**TITLE**

**Role of Artificial Intelligence in Solving Missing Persons Using K-Nearest Neighbor Algorithm and Comparing with Random forest Algorithm**

R.Rupasridevi1, C. Clement Raj2

R.Rupasridevi¹

Research Scholar,

Department of Computer Science and Engineering,

Saveetha School of Engineering,

Saveetha Institute of Medical And Technical Sciences,

SaveethaUniversity,Chennai,TamilNadu,India,Pincode:602105. [ravurisridevi0092.sse@saveetha.com](mailto:ravurisridevi0092.sse@saveetha.com)

Clement Raj C

Project Guide, Corresponding Author,

Department of Computer Science and Engineering,

Saveetha School of Engineering,

Saveetha Institute of Medical And Technical Sciences,

SaveethaUniversity,Chennai,TamilNadu,India,Pincode:602105.

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

**Keywords:** Supervised Learning, Feature Extraction, Location-Based Services, Emergency Response

**ABSTRACT**

**Aim:** The important goal of this research experiment is to focus on the importance of role of artificial intelligence in solving missing persons using the K-Nearest Neighbor (K-NN) algorithm and compare it with the Random forest Materials **and Methods:** The K-Nearest Neighbor is used with 20 sets as sample size, and Random forest has been used with a sample mean size of 20 sets with a total of 40 sets being compared to improve the accuracy of the present research. The mean accuracy of the present research has been calculated using the ClinCalc software appliance under supervised learning with 0.8 as the alpha value, a G-Power value of 0.8, and CI of 95%**.Results:** After performing this research, theK-Nearest Neighbor has attained an accuracy of 87.20% and the Random forest has achieved an accuracy of 74.92%. An Independent samples T-Test analysis has been executed, and its significance value is found to be p=0.000 (p<0.05), suggesting statistical significance**. Conclusion:** In this present research, K-Nearest Neighbor the algorithm is collated with the Random forest After performing the current research experiment, the K-Nearest Neighbor has been found to have more perfection than the Random forest

**Keywords:** Supervised Learning, Feature Extraction, Location-Based Services, Emergency Response

**INTRODUCTION:**

The efficacy of the KNN algorithm with the Random Forest algorithm in the context of solving missing person’s cases. Random Forest, an ensemble learning technique, combines multiple decision trees to make more accurate and robust predictions. By contrasting the two algorithms, this research seeks to identify which approach offers superior performance in terms of accuracy, efficiency, and adaptability to different types of missing person’s cases. The K-Nearest Neighbor algorithm, a supervised learning method, holds particular relevance in missing person’s cases. This algorithm works by classifying data points based on the majority class of their nearest neighbors. In the context of locating missing persons, KNN can be applied to analyze patterns in geographical data, behavioral patterns, and other relevant features that contribute to predicting the likely location of an individual.

The research papers are collected from the last 5 years i.e.; 2018-2022 and almost 300 research articles have been reported in “IEEE Xplore” and all over 390 research papers are published in the “Science Direct” on Finding Missing Persons Using AI. The IEEE Explore and Science Direct are considered as the main databases in collecting the research papers for this research experiment. Ultimately, this research seeks to offer insights into the applicability and effectiveness of AI-driven approaches in aiding law enforcement agencies, search and rescue teams, and relevant authorities in their efforts to locate missing persons. The findings aim to contribute to the discourse on leveraging technology for societal welfare, potentially facilitating the reunion of missing individuals with their families and promoting community safety and well-being-NN, a foundational machine learning algorithm, operates based on the principle of similarity assessment. In the context of missing persons, K-NN analyzes facial features or biometric data to identify similarities with known individuals from a database, aiding in potential matches or identifications. Conversely, Random Forest, a subset of deep learning models, excel in learning intricate patterns and representations directly from image data. With their prowess in image analysis and feature extraction, Random forest holds promise in recognizing missing individuals based on visual cues.

The accuracy of the present research is not exactly existing in the system. The present research experiment is done to improve the accuracy of the K-Nearest Neighbor algorithm and contrast it with the existing algorithm. The K-Nearest Neighbor algorithm has been introduced to improve the accuracy of finding missing persons using ai and it is compared with the Random Forest. The advantages and disadvantages of both Machine Learning algorithms have been considered for the current research. In the present article the K-Nearest Neighbor algorithm is compared with the Random Forest algorithm for finding missing persons using ai.

**MATERIALS AND METHODS**

The current experimentation work has been carried out in the Machine Learning Laboratory at Saveetha School of Engineering, Saveetha Institute of Medical And Technical Sciences (SIMATS), Chennai. The sample size has been calculated using the ClinCalc tool under supervised learning with an alpha value of 80% or 0.8 and G-Power value of 0.8 and with a significance value of 0.05 at Confidential Interval (CI) of 95%. The sample size of 20 sets has been used for both the Group 1 i.e.; K-Nearest Neighbor algorithm and Group 2 i.e.; Random Forest algorithm, with a total of 40 sets being considered for this research (S. Ayyappan and S. Matilda)

**K-Nearest Neighbor algorithm**

K-NN is straightforward to understand and implement, making it an ideal starting point for beginners in machine learning. Its intuitive nature stems from the principle of similarity measurement—instances are classified based on their proximity to other instances in the dataset. It can be applied to both classification and regression tasks. In classification, it assigns a class label to an unseen instance based on the majority vote of its K nearest neighbors. In regression, it predicts a continuous value based on the average or weighted average of the K nearest neighbors.it doesn't involve a separate training phase; it stores all the training data instances, making predictions directly from the stored data. This can be advantageous for dynamic datasets where new data points are frequently added. The decision-making process in K-NN is easily interpretable, as predictions are based on nearby instances in the feature space. This transparency can be beneficial in understanding model predictions and explaining outcomes to stakeholders.

Figure 1 represents the block diagram for K-Nearest Neighbor algorithm. The working of K-Nearest Neighbor algorithm is:

1. Data which has been given as input is read and preprocessed.

2. Decision trees are trained on the preprocess data using the Knn.

3. All the decision trees which are trained are combined to prepare a final model.

4. The final output model achieved is used to make the predictions on the dataset.

**Pseudocode**

Step 1. Load the dataset of known individuals, including features like facial features, age, etc.

Step 2. Preprocess the data, ensuring it is in a suitable format for KNN (e.g., scaling features).

Step 3. Input the features of the missing person.

Step 4. Calculate the distance between the features of the missing person and all individuals in

The dataset

Step 5. Sort the distances in ascending order and select the top k nearest neighbors.

Step 6. Determine the majority class among the k nearest neighbors (e.g., using voting).

Step 7. Output the predicted class as the potential identity of the missing person.

**Random Forest**

The Random Forest algorithm operates by constructing a multitude of decision trees during the training phase. Each tree in the ensemble is built using a random subset of the available data and features, introducing diversity to the model. During the prediction phase, the algorithm aggregates the results of individual trees to make a more robust and accurate prediction. This ensemble approach helps mitigate overfitting and enhances the algorithm's generalization capabilities. In the realm of missing person’s cases, the application of Random Forest extends to various facets of the problem. The algorithm excels in analyzing complex datasets that may include diverse features such as geographical information, temporal patterns, behavioral data, and other relevant factors. By leveraging this ensemble method, Random Forest can provide a more comprehensive and nuanced understanding of the underlying patterns associated with missing persons.

**Pseudocode**

Step 1. Load the dataset of known individuals, including features like facial features, age, etc.

Step 2. Preprocess the data, ensuring it is in a suitable format for Random forest (e.g., scaling features).

Step 3. Input the features of the missing person.

Step 4. Calculate the distance between the features of the missing person and all individuals in

The dataset

Step 5. Sort the distances in ascending order and select the top Random forest neighbors.

Step 6. Determine the majority class among the Random forest (e.g., using voting).

Step 7. Output the predicted class as the potential identity of the missing person.

The present research has been carried out in the system which has hardware specification of Intel i5 as the core processor, RAM of 8 GB, and storage of 512 GB SSD followed by the software specifications which includes Windows 11, Jupyter Notebook, Chrome web browser and SPSS Software for the result analysis. Python programming language with a version of 3.10 has been implemented to obtain the required accuracy for the current research. The program has been executed in the Jupyter Notebook compiler in the current system.

The execution procedures given below,

1. Install the Jupyter Notebook with the help of Anaconda software.

2. A Jupyter notebook can be opened by using the command “jupyter notebook”.

3. Now create a new notebook by clicking on the top right corner.

4. Now write the required python code in the first cell of the notebook.

5. Import some of the required libraries such as numpy, pandas, matplotlib and seaborn.

6. Run the program by clicking the run button.

7. The accuracy should be noted in the excel sheet and run with the help of SPSS software.

The dataset has been collected from the Kaggle website which is an freely available platform which has been used by many of the machine learning and the data scientists students for various research purposes. The present dataset is named Finding Missing persons using Ai The Data source link is: (“Kaggle” 2022). The present Dataset consists of 11 columns (attributes) and 10684 rows. The column names in the Dataset are as below:

* Name
* Gender
* Relative
* Age Start
* AgeEnd
* Height Start
* HeightEnd
* Built
* Dist.
* State
* Found

The above dataset has both dependent and independent columns. The independent columns from the above dataset are: Name, Gender, Dist, AgeStart, HeightStart,Built,Dist. The dependent columns which have been collected from the above dataset are: Relative, Found, State,AgeEnd,HeightEnd. The dataset which has been considered here is in the text form and there are no images or audio files included in the complete dataset. The dataset training and testing are planned in the ratio of 80:20 respectively.

**Statistical Analysis**

The statistical analysis for the current research is carried out using the SPSS software of version 26. The values which are collected in excel sheet are inserted into the SPSS software for the analysis of Independent samples T-test among the K-Nearest Neighbor and Random forest algorithm. The independent variables are Name, Gender, Dist, AgeStart, Height Start, Built, and Dist. The dependent variables are Relative, Found, State, AgeEnd, HeightEnd. The Independent samples T-Test analysis has been performed by analyzing the above collected data between the Novel KNN algorithm and RF algorithm (S. Ayyappan and S. Matilda 2020).

**RESULTS**

Table 1 shows the T-Test results of the proposed K-Nearest Neighbor algorithm and the Convolutional Neural Network algorithm which has been run numerous times in the Jupyter notebook with a sample size of 20. From Table 1, it has been observed that the accuracy of the algorithm is K-Nearest Neighbor algorithm 87.20% and for the Random forest algorithm the accuracy is found to be 74.92%. The standard deviation and the Standard Error Mean has also been calculated for the K-Nearest Neighbor algorithm and Random forest algorithm

Table 2 represents the outcome of the analysis of the Independent samples test which has been performed for the K-Nearest Neighbor algorithm and the Random forest algorithm. From Table 2, the significance value for the one tailed test is found to be 0.937, two-tailed is 0.000 and it is found

That the Independent samples test has been carried out at Confidence Interval of 95%.

Figure 2 shows the comparison graph between the K-Nearest Neighbor algorithm and the Random forest algorithm. From the graph, it is concluded that our proposed K-Nearest Neighbor algorithm has an accuracy of 87.20% and the Random forest algorithm has an accuracy of 74.92%. The plots of the graph are shown in the figure below in Fig. 1

**DISCUSSION**

In this research the K-Nearest Neighbor algorithm is compared with the algorithm to predict the future missing persons using ai and to enhance the accuracy of the existing system. By performing the experiment K-Nearest Neighbor algorithm has achieved an accuracy of 87.20% and Random forest has achieved an accuracy of 74.92%. The significance value for this research is found to be 0.000 after performing the Independent samples T-test analysis. The K-Nearest Neighbor algorithm (87.20%) has been found to be more accurate than the Random forest (74.92%). The significance value is found to be 0.000 which is lesser than 0.05 (p<0.05), therefore it is observed that the two groups are statistically significant.

In the recent survey, the K-Nearest Neighbor algorithm has been found to have more promising accuracy than the other real world algorithms (S. Ayyappan and S. Matilda 2020). The present framework will combine the two datasets with the data collected from the users and found that theK-Nearest Neighbor has provided the best accuracy (Shefali patil, Pratiksha Gaikar 2021).K-Nearest Neighbor algorithm has been significantly faster than the other gradient boosting methods and has more precise accuracy by all means. The results can be developed by implementing new features and picking the best data set (Sayan Deb Sarkar and Ajitha Shenoy). Some of the researchers have proposed theK-Nearest Neighbor algorithm in some of their research articles and concluded that the K-Nearest Neighbor algorithm has provided better results than the other Machine Learning algorithms (Y. Lin, L. Zheng, and Z. Zheng 2019). Some of the articles have proposed the Random Forest algorithm to forecast the future charges of finding missing persons using ai and found it had provided better accuracy in some cases than our proposed (S. Abhilash and V. M. Nookala 2022). In some research surveys some of the researchers have implemented the K-Nearest Neighbor (KNN) algorithm to provide future prices of flight tickets to the customers in an efficient way and it has been implemented and found out that it provided more accurate results than our Novel XGBRegression algorithm (X. Huang, S. Hu 2021).

**CONCLUSION**

The aim of the present experimentation research is to improve the accuracy of finding missing persons using ai and help the customers to book the tickets accordingly. In this research article the K-Nearest Neighbor algorithm is compared with the Convolutional Neural Network algorithm. Results which had been obtained showed that the K-Nearest Neighbor algorithm has provided an accuracy of 87.20% and Convolutional Neural Network algorithm has recorded 75.25% of accuracy

**DECLARATIONS**

**Conflict of Interests**

No conflict of Interest in this manuscript.

**Authors Contributions**

Author R.Rupasridevi was involved in data collection, data analysis and manuscript writing. Author C.Clement raj was involved in the 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.

**Funding**

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

1. Mind Play Software Services Pvt. Ltd, Chennai

2. Saveetha School of Engineering

3. Saveetha Institute of Medical and Technical Sciences

4. Saveetha University

**REFERENCES**

Abhinay Chaukade, “Finding and Matching the Lost Victim Using AI and ML” International Journal for Research in Applied science and technology, 2022

D.J. Samatha Naidu , R. Lokesh, “Missing Child Identification System using Deep Learning with VGG FACE Recognition Technique” SSRG International Journal of Computer Science and Engineering, 2022

G. Srikanth, Adurti Swarnalatha ,Thalari Abhishek , Ravula Sai Akhil Patel ,Thalari Swamy,” Missing Person Identification using Machine Learning with Python”, International Journal for Research in Applied Science & Engineering Technology, 2022

Roshin John, Basil Kuriakose , “Face Detection and Tracking to Find the Missing Person”, International Journal of Research in Engineering, Science and Management, 2020 [10] Sayan Deb Sarkar and Ajitha Shenoy. “Face Recognition using Artificial Neural Network and Feature Extraction”, 2020

Sanskar Pawar, Lalit Bhadane, Amanullah Shaikh, Atharv Kumbhejkar, Swati Jakkan, “FIND MISSING PERSON USING ARTIFICIAL INTELLIGENCE” In this paper we will use the ideas of the facial recognition algorithm which is based on Artificial Intelligence (AI) and will detect faces with the maximum accuracies to find the missing person, 2021.

S. Ayyappan and S. Matilda, “Criminals and missing children identification using face recognition and web scrapping” IEEE ICSCAN 2020.

Shefali patil, Pratiksha Gaikar, Divya Kare, Sanjay Pawar, “Find missing person using AI (ANDROID APPLICATION)”, International journal of progressive Research in Science and Engineering, Vol. 2, No. 6, June 2021.

Sarthak Babbar, Navroz Dewan, Kartik Shangle, Sudhanshu Kulshreshtra, Sanjeev Patel, “Cross Age Face recognition using Deep Residual Networks “. IEEE 2019 Fifth International Conference on Image Information Processing (ICIIP).

Vijay Deep,Govind Kaushik “Finding Missing Person Using Face Recognition System - Khoj” May 2022.

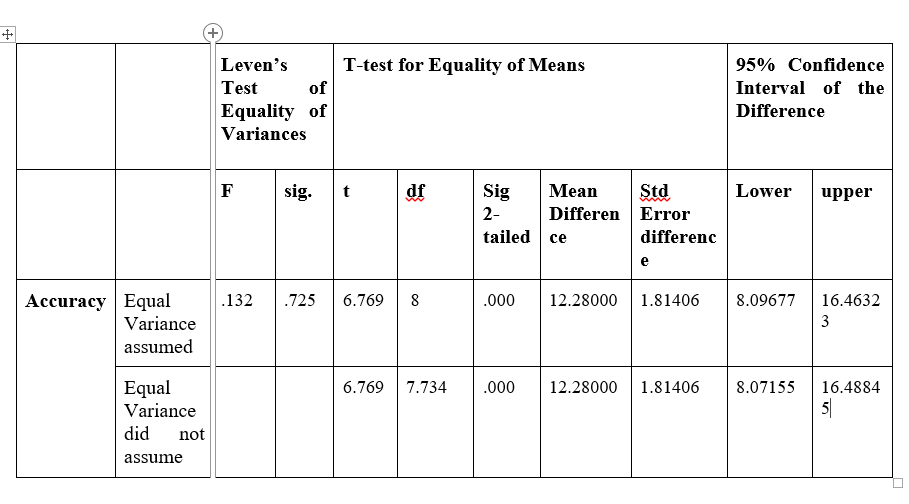
**TABLES AND FIGURES**

Table 1. Presents the statistical analysis results of the K-Nearest Neighbor algorithm and the Random forest algorithm, comparing the mean accuracy, standard deviation, and standard error mean values across 20 sample datasets.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Algorithm** | **N** | **Mean** | **Std.Deviation** | **Std.Error Mean** |
| **Accuracy** | **KNN** | **5** | **87.2000** | **2.58844** | **1.15758** |
| **CNN** | **5** | **74.9200** | **3.12316** | **1.39672** |

Table 2. An independent sample T-Test was conducted to determine the significance of the difference between the two groups, using a significance level of p=0.000 (p<0.05), indicating that the difference is statistically significant

**Independent Samples Test**



|  |
| --- |
|  |
|  |
|
|

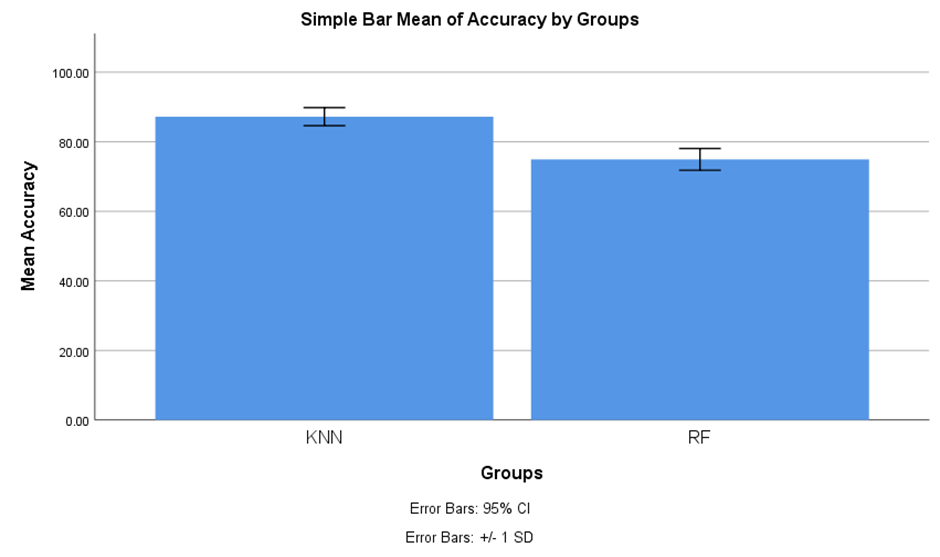


Fig. 1. This figure shows the comparison between the K-Nearest Neighbor algorithm and the Random forest algorithm in terms of Mean Accuracy. The Mean accuracy of the K-Nearest Neighbor is better than the Mean accuracy of the Random forest algorithm. X-axis: K-Nearest Neighbor algorithm vs Random forest algorithm, Y-axis: Mean Accuracy. Error Bar +/-1SD.