ENHANCING ACCURACY IN TEXT CLASSIFICATION ON CORONAVIRUS TWEETS USING NOVEL RANDOM FOREST CLASSIFIER COMPARED WITH DECISION TREE ALGORITHM

N. Naveen Reddy[1], K.Malarkodi[2*]

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

Abstract

Aim: To enhance accuracy in text classification on coronavirus tweets using Novel Random Forest Classifier compared with Decision Tree.
Materials and Methods: This study contains 2 groups i.e. Novel Random Forest Classifier and Decision tree algorithm. Each group consists of 10 sample sizes i.e. Novel Random Forest Classifier (Group 1=10) and Decision Tree (Group 2=10). The study parameters include alpha value 0.05, beta value 0.2, and the G power value 0.8. **Results:** The Novel Random Forest is (81.60%) more accurate than the Decision Tree of (80.14%) in classifying the coronavirus tweets with $p<0.717$.
Conclusion: The Novel Random Forest model is significantly better than the Decision Tree in identifying Text Classification on Coronavirus Tweets.

Keywords: Text Classification, Coronavirus Tweets, Novel Random Forest, Machine Learning, Accuracy, Decision Tree Algorithm.

[1]Research Scholar, Department of Information Technology, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamilnadu, India, Pincode: 602105.
[2*]Department of Information Technology, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamilnadu, India, Pincode: 602105.

1. Introduction

The coming of individual-to-individual contact organizations (Social Network Service, SNS) has cleared the way for people to interface and send messages in a quick and relaxed way to their companions, associates, and the overall population on various issues. They generally do it, as the expense of sending these short messages is little (Pelka, Nensa, and Friedrich 2019). Moreover, for instance, SNS likewise executed rich information about its clients, posting client-made text, client logs, and client relationships with clients. People share their perspectives, ideas, occasions, recordings, applications, and exercises utilizing SNS. Within the excess of 500,000,000 assorted clients on Twitter, SNS has ended up to be an eminent correspondence and information search device, and Twitter is perhaps the most famous web-based media platform today (Nay 2017). The tweet is a message restricted to 140 characters (presently 240 characters) that hoist clients to post data and messages about various interests while conventional publishing content to a blog seems to set aside more committed effort to compose a post (Miyamoto 2012). Most work has been done on text understanding and translation of feelings, yet critical exploration has been done on Twitter opinion examination of items, audits, breaking news, position recognition, and so on. The application of this research is that it can be applicable in Sarsov2, Covid-19, Test.csv, China virus, mental health (Kaila et al. 2016).

There are about 85 articles in IEEE xplore and in 130 Scopus related to this study. The most cited article was (Melamud, Goldberger, and Dagan 2016) performed on Twitter tweets utilizing feeling arrangement techniques with assessment vocabulary word references. Artificial Intelligence (Machine Learning) calculations to break down and imagine the impact of Covid all throughout the planet on dataset tweets that were crept by means of Twitter API and Google Tags. The Naive Bayes classifier was utilized as a base classifier and the five notable order procedures are proposed against the current gathering strategy, which coordinates five individual calculations into an ensemble-based grouping dynamic system, in particular Support Vector Machine (SVM) classifier, MaxEntropy classifier, Decision Tree classifier, Bagging classifier, and Random Forest classifier (Dasgupta and Gupta 2003). The grouping of the gathering considers the arrangement results of the five classifiers and the utilization of the procedure Majority-Vote to decide the last objective of the class of assessment (Kowsari et al. 2019). The applications of this research study additionally addresses a point by point exploratory investigation of (Kowsari et al. 2017) the five unique techniques for grouping feelings and basic comments.

Our institution is passionate about high quality evidence based research and has excelled in various domains (Vickram et al. 2022; Bharathiraja et al. 2022; Kale et al. 2022; Sumathy et al. 2022; Thanigaivel et al. 2022; Ram et al. 2022; Jothi et al. 2022; Anupong et al. 2022; Yaashikaa, Keerthana Devi, and Senthil Kumar 2022; Palanisamy et al. 2022).Drawback of the existing system is that it cannot detect the spam and non spam tweet messages. Understanding audience segments from social media and auto tagging the customer queries can be done in this research study (Gusenbauer 2008). The aim of this research is to enhance accuracy in text classification on coronavirus tweets using Novel Random Forest Classifier compared with Decision Tree.

2. Materials and Methods

This work is carried out in the Data Analytics Lab, Department of Information Technology at Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences. The study consists of two sample groups i.e Novel Random Forest and Decision Tree. Each group consists of 10 sample sizes i,e Novel Random Forest Classifier (Group 1=10) and Decision Tree (Group 2=10) with pretest power of 0.18. The sample size kept the threshold at 0.05, G power of 80%, confidence interval at 95%, and enrolment ratio as 1.

Data preparation

To perform Text Classification on Coronavirus Tweets the real time data sets used are detection of spam and non spam tweet messages. The input datasets for the proposed work is corona_NLP_test1.csv collected from kaggle.com (https://www.kaggle.com/vanshjatana). The dataset contains 7 columns and 1264 instances. The dataset was split into training and testing parts accordingly using the test size of 0.2.

Novel Random Forest

Arbitrary Forest created by Leo Breiman is a gathering of un-pruned order or relapse trees produced using the arbitrary determination of tests of the preparation information. Arbitrary provisions are chosen in the acceptance cycle. Expectation is made by collecting (larger part vote in favor of arrangement or averaging for relapse) the forecasts of the group. Each tree is developed as portrayed. By Sampling N haphazardly, If the quantity of cases in the preparation set is N yet with

substitution, from the first information (Liu 2012). The pseudocode for Random Forest is shown in Table 1.

For M number of information factors, the variable m is chosen with the machine learning end goal that m. Irregular Forest by and large displays a huge execution improvement when contrasted with single tree classifiers like C4.5. The speculation blunder rate that it yields thinks about well to Adaboost, anyway it is more strong to commotion.

Decision Tree

Choice Trees epitomize a regulated characterization approach. The thought came from the conventional tree structure which consists of a root and hubs (the positions where spots branches isolate), branches and leaves. Along these lines, a Decision Tree is developed from hubs which address circles and the branches are addressed by the sections that associate the hubs. A Choice Tree begins from the root, moves descending and by and large are attracted from left to right. The hub from where the tree begins is known as a root hub. The hub where the chain closes is known as the "leaf" hub. At least two branches can be reached out from each interior hub for example a hub that isn't leaf hub. A hub addresses a certain trademark while the branches address a scope of values. These scopes of qualities go about as a segment focuses for the arrangement of upsides of the given trademark. Figure 1 depicts the design of a tree.

Different strides for twitter opinion examination were inspected by Jain and Dandannavar utilizing AI calculations. A definite methodology has been accommodated in the feeling examination. First and foremost, information is gathered and afterward pre-handling is finished with the assistance of calculations enlivened from NLP. Then, at that point, unsheathing of components is finished. Opinion examination is finished by applying multinomial gullible bayes and choice tree models in the proposed technique. After assessments, best outcomes were acquired by that of the choice tree, showing 100% exactness, review, accuracy and F1 score.

At the point when the choice tree is utilized for text characterization it comprises a tree where interior hubs are marked by term, branches address weight and leaf address the class. Tree can characterize the record by going through the question structure from root until it arrives at a specific leaf, which addresses the objective for the arrangement of the report. The majority of preparing information will not fit in memory, choice tree development becomes wasteful due to trading of preparing tuples". The pseudocode for Random Forest is shown in Table 2.

1. Benefits: Decision trees competent to learn disjunctive articulations and their power to uproarious information appear to be helpful for report characterization.
2. Drawbacks: learning of choice tree calculations can't ensure to return the all around the world ideal choice tree.

STATISTICAL ANALYSIS

Statistical Package for the social sciences version 26 software tool was used for statistical analysis. An independent sample T-test was conducted for accuracy. Standard deviation, standard mean errors were also calculated using the SPSS Software tool. The significance values of proposed and existing algorithms contain group statistical values of proposed and existing algorithms. The independent variable which is represented as an independent variable with the intensity of tweets per minute represented by the dependent variable. Dependent on time parameters and the exponential by the NHPP.

3. Results

The group statistical analysis on the two gatherings shows Novel Random Forest has more mean accuracy than Decision Tree and the standard error mean is somewhat not exactly Novel Random Forest. Addresses the bar outline of correctnesses with standard deviation mistakes are plotted for both the calculations. The Novel Random Forest calculation scored an exactness of (81.60%) as displayed in table 5 and Decision Tree has scored (80.14%) as displayed in table 5. The correctnesses are recorded by testing the calculations with 10 distinctive example sizes and the normal precision is determined for every calculation Table 3 and Table 4. Table 5 shows the group statistic analysis, representing Novel Random Forest and Decision Tree.

4. Discussion

Text mining offers an intriguing blend of text order calculations. From them: Decision Tree, Support Vector Machine, K-Nearest Neighbors, Naïve Bayes also, covered up Markov model are the most fundamental five order calculations. And do a relative report for these five text characterization calculations with practically every one of the corrections which were finished on these calculations. We have depicted every calculation independently and concentrated on the adjustments made to something similar. These enhancements are characterized by Student (alter`ation), the fundamental calculation (change and expansion), and elements (extraction and decrease).

This review showed that the least demanding approach to progress the characterization is by utilizing highlight decrease (Surya and Senthilselvi 2022) which causes (I) fasting the characterization adjacent to of (ii) expanding the proficiency (Manning and Schutze 1999). Another explanation which is the adjustment of a few calculations is extremely challenging to reach (Ghosh and Meeden 1997). Additionally, this review showed that alteration of students is direct and can help for expanding the precision of calculation (Al-Turjman 2021). We can see every scientist has their own dataset for testing the improvement which makes the correlation more troublesome.

The limitation in this model is that the accuracy of NRF may get affected due to the inconsistent data and difficulty in getting the right datasets for analysis. Most of the data is simulated from nature which is far from reality. As with any technology, there are potential drawbacks to using facial recognition, such as threats to privacy, violations of rights and personal freedoms, potential data theft and other crimes. There's also the risk of errors due to flaws in the technology. The Future work can be concentrated on effective data processing techniques and usage of ensemble machine learning algorithms can be focussed. Table 6 shows independent 'T' sample tests for algorithms. The comparative accuracy analysis, mean of loss between two algorithms are specified.

5. Conclusion

Based on the experimental results, the Novel Random Forest has been proved to Coronavirus Tweets more significantly than Decision Tree. It can be used in Text Classification on coronavirus tweets. This review investigated and worked on the grouping of intestinal sickness vector information, a few works have been proposed in audits by analysts utilizing the presentation measurements displayed in Fig. 1 over, the outcomes have demonstrated that, dimensionality decrease model utilizing highlight extraction techniques, for example, PCA can assist with further developing grouping yield for example, Novel Random forest. It is fascinating to explore if late proposed work can further develop component extraction models and calculations.

DECLARATIONS

Conflicts of Interest

No conflicts of interest in this manuscript.

Authors Contribution

Author NR was involved in data collection, data analysis, data extraction, manuscript writing. Author MS was involved in conceptualization, data validation, and critical review of the manuscript.

Acknowledgement

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. Vee Eee Technologies Solution Pvt. Ltd., Chennai.
2. Saveetha University.
3. Saveetha Institute of Medical and Technical Sciences.
4. Saveetha School of Engineering.

6. References

Al-Turjman, Fadi. 2021. Artificial Intelligence and Machine Learning for COVID-19. Springer Nature.

Anupong, Wongchai, Lin Yi-Chia, Mukta Jagdish, Ravi Kumar, P. D. Selvam, R. Saravanakumar, and Dharmesh Dhabliya. 2022. "Hybrid Distributed Energy Sources Providing Climate Security to the Agriculture Environment and Enhancing the Yield." Sustainable Energy Technologies and Assessments. https://doi.org/10.1016/j.seta.2022.102142.

Bharathiraja, B., J. Jayamuthunagai, R. Sreejith, J. Iyyappan, and R. Praveenkumar. 2022. "Techno Economic Analysis of Malic Acid Production Using Crude Glycerol Derived from Waste Cooking Oil." Bioresource Technology 351 (May): 126956.

Dasgupta, Sanjoy, and Anupam Gupta. 2003. "An Elementary Proof of a Theorem of Johnson and Lindenstrauss." Random Structures and Algorithms. https://doi.org/10.1002/rsa.10073.

Ghosh, Malay, and Glen Meeden. 1997. Bayesian Methods for Finite Population Sampling. CRC Press.

Gusenbauer, Stefan. 2008. Relevance Feedback in Information Retrieval: A Comparison Including a Practical Evaluation of Several Approaches to Relevance Feedback in Information Retrieval. VDM Publishing.

Jothi, K. Jeeva, K. Jeeva Jothi, S. Balachandran, K. Mohanraj, N. Prakash, A. Subhasri, P. Santhana Gopala Krishnan, and K. Palanivelu. 2022. "Fabrications of Hybrid

Polyurethane-Pd Doped ZrO2 Smart Carriers for Self-Healing High Corrosion Protective Coatings." Environmental Research. https://doi.org/10.1016/j.envres.2022.113095.

Kaila, Vivek, Parminder Singh, K. N. Khatri, Atul D. Rane, and N. Prabhakar. 2016. "Conceptual Design of a Multirotor Aerial System Incorporating Secondary Rotors for Attitude Correction." 2016 International Conference on Unmanned Aircraft Systems (ICUAS). https://doi.org/10.1109/icuas.2016.7502515.

Kale, Vaibhav Namdev, J. Rajesh, T. Maiyalagan, Chang Woo Lee, and R. M. Gnanamuthu. 2022. "Fabrication of Ni–Mg–Ag Alloy Electrodeposited Material on the Aluminium Surface Using Anodizing Technique and Their Enhanced Corrosion Resistance for Engineering Application." Materials Chemistry and Physics. https://doi.org/10.1016/j.matchemphys.2022.125900.

Kowsari, Kamran, Donald E. Brown, Mojtaba Heidarysafa, Kiana Jafari Meimandi, Matthew S. Gerber, and Laura E. Barnes. 2017. "HDLTex: Hierarchical Deep Learning for Text Classification." 2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA). https://doi.org/10.1109/icmla.2017.0-134.

Kowsari, Kowsari, Jafari Meimandi, Heidarysafa, Mendu, Barnes, and Brown. 2019. "Text Classification Algorithms: A Survey." Information. https://doi.org/10.3390/info10040150.

Liu, Bing. 2012. Sentiment Analysis and Opinion Mining. Morgan & Claypool Publishers.

Manning, Christopher, and Hinrich Schutze. 1999. Foundations of Statistical Natural Language Processing. MIT Press.

Melamud, Oren, Jacob Goldberger, and Ido Dagan. 2016. "context2vec: Learning Generic Context Embedding with Bidirectional LSTM." Proceedings of The 20th SIGNLL Conference on Computational Natural Language Learning. https://doi.org/10.18653/v1/k16-1006.

Miyamoto, S. 2012. Fuzzy Sets in Information Retrieval and Cluster Analysis. Springer Science & Business Media.

Nay, John J. 2017. "Predicting and Understanding Law-Making with Word Vectors and an Ensemble Model." PloS One 12 (5): e0176999.

Palanisamy, Rajkumar, Diwakar Karuppiah, Subadevi Rengapillai, Mozaffar Abdollahifar, Gnanamuthu Ramasamy, Fu-Ming Wang, Wei-Ren Liu, Kumar Ponnuchamy, Joongpyo Shim, and Sivakumar Marimuthu. 2022. "A Reign of Bio-Mass Derived Carbon with the Synergy of Energy Storage and Biomedical Applications." Journal of Energy Storage. https://doi.org/10.1016/j.est.2022.104422.

Pelka, Obioma, Felix Nensa, and Christoph M. Friedrich. 2019. "Variations on Branding with Text Occurrence for Optimized Body Parts Classification." Conference Proceedings: ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference 2019 (July): 890–94.

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.

Sumathy, B., Anand Kumar, D. Sungeetha, Arshad Hashmi, Ankur Saxena, Piyush Kumar Shukla, and Stephen Jeswinde Nuagah. 2022. "Machine Learning Technique to Detect and Classify Mental Illness on Social Media Using Lexicon-Based Recommender System." Computational Intelligence and Neuroscience 2022 (February): 5906797.

Surya, V., and A. Senthilselvi. 2022. "Identification of Oil Authenticity and Adulteration Using Deep Long Short-Term Memory-Based Neural Network with Seagull Optimization Algorithm." Neural Computing and Applications. https://doi.org/10.1007/s00521-021-06829-3.

Thanigaivel, Sundaram, Sundaram Vickram, Nibedita Dey, Govindarajan Gulothungan, Ramasamy Subbaiya, Muthusamy Govarthanan, Natchimuthu Karmegam, and Woong Kim. 2022. "The Urge of Algal Biomass-Based Fuels for Environmental Sustainability against a Steady Tide of Biofuel Conflict Analysis: Is Third-Generation Algal Biorefinery a Boon?" Fuel. https://doi.org/10.1016/j.fuel.2022.123494.

Vickram, Sundaram, Karunakaran Rohini, Krishnan Anbarasu, Nibedita Dey, Palanivelu Jeyanthi, Sundaram Thanigaivel, Praveen Kumar Issac, and Jesu Arockiaraj. 2022. "Semenogelin, a Coagulum Macromolecule Monitoring Factor Involved in the First Step of Fertilization: A Prospective Review." International Journal of Biological Macromolecules 209 (Pt A): 951–62.

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.

Al-Turjman, Fadi. 2021. Artificial Intelligence and Machine Learning for COVID-19. Springer Nature.

A New Framework of Optimizing Keyword Weights in Text Categorization and Record Querying. 2008.

Dasgupta, Sanjoy, and Anupam Gupta. 2003. "An Elementary Proof of a Theorem of Johnson and Lindenstrauss." Random Structures and Algorithms. https://doi.org/10.1002/rsa.10073.

Ghosh, Malay, and Glen Meeden. 1997. Bayesian Methods for Finite Population Sampling. CRC Press.

Gupta, Vishal, and Gurpreet S. Lehal. 2009. "A Survey of Text Mining Techniques and Applications." Journal of Emerging Technologies in Web Intelligence. https://doi.org/10.4304/jetwi.1.1.60-76.

Gusenbauer, Stefan. 2008. Relevance Feedback in Information Retrieval: A Comparison Including a Practical Evaluation of Several Approaches to Relevance Feedback in Information Retrieval. VDM Publishing.

Heisei 2-Nen Kokusei Chōsa Hōkoku: 1990 Population Census of Japan. Volume 3, Results of the Second Basic Complete Tabulation. Dai 2-Ji Kihon Shūkei Kekka. Dai 3-Kan. 1992.

Horita, Luiz Ricardo Takeshi. n.d. "Visual Urban Road Features Detection Using Convolutional Neural Network with Application on Vehicle Localization."

https://doi.org/10.11606/d.18.2018.tde-10122018-152247.

Liu, Bing. 2012. Sentiment Analysis and Opinion Mining. Morgan & Claypool Publishers.

Manning, Christopher, and Hinrich Schütze. 1999. Foundations of Statistical Natural Language Processing. MIT Press.

Melamud, Oren, Jacob Goldberger, and Ido Dagan. 2016. "context2vec: Learning Generic Context Embedding with Bidirectional LSTM." Proceedings of The 20th SIGNLL Conference on Computational Natural Language Learning. https://doi.org/10.18653/v1/k16-1006.

Miyamoto, S. 2012. Fuzzy Sets in Information Retrieval and Cluster Analysis. Springer Science & Business Media.

Nay, John J. 2017. "Predicting and Understanding Law-Making with Word Vectors and an Ensemble Model." PloS One 12 (5): e0176999.

Pelka, Obioma, Felix Nensa, and Christoph M. Friedrich. 2019. "Variations on Branding with Text Occurrence for Optimized Body Parts Classification." Conference Proceedings: ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference 2019 (July): 890–94.

Surya, V., and A. Senthilselvi. 2022. "Identification of Oil Authenticity and Adulteration Using Deep Long Short-Term Memory-Based Neural Network with Seagull Optimization Algorithm." Neural Computing and Applications. https://doi.org/10.1007/s00521-021-06829-3.

TABLES AND FIGURES

Table 1. Pseudocode for Novel Random Forest

// I : Input dataset records
1. Import the required packages.
2. Convert the audio files into numerical values after the extraction feature.
3. Assign the data to X_train, y_train, X_test and y_test variables.
4. Using train_test_split() function, pass the training and testing variables.
5. Give test_size and the random_state as parameters for splitting the data using SVM training.
6. Importing the SVClassifier from sklearn library.
7. Using SVClassifier, predict the output of the testing data.

8. Calculate the accuracy of the model.
OUTPUT //Accuracy

Table 2. Pseudocode for Decision Tree

// I : Input dataset records
1. Import the required packages.
2. Convert the audio files and truncate the input sequences so that they are all the same length for modeling.
3. Assign the data to X_train, y_train, X_test and y_test variables.
4. Using train_test_split() function, pass the training and testing variables.
5. Give test_size and the random_state as parameters for splitting the data.
6. Adding Decision Tree, Dense Layer to the model
7. Compiling the model using metrics as accuracy.
7. Evaluate the output using X_test and y_test function
8. Get the accuracy of the model.
OUTPUT //Accuracy

Table 3. Accuracy of Text Classification on Coronavirus Tweets using Novel Random Forest for 10 samples out of 30 (Accuracy= 92.28%)

Test size	**Accuracy**
Test 1	74.27
Test 2	76.43
Test 3	78.38
Test 4	80.12
Test 5	82.73
Test 6	84.62
Test 7	87.65
Test 8	89.71
Test 9	90.31
Test 10	92.28

Table 4. Accuracy of Text Classification on Coronavirus Tweets using Decision Tree for 10 samples out of 30 samples (Accuracy= 82.36%)

Test size	**Accuracy**

Test 1	70.16
Test 2	72.32
Test 3	74.35
Test 4	76.96
Test 5	78.21
Test 6	79.62
Test 7	80.22
Test 8	80.87
Test 9	81.64
Test 10	82.36

Table 5. Group Statistic analysis, representing Novel Random Forest (mean accuracy 92.28%, standard deviation 6.45617) and Recurrent Neural Network (mean accuracy 82.36%, standard deviation 6.62072)

Algorithm	**N**	**Mean**	**Std. Deviation**	**Std.Error Mean**
Accuracy NRF	10	81.6000	6.45617	2.04162
Accuracy DT	10	80.1450	6.62072	2.09366

Table 6. Independent Sample Tests results with confidence interval as 95% and level of significance as 0.05 (Support Vector Machine appears to perform significantly better than Recurrent Neural Network with the value of p=0.18).

Accuracy	**Levene's Test for Equality of Variances**	**T-test for Equality of Means**

	F	Sig.	t	df	Sig	Mean Difference	Std. Error Difference	95% Conf. Interval Lower	95% Conf. Interval Upper
Equal Variances assumed	.136	.717	498	18	.625	1.45500	2.92431	-4.68875	7.59875
Equal Variances not assumed	-	-	498	17.98	.625	1.45500	2.92431	-4.68903	7.59903

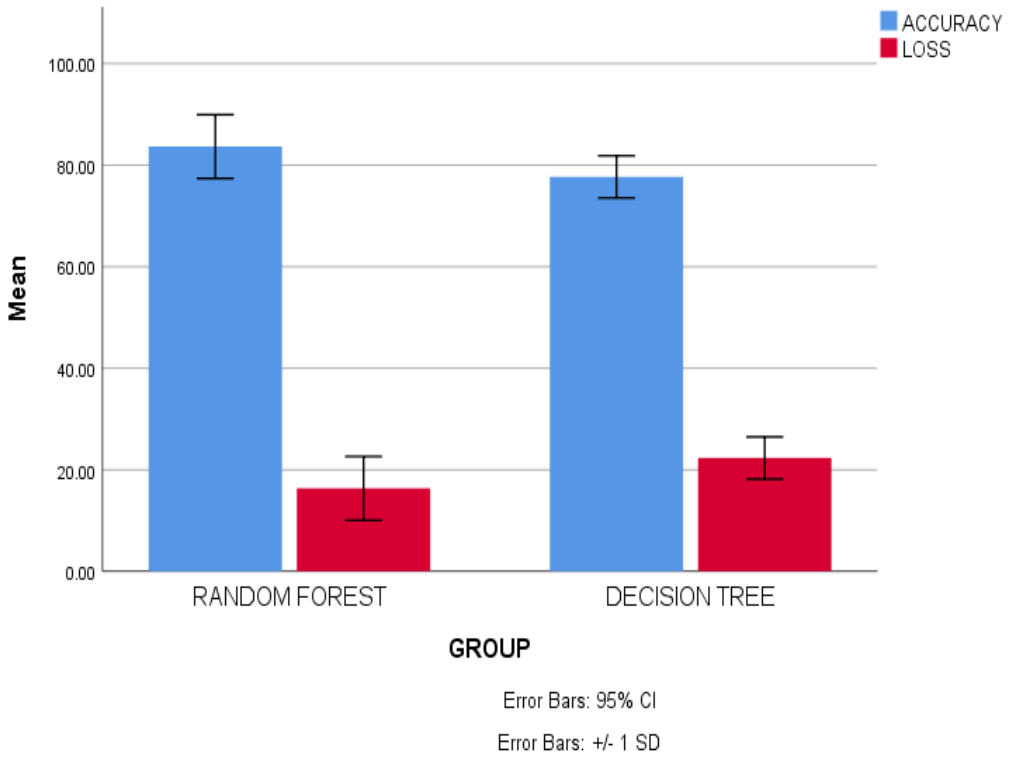

Fig. 1. Comparison of Novel Random Forest and Decision Tree in terms of accuracy. The mean accuracy of Random Forest is greater than Decision Tree and standard deviation is also slightly higher than Decision. X-axis: Novel Random Forest vs Decision Tree. Y-axis: Mean accuracy of detection ±1 SD.