, and Sonia Abdellatif. 2019. *Modern Techniques for Agricultural Disease Management and Crop Yield Prediction*. IGI Global.
- 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.
- Schrauf, Matías F., Gustavo de Los Campos, and Sebastián Munilla. 2021. “Comparing Genomic Prediction Models by Means of Cross Validation.” *Frontiers in Plant Science* 12 (November): 734512.
- Singh, Pawan Kumar, Alok Kumar Pandey, Sahil Ahuja, and Ravi Kiran. 2021. “Multiple Forecasting Approach: A Prediction of CO<sub>2</sub> Emission from the Paddy Crop in India.” *Environmental Science and Pollution Research International*, November. <https://doi.org/10.1007/s11356-021-17487-2>.
- Stanley, A. E., A. Menkir, B. Ifie, A. A. Paterne, N. N. Unachukwu, S. Meseke, W. A. Mengesha, et al. 2021. “Association Analysis for Resistance to Striga Hermonthica in Diverse Tropical Maize Inbred Lines.” *Scientific Reports* 11 (1): 24193.
- Teufelberger, Peter. 2019. *Crop Yield Prediction in Agriculture Based on Long Short-Term Memory*.
- Tomar, Vipin, Daljit Singh, Guriqbal Singh Dhillon, Yong Suk Chung, Jesse Poland, Ravi Prakash Singh, Arun Kumar Joshi, Yogesh Gautam, Budhi Sagar Tiwari, and Uttam Kumar. 2021. “Increased Predictive Accuracy of Multi-Environment Genomic Prediction Model for Yield and Related Traits in Spring Wheat (L.)” *Frontiers in Plant Science* 12 (October): 720123.
- Vivek, J., T. Maridurai, K. Anton Savio Lewis, 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>.
- Wilson, Stefan, Marcos Malosetti, Chris Maliepaard, Han A. Mulder, Richard G. F. Visser, and Fred van Eeuwijk. 2021. “Training Set Construction for Genomic Prediction in Auto-Tetraploids: An Example in Potato.” *Frontiers in Plant Science* 12 (November): 771075.
- 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. Descriptive statistics show the output of the descriptive statistics of the dataset. It consists of accuracies of both random forest 92.50% and k-nearest neighbor 84.03%

No of Experiments	Number of reviews in the training dataset	Random Forest			K-Nearest Neighbor		
		Accuracy	Precision	Recall	Accuracy	Precision	Recall
1	100	89.16	1.006	1.008	79.91	1.007	1.002
2	200	89.95	1.004	0.964	80.92	1.009	1.004
3	300	90.93	1.002	1.003	81.94	1.005	0.940

4	400	91.87	1.008	1.009	82.93	1.003	1.006
5	500	92.89	0.901	0.907	83.96	0.842	0.762
6	600	93.95	0.926	0.923	86.46	0.861	0.920
7	700	94.97	0.917	0.919	89.39	0.856	1.007
8	800	95.93	1.009	0.940	90.97	1.001	1.009
9	900	97.92	1.005	1.006	95.99	0.956	1.008
10	1000	93.06	0.896	1.000	85.80	0.750	0.945

Table 2. T-Test comparison novel Random Forest has higher accuracy than support vector machines. Descriptive Statistics minimum, maximum, mean and standard deviation of two groups: novel random forest and k-nearest neighbor 10 sample size is taken for both proposed and existing.

	Group	N	Mean	Std.Deviation	Std.Error Means
Accuracy	Random Forest	10	93.0630	2.73172	0.86385
	K-Nearest Neighbor	10	85.8270	5.04326	1.59482

Table 3. Group statistics T-Test has the dependent values for k-nearest neighbor standard error mean and random forest independent sample t-test is applied for the data set fixing confidence interval as 95% and level of significance as 0.02. There is a significant difference in accuracy(p=0.01).

		Levene's test for equality of variances		T-test for equality of means						
				t	df	sig(2 tailed)	Mean difference	Std error difference	95% confidence interval of the difference	
		f	sig						Lower	Upper
Accuracy	Equal variances assumed	0.115	0.010	3.990	18	0.01	7.23600	1.81375	3.42546	11.04654

	Equal variances not assumed	0.115	0.010	3.990	13.863	0.01	7.23600	1.81375	3.34228	11.12972
--	--------------------------------------	-------	-------	-------	--------	------	---------	---------	---------	----------

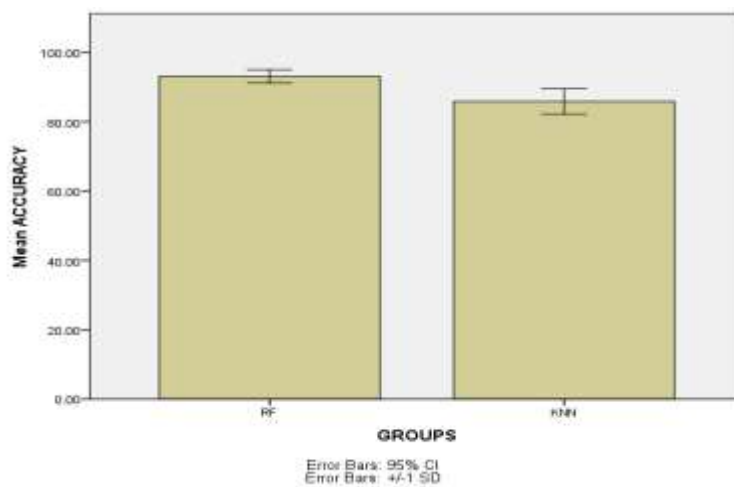


Fig. 1. Bar chart representing the comparison mean accuracy of random forest algorithm and k-nearest neighbor (KNN). The mean accuracy of novel random forest is better than k-nearest neighbor and standard deviation of novel k-nearest neighbor (KNN) is slightly better than the random forest. X-axis: random forest algorithm vs knn algorithm. Y-Axis: Mean accuracy of detection +/- 1SD.