Comparing and Improving the Novel Random Forest Algorithm to Logistic Regression Classifier for the Analysis and Accuracy of Fake Notes

Renangi.Supriya[1], S.Kalaiarasi[2*]

[1]Research Scholar,Saveetha School of Engineering,Saveetha Institute of Medical And Technical Sciences, Saveetha University, Chennai, Tamil Nadu , India, Pincode: 602105

[2*]Project Guide, Corresponding Author, Saveetha School of Engineering, Saveetha Institute of Medical And Technical Sciences, Saveetha University, Chennai, Tamilnadu, India, Pincode: 602105

ABSTRACT

Aim: To enhance the accuracy in classifying fake currency detection using Novel Random Forest Algorithm and Logistic Regression Classifier. **Materials and Methods:** This study contains 2 groups such as Novel Random Forest Algorithm and Logistic Regression Classifier. Each group consists of a sample size of 10 and the study parameters include alpha value 0.05, beta value 0.2. The SPSS was used for predicting significance value of the dataset considering G-Power value 80%. **Results and Discussions:** The Novel Random Forest Algorithm is 65.55% more accurate than the Logistic Regression Classifier of 35.5% in classifying the fake currency notes with significance value p=1.000. **Conclusion:** The Random forest algorithm is significantly better than the Logistic regression classifier in identifying fake notes. It can be also considered as a better option for the classification model of fake currency.

Keywords: Fake Currency Detection, Classification Model, Machine Learning, Logistic Regression Classifier, Novel Random Forest Algorithm, Image Processing, counterfeit.

INTRODUCTION

Counterfeiting is the intentional reproduction of genuine currency. It is done to trick the receiver into thinking the entity received is genuine. The spread of counterfeit currency is posing a serious threat to the global economy, affecting each country severely. (Greifeneder et al. 2020) According to the National Crime Records Bureau NCRB, India, counterfeiting is one of the Indian Panel Code economic crimes NCRB Compendium 2015-16. (Chhibbar 2020). The rise in counterfeit currency is the result of technological advancement and a variety of other factors. One of the innocuous effects of counterfeit money is that it has a direct impact on the value of real money causing a decrease in value. Because of the rapid acquisition of technology, the rate of counterfeiting is rapidly increasing. The reason for the rapid adoption is the low cost, wide availability, and efficiency of technological equipment (Shiny 2021). For many years, there has been a race between counterfeiters and banks.To address this serious issue, various researchers came up with a variety of techniques and proposed solutions from the fields of machine learning and image processing. People in today's new digitalization environment are constrained by technology, which is rapidly evolving. Of course, such inventions make our lives

a lot easier. People can now complete tasks with minimal effort thanks to technological advancements. However, some people are abusing the benefits of such technologies to achieve their nefarious goals. There are numerous examples of this type all around us. One of the most obvious examples is a counterfeit note. Counterfeit currency is defined as currency produced without the government's (OECD 2008) legal approval. Dishonest people use cutting-edge scanning and printing techniques to produce counterfeit money. The creation of such counterfeit money has an impact on the economy of any country. Indian currency is also a component.

This study is referenced in approximately 34 articles in IEEE Xplore and 10 in Scopus. The use of sequential data information results in a Random forest algorithm. Deep Learning-Based Currency Detection and Image Recognition Classification models based on an examination of fake currency detection techniques. Various researchers in the fields of machine learning and image processing have proposed numerous currency detection techniques. (Shih 2010). Techniques for machine learning: It can be classified as either supervised or unsupervised learning, or as a combination of the two. It is entirely dependent on the type of data collection to determine which technique should be used to achieve good results. Supervised machine learning techniques (Jo 2021) are those in which the value of an output variable is predicted using an input variable. It is classified as supervised learning because it uses both a training and a testing dataset.These techniques are primarily (Saini, Lata, and Sinha 2022) When the problem has an output variable with a real value, regression is used, whereas classification model (Shu and Liu 2019) is used when the output variable has a category. With the advancement of modern banking services, automatic methods for detecting paper currency have become important in most applications, such as automated teller machines and automatic goods seller machines. Images are processed using various image processing techniques, and various features are extracted from the images. (National Research Council et al. 1995) Many applications, such as automatic selling goods and vending machines, require automatic methods of banknote recognition.

(Bhavikatti et al. 2021; Karobari et al. 2021; Shanmugam et al. 2021; Sawant et al. 2021; Muthukrishnan 2021; Preethi et al. 2021; Karthigadevi et al. 2021; Bhanu Teja et al. 2021; Veerasimman et al. 2021; Baskar et al. 2021)

Some datasets are intended for theoretical research rather than real-world application processing. The disadvantage is that distinguishing between fake and original notes is extremely difficult. Researchers developed their own feature set because most existing standard feature extraction processes are designed for short-term analysis. The goal is to improve the detection of counterfeit currencies.

MATERIALS AND METHODS

This work is Carried out in the Data Analytics lab, Department of Computer Science at Saveetha School of Engineering. This Study consists of two groups i.e., Novel random Forest and Logistic Regression Classifier. Each group consists of 10 samples with a test size beta value of 0.2. The sample size kept the threshold at Alpha value of 0.05, G power

of 80%, confidence interval at 95%, and enrolment ratio as 1.

The data collection is taken from the open source access website IEEE-dataport.org that is used for software effort estimation using Novel Random Forest Algorithm and Logistic Regression Classifier. The open access dataset consists of 108 rows and 10 columns. The Jupyter software with windows 10.1 system has been used to develop this software effort estimation. The proposed system uses two groups Novel Random Forest Algorithm and Logistic Regression Classifier technique (Shrivastava, Mahalingam, and Dutta 2017) where this algorithms are fitted into the dataset which is then tested and trained for the process of estimating the software effort where the cost estimation and the time estimation is known.

Novel Random Forest Algorithm

Random Forest Algorithm is a well-known machine learning algorithm from the supervised learning technique. It can be applied to both classification and regression problems in machine learning . It is based on the concept of ensemble learning, which is the process of combining multiple classifiers to solve a complex problem and improve the model's performance. Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset," as the name implies. Instead of relying on a single decision tree, the random forest takes each tree's prediction and, based on the majority vote of predictions.

Pseudocode for Novel Random Forest Algorithm

Input: Training dataset K.

Output: A class of testing dataset

Step 1: Choose K data points at random from the training set.

Step 2: Create decision trees for the data points you've chosen (Subsets).

Step 3: Choose a N for the number of decision trees you want to make.

Step 4: Repeat of Steps 1 and 2.

Step 5: Find the forecasts of each decision tree for new data points, and allocate the new data points to the category with the most votes.

Logistic Regression Classifier

A fundamental classification technique is logistic regression . It belongs to the linear classifier family and is related to polynomial and linear regression. Logistic regression is quick and easy to understand, and the results are easy to interpret. Although it is primarily a binary classification method, it can also be applied to multiclass problems.

Pseudocode for Logistic Regression Classifier

Input: Training dataset K.

Output: A class of testing dataset.

Step 1: To train the model, first fit the data to the Logistic Regression model.

Step 2: Let's simulate a single banknote prediction.

Step 3: All we have to do now is extract the features, scale them, and integrate them into our model that has already been trained.

Step 4: The chances of a banknote belonging to each target class.

The minimum requirements to run the softwares used here are Intel Core i3 7th generation CPU with 12GB of RAM, an x86-based processor, a 64-bit operating system, and a hard drive. Currently, the software runs on Windows 10 and is programmed in Python.

The data collection is taken from the open source access website IEEE-dataport.org that is used for software effort estimation using Novel Random Forest and Logistic Regression Classifier technique. The dataset consists of 108 rows and 10 columns. The Jupyter software with windows 10.1 system has been used to develop this software effort estimation. The proposed system uses two groups Novel Random Forest and Logistic Regression Classifier technique where this algorithms are fitted into the dataset which is then tested and trained for the process of estimating the software effort where the cost estimation and the time estimation is known, the sample size of the dataset is 74.

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. In fake currency detection parameters are independent variables and counterfeit prediction is dependent variable. Dependent variable Class Independent variable Variance, skewness, kurtosis.

RESULTS

Table 1 shows the comparison of accuracy of Random Forest Algorithm and Logistic Regression Classifier. Table 2 shows the Group Statistics Results with RFA having mean accuracy of 65.50% and std.deviation of 3.028, and the LRC with mean accuracy of 30.50% and std.deviation of 3.028.

Table 3 shows the Independent sample t-test of the significance level 0.001 (2-tailed) for Random Forest and Logistic Regression Classifier. The accuracies are recorded by testing the algorithms with different sample sizes and the average accuracy is calculated for each algorithm.

Fig. 1 represents the bar chart representing the mean accuracies and standard deviation error for Random Forest Algorithm and Logistic Regression Classifier. The group statistical analysis on the two groups shows Random Forest Algorithm has more mean accuracy than Logistic Regression Classifier, and the standard error mean is slightly less than Random Forest Algorithm.

DISCUSSION

In this study, detecting fake currency using the RF algorithm has significantly higher accuracy of 65% in comparison to LRC with 35% accuracy. RFA appears to produce more consistent results with minimal standard deviation.

The similar findings of the paper (Madhusudan 2020) had an accuracy of 65% with RFA which was used to detect the currency. The proposed work of (Bosco and Khan 2018) reported RFA has 65% accuracy which is used to predict the accuracy of fake currency notes and performance of fake notes ((Noel-Tod 2015). The work proposed by Rathore 2020 shows the RFA has a better accuracy of 65%. LRC is a parameter to measure fake currency which is used in both traditional and modern methods as per their research it opposes DTC (*Payment and Settlement Systems in Selected Countries* 2003) has highest accuracy and LRC will get least accuracy compared to other machine learning techniques which ranges between 60% when compared to other machine learning algorithms will get more accuracy than this. By using RFA

for forecasting fake currency it will have key issues to pretend (Mehta, Pandya, and Kotecha 2021) in this paper shows RFA has the least accuracy of 65%. Increasing the dataset's value only tends to get desired accuracy. RFA performs better with a combination of other machine learning algorithms (Guru et al. 2017).

The limitation of this research is that it cannot give appropriate results for smaller data. In this model it is not able to consider all given feature variable parameters for training. The future scope of proposed work will be prediction of fake notes based on classification model using class labels for lesser time complexity (Cristianini, Shawe-Taylor, and Department of Computer Science Royal Holloway John Shawe-Taylor 2000). Our future work will focus on faster and more accurate fake currency detection through the use of advanced image processing techniques.

CONCLUSION

The implementation of Random Forest Algorithm and Logistic Regression Classifier were completed, and its results were statistically compared for the detection of the fake currency notes. The accuracy value of the Random Forest Algorithm is 65% whereas the accuracy value of the Logistic Regression Classifier is 35%. The quality of detecting fake notes accuracy using Random Forest Algorithm appears to be better than Logistic Regression Classifier.

DECLARATION

Conflict of Interests

No conflict of interests in this manuscript

Authors Contribution

Author RS as involved in data collection, data analysis, manuscript writing. Author SKA was involved in conceptualization, data validation, and critical review of 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: Thanks to the following organizations for providing financial support that enabled us to complete the study.

1. HKM Info Systems Pvt.Ltd.
2. Saveetha University.
3. Saveetha Institute of Medical and Technical Sciences.
4. Saveetha School of Engineering

REFERENCES

1. Baskar, M., R. Renuka Devi, J. Ramkumar, P. Kalyanasundaram, M. Suchithra, and B. Amutha. 2021. “Region Centric Minutiae Propagation Measure Orient Forgery Detection with Finger Print Analysis in Health Care Systems.” *Neural Processing Letters*, January. https://doi.org/10.1007/s11063-020-10407-4.
2. Bhanu Teja, N., Yuvarajan Devarajan, Ruby Mishra, S. Sivasaravanan, and D. Thanikaivel Murugan. 2021. “Detailed Analysis on Sterculia Foetida Kernel Oil as Renewable Fuel in Compression Ignition Engine.” *Biomass Conversion and Biorefinery*, February. https://doi.org/10.1007/s13399-021-01328-w.
3. Bhavikatti, Shaeesta Khaleelahmed, Mohmed Isaqali Karobari, Siti Lailatul Akmar Zainuddin, Anand Marya, Sameer J. Nadaf, Vijay J. Sawant,

Sandeep B. Patil, Adith Venugopal, Pietro Messina, and Giuseppe Alessandro Scardina. 2021. "Investigating the Antioxidant and Cytocompatibility of Mimusops Elengi Linn Extract over Human Gingival Fibroblast Cells." *International Journal of Environmental Research and Public Health* 18 (13). https://doi.org/10.3390/ijerph18137162.

4. Bosco, Joish, and Fateh Khan. 2018. *Stock Market Prediction and Efficiency Analysis Using Recurrent Neural Network*. GRIN Verlag.
5. Chhibbar, Ashish. 2020. *Navigating the Indian Cyberspace Maze: Guide for Policymakers*. K W Publishers Pvt Limited.
6. Cristianini, Nello, John Shawe-Taylor, and Department of Computer Science Royal Holloway John Shawe-Taylor. 2000. *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*. Cambridge University Press.
7. Greifeneder, Rainer, Mariela Jaffe, Eryn Newman, and Norbert Schwarz. 2020. *The Psychology of Fake News: Accepting, Sharing, and Correcting Misinformation*. Routledge.
8. Guru, D. S., T. Vasudev, H. K. Chethan, and Y. H. Sharath Kumar. 2017. *Proceedings of International Conference on Cognition and Recognition: ICCR 2016*. Springer.
9. Jo, Taeho. 2021. *Machine Learning Foundations: Supervised, Unsupervised, and Advanced Learning*. Springer Nature.
10. Karobari, Mohmed Isaqali, Syed Nahid Basheer, Fazlur Rahman Sayed, Sufiyan Shaikh, Muhammad Atif Saleem Agwan, Anand Marya, Pietro Messina, and Giuseppe Alessandro Scardina. 2021. "An In Vitro Stereomicroscopic Evaluation of Bioactivity between Neo MTA Plus, Pro Root MTA, BIODENTINE & Glass Ionomer Cement Using Dye Penetration Method." *Materials* 14 (12). https://doi.org/10.3390/ma14123159.
11. Karthigadevi, Guruviah, Sivasubramanian Manikandan, Natchimuthu Karmegam, Ramasamy Subbaiya, Sivasankaran Chozhavendhan, Balasubramani Ravindran, Soon Woong Chang, and Mukesh Kumar Awasthi. 2021. "Chemico-Nanotreatment Methods for the Removal of Persistent Organic Pollutants and Xenobiotics in Water - A Review." *Bioresource Technology* 324 (March): 124678.
12. Madhusudan, Desai Mitesh. 2020. "Stock Closing Price Prediction Using Machine Learning SVM Model." *International Journal for Research in Applied Science and Engineering Technology*. https://doi.org/10.22214/ijraset.2020.32154.
13. Mehta, Pooja, Sharnil Pandya, and Ketan Kotecha. 2021. "Harvesting Social Media Sentiment Analysis to Enhance Stock Market Prediction Using Deep Learning." *PeerJ. Computer Science* 7 (April): e476.
14. Muthukrishnan, Lakshmipathy. 2021. "Nanotechnology for Cleaner Leather Production: A Review." *Environmental Chemistry Letters* 19 (3): 2527–49.
15. National Research Council, Division on Engineering and Physical Sciences, National Materials Advisory Board, Commission on Engineering and

Technical Systems, and Committee on Currency Features Usable by the Visually Impaired. 1995. *Currency Features for Visually Impaired People*. National Academies Press.

16. Noel-Tod, Jeremy. 2015. "Mass Illuminations: Jennings, Madge, Rimbaud and the 'popular' Prose Poem." *Critical Quarterly*. https://doi.org/10.1111/criq.12209.
17. OECD. 2008. *The Economic Impact of Counterfeiting and Piracy*. OECD Publishing.
18. *Payment and Settlement Systems in Selected Countries*. 2003.
19. Preethi, K. Auxzilia, K. Auxzilia Preethi, Ganesh Lakshmanan, and Durairaj Sekar. 2021. "Antagomir Technology in the Treatment of Different Types of Cancer." *Epigenomics*. https://doi.org/10.2217/epi-2020-0439.
20. Saini, Sandeep, Kusum Lata, and G. R. Sinha. 2022. *VLSI and Hardware Implementations Using Modern Machine Learning Methods*. CRC Press.
21. Sawant, Kashmira, Ajinkya M. Pawar, Kulvinder Singh Banga, Ricardo Machado, Mohmed Isaqali Karobari, Anand Marya, Pietro Messina, and Giuseppe Alessandro Scardina. 2021. "Dentinal Microcracks after Root Canal Instrumentation Using Instruments Manufactured with Different NiTi Alloys and the SAF System: A Systematic Review." *NATO Advanced Science Institutes Series E: Applied Sciences* 11 (11): 4984.
22. Shanmugam, Vigneshwaran, Rhoda Afriyie Mensah, Michael Försth, Gabriel Sas, Ágoston Restás, Cyrus Addy, Qiang Xu, et al. 2021. "Circular Economy in Biocomposite Development: State-of-the-Art, Challenges and Emerging Trends." *Composites Part C: Open Access* 5 (July): 100138.
23. Shiby, Ashik. 2021. "Detection of Fake Currency Using Image Processing." *International Journal for Research in Applied Science and Engineering Technology*. https://doi.org/10.22214/ijraset.2021.36348.
24. Shih, Frank Y. 2010. *Image Processing and Pattern Recognition: Fundamentals and Techniques*. John Wiley & Sons.
25. Shrivastava, Rahul, Hari Mahalingam, and N. N. Dutta. 2017. "Application and Evaluation of Random Forest Classifier Technique for Fault Detection in Bioreactor Operation." *Chemical Engineering Communications*. https://doi.org/10.1080/00986445.2017.1292259.
26. Shu, Kai, and Huan Liu. 2019. *Detecting Fake News on Social Media*. Morgan & Claypool Publishers.
27. Veerasimman, Arumugaprabu, Vigneshwaran Shanmugam, Sundarakannan Rajendran, Deepak Joel Johnson, Ajith Subbiah, John Koilpichai, and Uthayakumar Marimuthu. 2021. "Thermal Properties of Natural Fiber Sisal Based Hybrid Composites – A Brief Review." *Journal of Natural Fibers*, January, 1–11.

TABLES AND FIGURES

Table 1. Comparison of accuracy values of Random Forest Algorithm and Logistic Regression Classifier in various iterations

Iterations	RFA	LRC
1	65	35
2	64	34
3	63	33
4	62	32
5	61	31
6	60	30
7	59	29
8	58	28
9	57	27
10	56	26

Table 2. Group Statistics Results- RFA has an mean accuracy (65.50%), std.deviation (3.028), whereas for LRC has mean accuracy (30.50%), std.deviation (3.028).

Group Statistics					
	Groups	**Sample (N)**	**Mean**	**Std deviation**	**Std. Error Mean**
Accuracy	RFA	10	65.50	3.028	0.957
	LRC	10	30.50	3.028	0.957

Table 3. The Independent sample t-test of the significance level 0.001 (2-tailed) for RFA and LRC.

Accuracy	**Independent Samples Test**							
	Levene's Test for Equality of Variances					**T-test for Equality of Means**		
	F	**Sig**	**t**	**df**	**Sig(2-tailed)**	**Mean Difference**	**Std.Error Difference**	**95% Confidence Interval of the Difference**

								Lower	Upper
Equal variances assumed	0.000	1.000	25.849	18.000	0.001	35.000	1.354	32.155	37.845
Equal variances not assumed			25.849	18.000	0.001	35.000	1.354	32.155	37.845

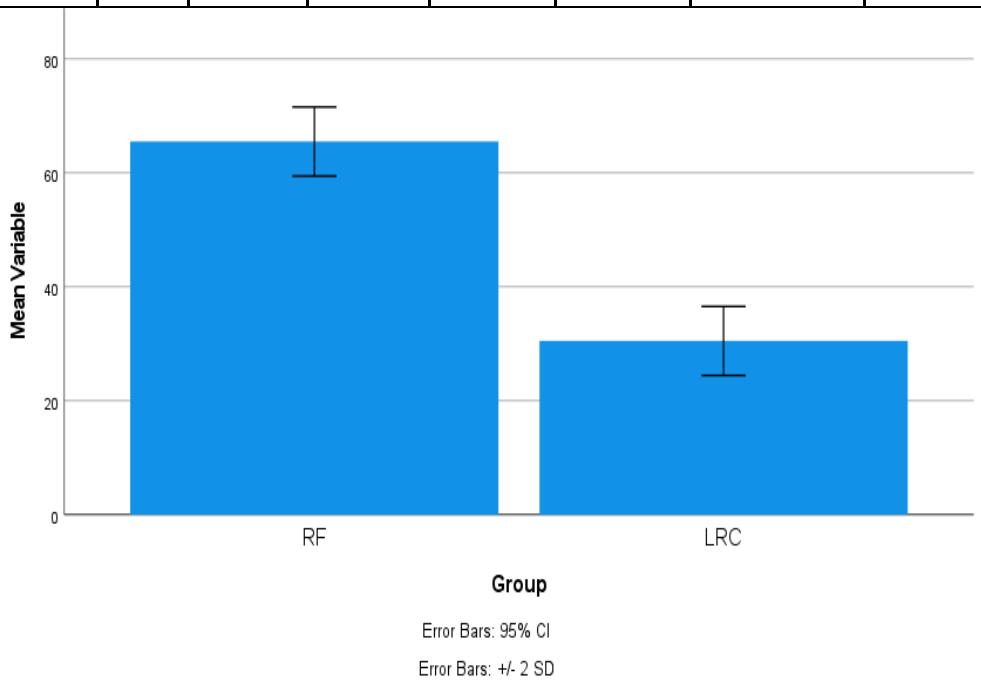

Fig. 1 Bar Graph showing comparison on mean accuracy of RFA (65%) and the LRC (35%). X-axis: RFA vs LRC and Y-axis: Mean Accuracy with $\pm$2 SD.