IMPROVED ACCURACY FOR INFORMATION EXTRACTION AND DIGITALIZATION OF HANDWRITTEN DOCUMENTS USING DECISION TREE COMPARED OVER RANDOM FOREST

B.Upendra[1], P.V.Pramila[2]

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

Abstract

Aim: The research work aims to improve the accuracy of the information extraction and digitalization of human handwritten character documents using Decision Trees with machine learning algorithms. **Materials and Methods:** The categorizing is performed by adopting a sample size of N=20 in the Decision Tree and sample size of N=20 in Random Forest algorithms. **Results:** Novel Decision Tree delivered significant results with 95.00% accuracy, compared to Random Forest 93.33% accuracy. Novel Decision Tree and Random Forest statistical significance is $p = 0.57$ ($p < 0.05$). The Independent sample T-test value states that the results in the study are significantly not achieved with a 95% confidence level. **Conclusion:** Information extraction and digitalization of human handwritten character documents with the Novel Decision Tree provides better accuracy than the Random Forest algorithm.

Keywords: Handwritten Characters, Information Extraction, Digitalization, Novel Decision Tree, Random Forest, Machine Learning, Accuracy.

[1]Research Scholar, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical And Technical Sciences, Chennai, Tamil Nadu, India, Pincode: 602105
[2*]Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical And Technical Sciences, Chennai, Tamil Nadu, India, Pincode: 602105

1. Introduction

The recognition of handwritten characters is a part of computer vision and pattern recognition research. (Mainkar et al. 2020) A human being's intelligence, learning skills, and interpretation abilities help him to understand different forms of handwriting. (Dessai and Patil 2019) People can easily identify letters and printed records. It is much more difficult to identify ancient documents, especially those written in many different languages and with a variety of handwriting styles. (Raundale and Maredia 2021) To overcome this challenge, a method that can be used is converting handwritten characters into a machine-readable format with the help of some effective methods, techniques, and machine learning algorithms. A wide range of applications exists for human handwriting recognition, such as recognizing postal addresses, bank check amounts, forms, and handwritten paper documents (Oktaviani et al. 2020).

Human handwritten character recognition documents were implemented by many researchers. There were around 495 Conferences, 20 Journal articles were published in IEEE, and 85 papers in Science direct. (Dessai and Patil 2019) Handwritten Character Recognition is one of the most difficult areas of study for pattern recognition and image processing. (Negi and Rao 2019) The research also proposes Convolutional Neural Network methods, along with Visual Attention Network approaches, are used to recognise Telugu handwritten characters. (Raundale and Maredia 2021) Data can be transferred from paper documents into a digital storage system by scanning them and storing them as images. But the difficulty arises when we need to re-use this data since it becomes difficult to interpret particular data from these documents. (Raundale and Maredia 2021; Bhagyasree, James, and Saravanan 2019) OCR is the identification of human handwritten documents and printed documents using the computer. Pattern Recognition and optical character recognition use Handwritten Character Recognition. It has a wide variety of applications in different fields. Handwritten Character Recognition helps significantly to the advancement of automation and is relevant in the areas of tax filings, medical prescriptions, and bank checks. Reading written or printed documents is easy for humans, this ability can be induced in the machine using Optical Character Recognition technique (Shelke and Apte 2016).

Our institution is keen on working on latest research trends and has extensive knowledge and research experience which resulted in quality publications (Rinesh et al. 2022; Sundararaman et al. 2022; Mohanavel et al. 2022; Ram et al. 2022; Dinesh Kumar et al. 2022; Vijayalakshmi et al. 2022; Sudhan et al. 2022; Kumar et al. 2022; Sathish et al. 2022; Mahesh et al. 2022; Yaashikaa et al. 2022). The biggest challenging problem in the recognition of a wide variety of human handwritten characters is that every individual has different sizes and styles. Additionally, there are a variety of other factors that create differences in human handwriting, such as connected components, multi-oriented letters, overlapping characters, skewness of text lines, and pressure points among other factors. Several scripts have intrinsic variations. Even a simple character can be written differently. Thus, recognizing a particular handwritten character is a challenging task. Nowadays, machine learning algorithms are able to recognize human handwritten character documents in several ways and convert them into digitalized documents, which are in a variety of handwriting styles.

2. Materials And Methods

This research was conducted in the Image Processing Laboratory of the Saveetha School of Engineering, an institute under the Saveetha Institute of Medical and Technical Sciences. The algorithms are evaluated using two group classifiers, Decision Tree and Random Forest which are used to convert human handwritten character documents into digitalization (BahetiI 2013; Rath and Manmatha 2007; Sunderiyal and Yadav 2013)). The Decision Tree technique is used in Group 1 with a sample size of 20. The Random Forest technique is used in Group 2 with a sample size of 20. Using G power, 20 Sample sizes and a total of 40 sample sizes are carried out with 95% confidence and 80% pretest power.

An Intel Core i5 CPU, 8GB Random Access Memory, 512GB Solid State Drive, X64 based processor, and 64Bit operating system are used in the hardware configurations. Windows 10 and a jupyter notebook with the Python programming language were used as software configurations. IBM Statistical Package for the Social Sciences (SPSS) is used to implement and predict it.

This study used the "A_Z Handwritten Alphabets dataset" from www.kaggle.com. English alphabets are Dependent variables and their styles, strokes, and sizes are Independent variables. It was compared for more accuracy scores and less loss values for choosing the best algorithm suitable for the digitalization of human handwritten characters into digitalized documents (Noubigh, Mezghani, and Kherallah 2020).

Decision Tree

Novel Decision Trees are decision-making tools that utilize a tree-like model of alternatives and prospective outcomes, such as chance event outcomes, usable resource costs, and utility. It is one way of showing an algorithm that mainly consists of conditional control statements. Decision trees are frequently used in operations research, specifically in decision analysis and operation management, to help in finding the most likely approach to achieving a goal. Indecision analysis, a selection tree, and the influence diagram are used as visual and analytical decision support tools. Each internal node represents a "test" on an attribute, each branch shows the test's outcome, and each leaf node represents a class label. Because of the great preferred version or online choice version set of rules, the pathways from root to leaf indicate classifications rules, and options must be made online with no recall under insufficient knowledge. Because of the good wanted version or online choice version set of rules, the Decision Tree must be mirrored by a chance version. It is a descriptive method of estimating conditional probabilities.

Pseudocode for Decision Tree

Step 1 : Keras Library has been used to import the essential classes, Such as Sequential, Dense, Flatten, Conv2D, MaxPool2D, and Dropout
Step 2: Import the dataset as alphabets.csv
Step 3: Analyze the handwritten alphabets stored in images.
Step 4: After loading the dataset, it was splitted into training and testing samples.
Step 5: Preparing the data for training by using a NumPy array data and reshaping it.
Step 6: Imported relevant libraries and modules. Next, we splitted the data into half training and half testing using the function train_test_split.
Step 7: The model is trained and tested to ensure accuracy.
Step 8: Import train_test_split into the coding.
Step 9: We use Decision Tree algorithms as a classifier and make predictions to print a Decision Tree Classification Report and Accuracy Score.

Random Forest

Random forests are enhanced decision tree models that provide supervised learning. Random forests use a large number of decision trees that work together to predict the outcome of a class, with the final prediction based on the class with the most votes and correcting the decision tree's tendency to adapt to their training set. Random forests outperform decision trees on average, although their accuracy is lower than that of gradient-enhanced trees. There are multiple ways for selecting a division in a decision tree based on a regression or classification problem, and the features of the data might influence their performance. Because of the poor association among trees, random forests have a low error rate when compared to other models. Our random forest model was trained in a grid search using various parameters, such as altering the number of estimators, to discover the best model that can reliably predict the outcome. Random forests represent a shift towards bagged decision trees to create a large number of decor-related trees to improve predictive performance and to have few hyperparameters. Many random forest implementations are available. Random forests produce an average predictive value as a result throughout the regression process.

Pseudocode for Random Forest

Step 1 : Keras Library has been used to import the essential classes, Such as Sequential, Dense, Flatten, Conv2D, MaxPool2D, and Dropout
Step 2: Import the dataset as alphabets.csv
Step 3: Analyze the handwritten alphabets stored in images.
Step 4: After loading the dataset, it was splitted into training and testing samples.
Step 5: Preparing the data for training by using a NumPy array data and reshaping it.
Step 6: Imported relevant libraries and modules. Next, we splitted the data into half training and half testing using the function train_test_split.
Step 7: The model is trained and tested to ensure accuracy.
Step 8: Import train_test_split into the coding.
Step 9: We use Random Forest algorithms as a classifier and make predictions to print a Random Forest Classification Report and Accuracy Score.

Statistical Analysis

The datasets are prepared in SPSS using a sample size of N=20 for Decision Tree along with Random Forest. The testing variable is specified as GroupID, Accuracy, and Loss. GroupID for Decision Tree is Group 1 and Group 2 for Random Forest. Group Statistics is used for the Statistical Package for the Social Sciences (SPSS) dataset. By performing the statistical analysis of group statistics, By using Decision Tree and Random Forest, we evaluate the accuracy and loss of information extraction and digitalization of human handwritten documents (Tabassum et al. 2021).

3. Results

The Decision Tree algorithm is used to develop a system that automatically classifies without the need for human intervention. Each algorithm's performance is measured using metrics. After performing the training and classification operations with the collected dataset, because of the initialization of a sample size of N=20, the accuracy value changes with the duration of the running time and provides the accuracy and loss for the period. The Decision Tree algorithm analyzes the data of each alphabet character individually and provides the final result, which performed better than the Random Forest algorithm. This analysis only uses the bare minimum preprocessing required to execute the algorithm, however this research takes it to the next level, resulting in higher accuracy and less loss, which has proven that the Novel Decision Tree is better than Random Forest.

Table1 represents the N=20 samples of the dataset for Decision Tree with gain accuracy (%) and loss reduction (%) and Random Forest with gain accuracy (%) and loss reduction (%).

Table 2 represents a comparison of the accuracy and loss of information extraction and digitalization of human handwritten character documents of Decision Tree and Random forest. The Decision Tree algorithm got the highest accuracy of 95.00% and the lowest loss of 1.33%. Random Forest had the lowest accuracy of 93.33% and the highest loss of 2.34%.

Table 3 represents the independent sample test is used to determine significance and standard error. A p value of less than 0.05 was considered statistically significant, and 95% confidence intervals were calculated. The p value is 0.57, the mean difference is 1.67, and the confidence interval is (0.65 - 0.06). Decision Tree and Random Forest are significantly different.

4. Discussion

The prediction of a word and identifying the word accuracy were aimed in this work. The overall results show that there exists some differences observed in the accuracy and loss values of the models used in this analysis. The Novel Decision Tree showed a better accuracy of 95.00% than the Random Forest accuracy of 93.33% in information extraction and digitalization of human handwritten documents. There is a statistically significant difference in the accuracies of the two algorithms in the innovative digitization of human handwritten character documents, with a $p = 0.57$ ($p<0.05$ Independent Sample t-Test) states that the results in the study are significantly not achieved with a 95% confidence level.

(Darmatasia and Fanany 2017) presented a methodology and a machine learning model for handwritten characters on form paper documents. The learning model is based on Convolutional Neural Networks and Support Vector Machines as a high-end classifier. The output of the system is converted into editable text. Preprocessing, segmentation, and character recognition are all handled by a single system. Ten separate test form papers are correct to 83.37%, according to the system. (Alkhateeb, Turani, and Alsewari, 2020) uses machine learning and deep learning methodologies to recognise Arabic handwritten text. They conducted several experiments on two distinct datasets, AHCR and ADBase, employing both machine learning and deep learning. Deep learning surpassed machine learning by 91% in both datasets, according to the experimental results. (Darapaneni et al. 2020) the handwritten forms are scanned, preprocessed, and handwritten fields are extracted. OpenCV is used to get the contours of the characters in the extracted images. This approach gives better accuracy than using plain CNN without counters. The CNN model gives an accuracy of 90% on the merger of numbers, uppercase, and lowercase alphabets of the EMINST dataset. (Gupta and Bag 2021) used deeper neural networks and they obtained an accuracy of 96.2%. But their network model usually required more time to train data. (Arora and Chandratre 2018) specifically proposed a localized zonal technique to character identification, which has been found to significantly improve recognition accuracy levels. Based on the EMNIST balanced dataset, this method accounts for character contextual placement by using two independent convolutional neural networks (alphanumeric and numeric) with test accuracies of 97.2% and 76.8%, respectively.

5. Conclusion

In this research, the advanced innovative digitalization of human handwritten character documents was performed by using the A_Z Handwritten Alphabets dataset using Decision Tree over the Random Forest. The present study focuses on the machine learning techniques for higher classification in digitalization of human handwritten documents. In the future, it can be slightly enhanced based on the study of random data sets. The outcome of the study Decision Tree has higher accuracy (95.00%) than Random Forest (93.33%). The clarity of information extraction and digitalization of human handwritten characters is found with good accuracy and less loss.

DECLARATIONS

Conflict of Interests

No conflict of interest.

Author's Contribution

Author BU was involved in data analysis, data collecting, and paper writing. Author PVP was involved in the data validation, concept development, and critical review of the manuscript.

Acknowledgments

The author's would like to thank Saveetha School of Engineering, an institute under the Saveetha Institute of Medical And Technical Sciences (Formerly known as Saveetha University) for providing all the facilities required to complete this research work successfully.

Funding

We would like to thank the following organization for providing financial support that helped us to complete the study.

1. SNEW.AI Technologies, Chennai.
2. Saveetha University.
3. Saveetha Institute of Medical And Technical Sciences.
4. Saveetha School of Engineering.

6. References

Arora, Anushri, and Aniruddh Chandratre. 2018. "Optical Character Recognition for Handwritten Forms with Dynamic Layout." In *2018 4th International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT)*. IEEE. https://doi.org/10.1109/icatcct44854.2018.9001954.

BahetiI, Mamta. 2013. *Handwritten Character Recognition: A Milestone to Achieve*. LAP Lambert Academic Publishing.

Bhagyasree, P. V., Ajay James, and Chandran Saravanan. 2019. "A Proposed Framework for Recognition of Handwritten Cursive English Characters Using DAG-CNN." In *2019 1st International Conference on Innovations in Information and Communication Technology (ICIICT)*. IEEE. https://doi.org/10.1109/iciict1.2019.8741412.

Darapaneni, Narayana, Malarvizhi Subramaniyan, Aafia Mariam, Sai Venkateshwaran, Nandini Ravi, Anwesh Reddy Paduri, Sumathi Gunasekaran, and Asha. 2020. "Handwritten Form Recognition Using Artificial Neural Network." In *2020 IEEE 15th International Conference on Industrial and Information Systems (ICIIS)*. IEEE. https://doi.org/10.1109/iciis51140.2020.9342638.

Darmatasia, and Mohamad Ivan Fanany. 2017. "Handwriting Recognition on Form Document Using Convolutional Neural Network and Support Vector Machines (CNN-SVM)." In *2017 5th International Conference on Information and Communication Technology (ICoIC7)*. IEEE. https://doi.org/10.1109/icoict.2017.8074699.

Dessai, Brijeshwar, and Amit Patil. 2019. "A Deep Learning Approach for Optical Character Recognition of Handwritten Devanagari Script." In *2019 2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT)*. IEEE. https://doi.org/10.1109/icicict46008.2019.8993342.

Dinesh Kumar, M., V. Godvin Sharmila, Gopalakrishnan Kumar, Jeong-Hoon Park, Siham Yousuf Al-Qaradawi, and J. Rajesh Banu. 2022. "Surfactant Induced Microwave Disintegration for Enhanced Biohydrogen Production from Macroalgae Biomass: Thermodynamics and Energetics." *Bioresource Technology* 350 (April): 126904.

Gupta, Deepika, and Soumen Bag. 2021. "CNN-Based Multilingual Handwritten Numeral Recognition: A Fusion-Free Approach." *Expert Systems with Applications*. https://doi.org/10.1016/j.eswa.2020.113784.

Kumar, J. Aravind, J. Aravind Kumar, S. Sathish, T. Krithiga, T. R. Praveenkumar, S. Lokesh, D. Prabu, A. Annam Renita, P. Prakash, and M. Rajasimman. 2022. "A Comprehensive Review on Bio-Hydrogen Production from Brewery Industrial Wastewater and Its Treatment Methodologies." *Fuel*. https://doi.org/10.1016/j.fuel.2022.123594.

Mahesh, Narayanan, Srinivasan Balakumar, Uthaman Danya, Shanmugasundaram Shyamalagowri, Palanisamy Suresh Babu, Jeyaseelan Aravind, Murugesan Kamaraj, and Muthusamy Govarthanan. 2022. "A Review on Mitigation of Emerging Contaminants in an Aqueous Environment Using Microbial Bio-Machines as Sustainable Tools: Progress and Limitations." *Journal of Water Process Engineering*. https://doi.org/10.1016/j.jwpe.2022.102712.

Mainkar, Vaibhav V., Jyoti A. Katkar, Ajinkya B. Upade, and Poonam R. Pednekar. 2020. "Handwritten Character Recognition to Obtain Editable Text." In *2020 International Conference on Electronics and Sustainable*

Communication Systems (ICESC). IEEE. https://doi.org/10.1109/icesc48915.2020.9155786.

Mohanavel, Vinayagam, K. Ravi Kumar, T. Sathish, Palanivel Velmurugan, Alagar Karthick, M. Ravichandran, Saleh Alfarraj, Hesham S. Almoallim, Shanmugam Sureshkumar, and J. Isaac JoshuaRamesh Lalvani. 2022. "Investigation on Inorganic Salts K2TiF6 and KBF4 to Develop Nanoparticles Based TiB2 Reinforcement Aluminium Composites." *Bioinorganic Chemistry and Applications* 2022 (January): 8559402.

Negi, Atul, and Anish M. Rao. 2019. "Offline Handwritten Telugu Character Dataset and Recognition." In *2019 IEEE 16th India Council International Conference (INDICON)*. IEEE. https://doi.org/10.1109/indicon47234.2019.9028977.

Noubigh, Zouhaira, Anis Mezghani, and Monji Kherallah. 2020. "Transfer Learning to Improve Arabic Handwriting Text Recognition." In *2020 21st International Arab Conference on Information Technology (ACIT)*. IEEE. https://doi.org/10.1109/acit50332.2020.9300105.

Oktaviani, Selly, Christy Atika Sari, Eko Hari Rachmawanto, and De Rosal Ignatius Moses Setiadi. 2020. "Optical Character Recognition for Hangul Character Using Artificial Neural Network." In *2020 International Seminar on Application for Technology of Information and Communication (iSemantic)*. IEEE. https://doi.org/10.1109/isemantic50169.2020.9234215.

Ram, G. Dinesh, G. Dinesh Ram, S. Praveen Kumar, T. Yuvaraj, Thanikanti Sudhakar Babu, and Karthik Balasubramanian. 2022. "Simulation and Investigation of MEMS Bilayer Solar Energy Harvester for Smart Wireless Sensor Applications." *Sustainable Energy Technologies and Assessments*. https://doi.org/10.1016/j.seta.2022.102102.

Rath, Toni M., and R. Manmatha. 2007. "Word Spotting for Historical Documents." *International Journal of Document Analysis and Recognition (IJDAR)*. https://doi.org/10.1007/s10032-006-0035-8.

Raundale, Pooja, and Hadi Maredia. 2021. "Analytical Study of Handwritten Character Recognition: A Deep Learning Way." In *2021 International Conference on Intelligent Technologies (CONIT)*. IEEE. https://doi.org/10.1109/conit51480.2021.9498347.

Rinesh, S., K. Maheswari, B. Arthi, P. Sherubha, A. Vijay, S. Sridhar, T. Rajendran, and Yosef Asrat Waji. 2022. "Investigations on Brain Tumor Classification Using Hybrid Machine Learning Algorithms." *Journal of Healthcare Engineering* 2022 (February): 2761847.

Sathish, T., V. Mohanavel, M. Arunkumar, K. Rajan, Manzoore Elahi M. Soudagar, M. A. Mujtaba, Saleh H. Salmen, Sami Al Obaid, H. Fayaz, and S. Sivakumar. 2022. "Utilization of Azadirachta Indica Biodiesel, Ethanol and Diesel Blends for Diesel Engine Applications with Engine Emission Profile." *Fuel*. https://doi.org/10.1016/j.fuel.2022.123798.

Shelke, Sushama, and Shaila Apte. 2016. "Performance Optimization and Comparative Analysis of Neural Networks for Handwritten Devanagari Character Recognition." *2016 International Conference on Signal and Information Processing (IConSIP)*. https://doi.org/10.1109/iconsip.2016.7857482.

Sudhan, M. B., M. Sinthuja, S. Pravinth Raja, J. Amutharaj, G. Charlyn Pushpa Latha, S. Sheeba Rachel, T. Anitha, T. Rajendran, and Yosef Asrat Waji. 2022. "Segmentation and Classification of Glaucoma Using U-Net with Deep Learning Model." *Journal of Healthcare Engineering* 2022 (February): 1601354.

Sundararaman, Sathish, J. Aravind Kumar, Prabu Deivasigamani, and Yuvarajan Devarajan. 2022. "Emerging Pharma Residue Contaminants: Occurrence, Monitoring, Risk and Fate Assessment – A Challenge to Water Resource Management." *Science of The Total Environment*. https://doi.org/10.1016/j.scitotenv.2022.153897.

Sunderiyal, Vinita Dutt, and Ajit Kumar Singh Yadav. 2013. *Handwritten Character Recognition Using Artificial Neural Network: System for Handwritten Character Recognition, Bring Out Accurate Results Even for Noisy Image*. LAP Lambert Academic Publishing.

Tabassum, Shaira, Ryo Takahashi, Md Mahmudur Rahman, Yosuke Imamura, Luo Sixian, Md Moshiur Rahman, and Ashir Ahmed. 2021. "Recognition of Doctors' Cursive Handwritten Medical Words by Using Bidirectional LSTM and SRP Data Augmentation." In *2021 IEEE Technology & Engineering Management Conference - Europe (TEMSCON-EUR)*. IEEE. https://doi.org/10.1109/temscon-eur52034.2021.9488622.

Vijayalakshmi, V. J., Prakash Arumugam, A. Ananthi Christy, and R. Brindha. 2022. “Simultaneous Allocation of EV Charging Stations and Renewable Energy Sources: An Elite RERNN-m2MPA Approach.” *International Journal of Energy Research*. https://doi.org/10.1002/er.7780.

Yaashikaa, P. R., P. Senthil Kumar, S. Jeevanantham, and R. Saravanan. 2022. “A Review on Bioremediation Approach for Heavy Metal Detoxification and Accumulation in Plants.” *Environmental Pollution* 301 (May): 119035.

Tables and Figures

Table 1: The N=20 samples of the dataset for Decision Tree with gain accuracy (%) and loss reduction (%) and Random Forest to gain accuracy (%) and loss reduction (%).

Samples	**Decision Tree**		**Random Forest**	
(N)	**Accuracy (%)**	**Loss (%)**	**Accuracy (%)**	**Loss (%)**
1	91.9	1.65	89.7	2.69
2	92.2	1.63	90.2	2.65
3	92.5	1.60	90.5	2.62
4	92.8	1.57	90.7	2.59
5	93.0	1.53	91.4	2.55
6	93.4	1.50	91.7	2.50
7	93.8	1.47	92.0	2.46
8	94.1	1.43	92.4	2.41
9	94.5	1.39	92.7	2.38
10	94.9	1.35	93.3	2.34
11	95.3	1.31	93.6	2.30
12	95.7	1.29	94.1	2.29
13	96.0	1.24	94.5	2.25
14	96.2	1.20	94.7	2.23
15	96.5	1.18	95.0	2.20
16	96.7	1.15	95.5	2.17
17	97.2	1.10	95.7	2.12
18	97.5	1.07	96.0	2.09
19	97.8	1.03	96.3	2.05
20	98.0	1.00	96.6	2.01

Table 2: Comparison of the accuracy and loss of information extraction and digitalization of human handwritten character documents of Decision Tree and Random forest. The Decision Tree algorithm got the highest accuracy of 95.00% and the lowest loss of 1.33%. Random Forest had the lowest accuracy of 93.33% and the highest loss of 2.34%.

	Groups	**N**	**Mean**	**Std.Deviation**	**Std Error Mean**
Accuracy	**Decision Tree**	20	95.00	1.96	0.44
	Random Forest	20	93.33	2.18	0.48
Loss	**Decision Tree**	20	1.33	0.20	0.04
	Random Forest	20	2.34	0.20	0.04

Table 3: Independent sample test is used to determine significance and standard error. A p value of less than 0.05 was considered statistically significant, and 95% confidence intervals were calculated. The p value is 0.57, the mean difference is 1.67, and the confidence interval is (0.65 - 0.06). Decision Tree and Random Forest are significantly different.

		Levene's Test for Equality of Variance		**t-test for Equality of Means**						
				t	**df**	**Sig (2-tailed)**	**Mean Difference**	**Std. Error Difference**	**95% Confidence Interval of the**	
		F	**Sig**						**Lower**	**Upper**
Accuracy	**Equal Variance Assumed**	0.327	0.57	2.53	38	0.04	1.67	0.65	0.33	3.00
	Equal Variance Not Assumed			2.53	37.58	0.04	1.67	0.65	0.33	3.00
Loss	**Equal Variance Assumed**	0.008	0.92	-15.32	38	0.00	-1.01	0.06	-1.14	-0.87
	Equal Variance Not Assumed			-15.32	38.00	0.00	-1.01	0.06	-1.44	-0.87

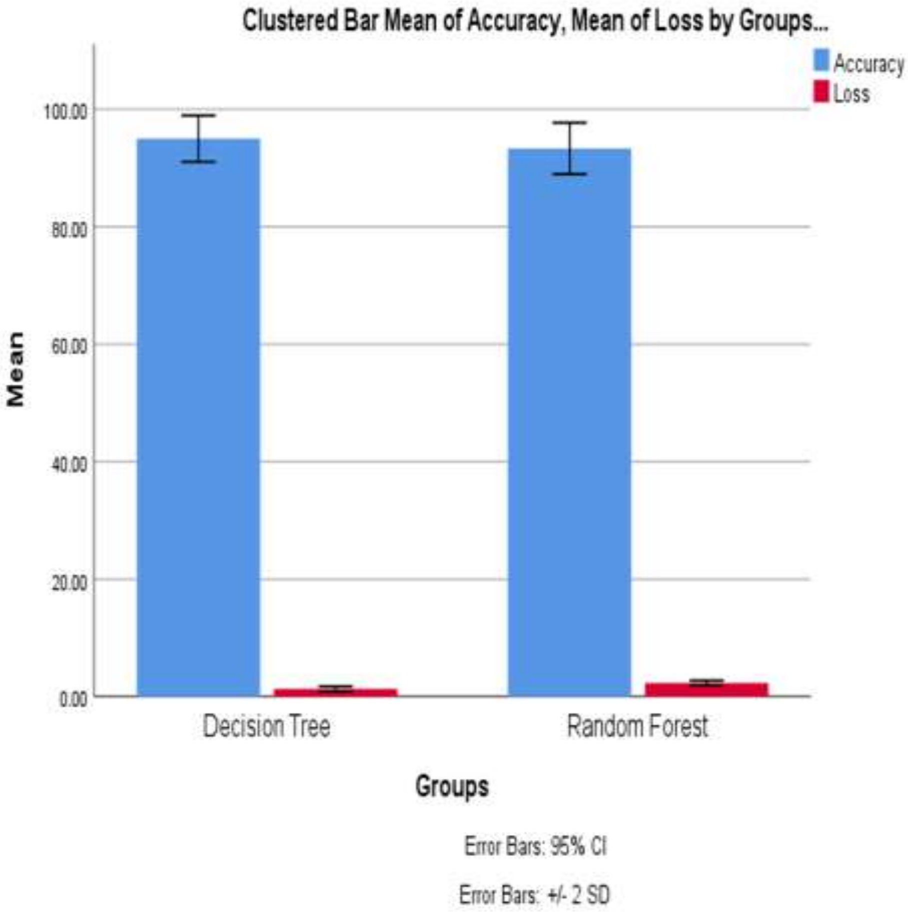

Fig 1: The mean accuracy of the Decision Tree and Random Forest Machine Learning algorithms was compared. The Decision Tree algorithm shows a better mean accuracy than the Random Forest model with moderately improved standard deviation. Graphically represented in X-axis: Decision Tree Vs Random Forest, Y-axis: Accuracy and Loss detection also depeting the error bars with +/- 2 Standard Deviation.