Original Article

Improving Accuracy in Malware Detection for Health Sensor Data by Novel Ada Boost M1Algorithm over Convolutional Neural Networks-Long Short-Term Memory Networks

Jyothi Thilak Kumar[1], P S.Uma Priyadarsini[2]

[1]Research Scholar, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamil Nadu, India, Pincode: 602105
[2]Project Guide, Corresponding Author, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamil Nadu, India, Pincode: 602105

Abstract

Aim: The Aim of research is to increase the accuracy of prediction the Malware detection for Health Sensor Data using the Convolutional Neural Networks-Long Short-Term(CNN-LSTM) in Comparison with Novel AdaBoostM1.
Materials and Methods: In the Convolutional Neural Networks-Long Short-Term algorithm, the sample size is n=10, while in the Novel AdaBoostM1Algorithm, the sample size is n=10 and the g-power value of 80% and datasets are collected from various web sources with recent research findings and threshold 0.05%, confidence interval 95% mean and standard deviation.
Results and Discussions: Convolutional Neural Networks-Long Short-Term algorithm provides mean accuracy of 79.6%. when compared to the Novel AdaBoostM1algorithm with a mean accuracy of 97.8%. Statistical insignificant difference was observed between Novel Novel AdaBoostM1 and Convolutional Neural Networks-Long Short-Term, $p = 0.893$ ($p>0.05$).
Conclusion: Novel AdaBoostM1achieved significantly better Accuracy than CNN-LSTM Algorithm in Malware Detection for Health sensor data comparison.

Keyword: Malware Detection ,Convolutional Neural Networks-Long Short-Term ,Novel AdaBoostM1, Health sensors, Machine Learning

https://doi.org/10.47750/pnr.2022.13.S04.197

INTRODUCTION

Malware Bloom grouping is utilized to characterize the Malware Blossom from its species like sesota, verginica, Versicolor.(Liu et al. 2020). It perceives the example as indicated by their sepal and petal lengths and widths. In fragrant healing, once in a while Malware Blossom oil is utilized as calming medication (Zhang and Ma 2012). Rhizomes of the Malwareare utilized for drugs, fragrances and are useful for infants getting teeth. Ideal bunching of information became troublesome through conventional methodologies when contrasted and neural organization grouping.(Jin et al. 2020) The machine can undoubtedly order the class of Malware Blossom while carrying out the current dataset of Malwarebloom for grouping.(Rajalakshmi and Anusha 2017). These days design acknowledgment and AI have been utilized in many fields. (Hu and Song 2015)design acknowledgment recognizes pictures, letters, voices, and different items. So this design acknowledgment has turned into a fundamental piece of this created innovation.(Ayu, Ayu, and Karyono 2014). It obviously depicts how AI functions in this design acknowledgment. (Kutlu, Kutlu, and Avcı 2019). Uses of this Malware Blossom are utilized in the water sanitization and fragrance industry. The application of malware detection serves as an early warning signal for your pc and informs you if you are on a secure platform.

In finding cost prediction in health care using novel ranking by comparing machine learning algorithms 560 journals from IEEE Xplore digital library, 280 articles from ScienceDirect, 1030 articles from google scholar and 798 articles are incorporated.A Multi-layer feed-forward network has different layers in it (Li and Wu 2015). Feedforward multi-facet neural organization is applied with Back Propagation preparing calculation then it becomes Backpropagation neural organization. (Zhu, n.d.). The backpropagation calculation gives the best precision with few mistakes.(Zhu, n.d.; Ayu, Ayu, and Karyono 2014) Multi-facet feed-forward neural

organizations give quicker and precise characterization for some example acknowledgment issues. Matlab orders are utilized as a hard copy of the MatLab code for reproduction of backpropagation neural organization for characterization of Malwarebloom dataset. By plotting the mistake versus the quantity of cycles execution was assessed for the created network. Neural organizations ordered the testing information with 100 percent acknowledgment.(Ayu, Ayu, and Karyono 2014) Gaussian-based order without tuning any boundaries accomplished rightness up to 96.67% and by applying nonlinear discriminant investigation to this grouping accomplished accuracy 98%. Assessment lists of review, accuracy, F-Measure, AUC, and Gini coefficients are taken to quantify the exhibition of the irregular timberland model and helping tree model for Malware Dataset characterization. Notwithstanding, the aftereffects of the arbitrary woods give somewhat preferable outcomes over those of helping tree models. With a more upgraded order there might be an opportunity to work on the presentation. The SVM order strategy is more successful than KNN, Logistic Regression techniques while contrasting the exactness and without cross-approval method for the Malwaredataset.(Mohanta and Saldanha 2020). To perceive Malware Species, administered learning of KNN calculation is utilized in Malware Blossom characterization and here some misclassified result is created because the expectation for class 1 is 4% off-base.

Our institution is passionate about high quality evidence based research and has excelled in various fields (Parakh et al. 2020; Pham et al. 2021; Perumal, Antony, and Muthuramalingam 2021; Sathiyamoorthi et al. 2021; Devarajan et al. 2021; Dhanraj and Rajeshkumar 2021; Uganya, Radhika, and Vijayaraj 2021; Tesfaye Jule et al. 2021; Nandhini, Ezhilarasan, and Rajeshkumar 2020; Kamath et al. 2020). For Pattern acknowledgment in Malwarebloom order utilizing AI calculations, many examination works are accessible (Vemparala, n.d.). Because of misclassification, a few works can't show great exact outcomes. (Ye et al. 2019)For this reason, an Innovative straight discriminant investigation calculation is proposed and contrasted with the Decision tree calculation. (Portegys 2010). This exploration plans to work on the exactness of the proposed model. The aim of the work is to detect suspicious users in the working environment.

MATERIALS AND METHODS

The proposed work is done in the Object-Oriented Analysis and Design Lab, Department of Computer Science and Engineering, Saveetha School of Engineering, SIMATS. In this review, In the proposed model, utilizing the Malware Blossom dataset which was downloaded from the Kaggle site https://www.kaggle.com/arshid/iris-bloom dataset. English analyst and researcher Ronald Fisher presented the Malwarerose dataset in his 1936 paper. The dataset has 150 columns and 5 ascribes. The dataset contains 50 examples from every species. In the Convolutional Neural Networks-Long Short-Term algorithm, the sample size is n=10, while in the Novel AdaBoostM1Algorithm, the sample size is n=10 and the g-power value of 80% and datasets are collected from various web sources with recent research findings and threshold 0.05%, confidence interval 95% mean and standard deviation.

There are three species in this dataset.(Alatabbi 2013) The example size was determined by utilizing clincalc.com by keeping G power (Saunders 2009) and the base force of the examination is fixed as 0.8 and the most extreme acknowledged mistake is fixed as 0.5 with an edge worth of 0.05% and Confidence Interval is 95%. They are Malware Sentosa, MalwareVirginica and MalwareVersicolor.Attributes utilized in this dataset are Petal Length, Petal Width, Sepal Length, Sepal Width, and Class (Species). Compute test size by keeping the G power. The base of force investigation and acknowledged blunder is fixed as 0.8 and 0.5 individually.

Novel AdaBoostM1

AdaBoost was initially known as AdaBoostM1 by the technique's creators, Freund and Schapire. It's been dubbed discrete AdaBoost recently because it's used for classification rather than regression.

Step 1: Assign Equal Weights to all the observations. Initially assign same weights to each record in the dataset
Step 2: Classify random samples using stumps.
Step 3: Calculate Total Error.
Step 4: Calculate Performance of the Stump.
Step 5: Update Weights.
Step 6: Update weights in iteration.
Step 7: Final Predictions.

CNN-LSTM

Although we will refer to LSTMs that employ a CNN as a front end as "CNN LSTM" in this course, this architecture was initially referred to as a Long-term Recurrent Convolutional Network or LRCN model. The duty of generating textual descriptions of photographs is handled by this architecture. The employment of a CNN that has been pre-trained on a difficult picture classification assignment and has been repurposed as a feature extractor for the caption producing challenge is crucial.
The hardware configuration includes an Intel i5 processor with a RAM size of 8GB. The system used was a 64-bit Windows operating system and Windows 11, Google Colab, and Microsoft Office are used for software specification.

Statistical Analysis
IBM SPSS version 23 was the statistical software used for analysis purposes. GetProcAddress, ExitProcess, WriteFile, GetLastError, CloseHandle, FreeLibrary, Sleep, GetStdHandle, MultiByteToWideChar, GetCurrentThreadId, FindClose are some of the independent variables used and malware(Christodorescu et al. 2007) is the dependent variable. The mean, standard deviation, and standard error mean statistical significance between the groups were determined using an independent sample T-Test.

RESULTS

To compare accuracy for the linear discriminant analysis and decision tree various sample datasets are collected with test size. Table 1 represents Group Statistics analysis for both algorithms based on Accuracy. The accuracy values are compared between the CNN-LSTM (79.6%) and Novel AdaBoostM1(97.0%) with standard deviation (2.89, 2.15) and standard error mean value (0.68, 0.91). It proves that linear discriminant analysis had the highest accuracy. It shows that Novel AdaBoostM1seems to have better accuracy than the CNN-LSTM.

Table 2 represents the independent sample test with f1 score, level of significance as 0.05, and 95% confidence interval differences for CNN-LSTM and AdaBoostM1. It found that Novel AdaBoostM1is insignificantly different from CNN-LSTM with p as 0.089 ($p>0.05$). In Fig. 1 Barcharts represents the comparison of mean accuracy and standard errors for CNN-LSTM and AdaBoostM1. Novel AdaBoostM1 is better than CNN-LSTM in terms of mean accuracy and standard deviation. From the graph, it can be seen that Novel AdaBoostM1performance is better than CNN-LSTM in Malware Detection for Health Sensor Data classification.

In Fig. 1 Bar charts represent the comparison of mean accuracy and standard errors for CNN-LSTM and AdaBoostM1. AdaBoostM1 is better than CNN-LSTM in terms of mean accuracy and standard deviation. From the graph, it can be seen that AdaBoostM1 performance is better than CNN-LSTM in Malware Detection for Health Sensor Data classification. Convolutional Neural Networks-Long Short-Term algorithm provides mean accuracy of 79.6%. when compared to the Novel AdaBoostM1algorithm with a mean accuracy of 97.8%.

DISCUSSIONS

In this research, the proposed model observed that the proposed model of Novel AdaBoostM1gives better accurate results in Malware Detection For Health Sensor Data classification than the existing model of the CNN-LSTM. Convolutional Neural Networks-Long Short-Term algorithm provides mean accuracy of 79.6%. when compared to the Novel AdaBoostM1algorithm with a mean accuracy of 97.8%. Statistical insignificant difference was observed between Novel Novel AdaBoostM1 and Convolutional Neural Networks-Long Short-Term, $p = 0.893$ ($p>0.05$).

We extricated memory helpers 1000 examples. (Vemparala, n.d.)The exploratory outcomes from include decrease utilizing mRMR and ANOVA are as displayed in. We got these results in the wake of arranging the examples utilizing SVM, AdaBoost, Random Woods, and J48. Five memory aide based models were built at a variable length, beginning from 40 to 120 at a time frame.(Vemparala, n.d.; Hu and Song 2015) Among these five models, ANOVA gives the best outcome with a solid positive probability proportion of 16.38 for the full length of 120 memory aides involving Novel AdaBoostM1J48 as base classifier. (Mohanta and Saldanha 2020)The principle benefits of this model are its low blunder rate and speed. Notwithstanding, mental aide based highlights can be without any problem changed utilizing code jumbling strategies.

We can picture the best result for the Novel AdaBoostM1classifier. (Jin et al. 2020) Nonetheless, this outcome is obtained with additional work i.e., including designing which is a basic undertaking in the AI pipeline. (Rajalakshmi and Anusha 2017). To dispense with the assignment of element designing, we make use of profound neural organization design, which is an assortment of layers, with each layer consisting of a few neurons(Ayu, Ayu, and Karyono 2014). In future enhancement various machine learning models are used to optimize the malware detection.

CONCLUSION

In this research, an efficient and accurate malware detection model was developed for detecting and comparing the performance of both algorithms. Novel AdaBoostM1(97.0%) better accuracy than CNN-LSTM (76.6%).

DECLARATIONS

Conflicts of interests

No conflicts of interest in this manuscript.

Authors Contribution

Author JTK was involved in conceptualization, data collection, data analysis, manuscript writing. Author UP was involved in conceptualization, guidance, 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 research.

1. Sri Design Card Technologies Solutions Pvt Ltd, Chennai.

2. Saveetha University.

3. Saveetha Institute of Medical and Technical Sciences.

4. Saveetha School of Engineering.

REFERENCES

1. Alatabbi, Ali. 2013. “Malware Detection Using Computational Biology Tools.” International Journal of Engineering and Technology. https://doi.org/10.7763/ijet.2013.v5.566.
2. Ayu, Indah Shekar Melati, Ayu Indah Shekar Ayu, and Kanisius Karyono Karyono. 2014. “Audio Detection (Audition): Android Based Sound Detection Application for Hearing-Impaired Using AdaBoostM1 Classifier with REPTree Weaklearner.” 2014 Asia-Pacific Conference on Computer Aided System Engineering (APCASE). https://doi.org/10.1109/apcase.2014.6924487.
3. Christodorescu, Mihai, Somesh Jha, Douglas Maughan, Dawn Song, and Cliff Wang. 2007. Malware Detection. Springer Science & Business Media.
4. Devarajan, Yuvarajan, Beemkumar Nagappan, Gautam Choubey, Suresh Vellaiyan, and Kulmani Mehar. 2021. “Renewable Pathway and Twin Fueling Approach on Ignition Analysis of a Dual-Fuelled Compression Ignition Engine.” Energy & Fuels: An American Chemical Society Journal 35 (12): 9930–36.
5. Dhanraj, Ganapathy, and Shanmugam Rajeshkumar. 2021. “Anticariogenic Effect of Selenium Nanoparticles Synthesized Using Brassica Oleracea.” Journal of Nanomaterials 2021 (July). https://doi.org/10.1155/2021/8115585.
6. Hu, Jintao, and Yurong Song. 2015. “The Model of Malware Propagation in Wireless Sensor Networks with Regional Detection Mechanism.” Communications in Computer and Information Science. https://doi.org/10.1007/978-3-662-46981-1_61.
7. Jin, Xiang, Xiaofei Xing, Haroon Elahi, Guojun Wang, and Hai Jiang. 2020. “A Malware Detection Approach Using Malware Images and Autoencoders.” 2020 IEEE 17th International Conference on Mobile Ad Hoc and Sensor Systems (MASS). https://doi.org/10.1109/mass50613.2020.00009.
8. Kamath, S. Manjunath, K. Sridhar, D. Jaison, V. Gopinath, B. K. Mohamed Ibrahim, Nilkantha Gupta, A. Sundaram, P. Sivaperumal, S. Padmapriya, and S. Shantanu Patil. 2020. “Fabrication of Tri-Layered Electrospun Polycaprolactone Mats with Improved Sustained Drug Release Profile.” Scientific Reports 10 (1): 18179.
9. Kutlu, Kutlu, and Avcı. 2019. “A Novel Method for Classifying Liver and Brain Tumors Using Convolutional Neural Networks, Discrete Wavelet Transform and Long Short-Term Memory Networks.” Sensors. https://doi.org/10.3390/s19091992.
10. Liu, Hanwen, Xiaohan Helu, Chengjie Jin, Hui Lu, Zhihong Tian, Xiaojiang Du, and Khalid Abualsaud. 2020. “A Malware Detection Method for Health Sensor Data Based on Machine Learning.” 2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT). https://doi.org/10.1109/iciot48696.2020.9089478.
11. Li, Xiangang, and Xihong Wu. 2015. “Long Short-Term Memory Based Convolutional Recurrent Neural Networks for Large Vocabulary Speech Recognition.” Interspeech 2015. https://doi.org/10.21437/interspeech.2015-648.

12. Mohanta, Abhijit, and Anoop Saldanha. 2020. "Malware Analysis Lab Setup." Malware Analysis and Detection Engineering. https://doi.org/10.1007/978-1-4842-6193-4_2.
13. Nandhini, Joseph T., Devaraj Ezhilarasan, and Shanmugam Rajeshkumar. 2020. "An Ecofriendly Synthesized Gold Nanoparticles Induces Cytotoxicity via Apoptosis in HepG2 Cells." Environmental Toxicology, August. https://doi.org/10.1002/tox.23007.
14. Parakh, Mayank K., Shriraam Ulaganambi, Nisha Ashifa, Reshma Premkumar, and Amit L. Jain. 2020. "Oral Potentially Malignant Disorders: Clinical Diagnosis and Current Screening Aids: A Narrative Review." European Journal of Cancer Prevention: The Official Journal of the European Cancer Prevention Organisation 29 (1): 65–72.
15. Perumal, Karthikeyan, Joseph Antony, and Subagunasekar Muthuramalingam. 2021. "Heavy Metal Pollutants and Their Spatial Distribution in Surface Sediments from Thondi Coast, Palk Bay, South India." Environmental Sciences Europe 33 (1). https://doi.org/10.1186/s12302-021-00501-2.
16. Pham, Quoc Hoa, Supat Chupradit, Gunawan Widjaja, Muataz S. Alhassan, Rustem Magizov, Yasser Fakri Mustafa, Aravindhan Surendar, Amirzhan Kassenov, Zeinab Arzehgar, and Wanich Suksatan. 2021. "The Effects of Ni or Nb Additions on the Relaxation Behavior of Zr55Cu35Al10 Metallic Glass." Materials Today Communications 29 (December): 102909.
17. Portegys, Thomas E. 2010. "A Maze Learning Comparison of Elman, Long Short-Term Memory, and Mona Neural Networks." Neural Networks. https://doi.org/10.1016/j.neunet.2009.11.002.
18. Rajalakshmi, B., and N. Anusha. 2017. "Sensor Based Application for Malware Detection in Android OS(Operating System) Devices." 2017 International Conference on Information Communication and Embedded Systems (ICICES). https://doi.org/10.1109/icices.2017.8070722.
19. Sathiyamoorthi, Ramalingam, Gomathinayakam Sankaranarayanan, Dinesh Babu Munuswamy, and Yuvarajan Devarajan. 2021. "Experimental Study of Spray Analysis for Palmarosa Biodiesel-diesel Blends in a Constant Volume Chamber." Environmental Progress & Sustainable Energy 40 (6). https://doi.org/10.1002/ep.13696.
20. Tesfaye Jule, Leta, Krishnaraj Ramaswamy, Nagaraj Nagaprasad, Vigneshwaran Shanmugam, and Venkataraman Vignesh. 2021. "Design and Analysis of Serial Drilled Hole in Composite Material." Materials Today: Proceedings 45 (January): 5759–63.
21. Uganya, G., Radhika, and N. Vijayaraj. 2021. "A Survey on Internet of Things: Applications, Recent Issues, Attacks, and Security Mechanisms." Journal of Circuits Systems and Computers 30 (05): 2130006.
22. Vemparala, Swapna. n.d. "Malware Detection Using Dynamic Analysis." https://doi.org/10.31979/etd.48fu-qckf.
23. Ye, W., J. Cheng, F. Yang, and Y. Xu. 2019. "Two-Stream Convolutional Network for Improving Activity Recognition Using Convolutional Long Short-Term Memory Networks." IEEE Access. https://doi.org/10.1109/access.2019.2918808.
24. Zhang, Cha, and Yunqian Ma. 2012. Ensemble Machine Learning: Methods and Applications. Springer Science & Business Media.
25. Zhu, Feng. n.d. "Integrity-Based Kernel Malware Detection." https://doi.org/10.25148/etd.fi14110701.

TABLES AND FIGURES

Table 1. Group Statistics analysis for both algorithms based on Accuracy. Accuracy values are compared between the Novel AdaBoostM1(97.0%) and Convolutional Neural Networks-Long Short-Term(79.6%). It shows that Novel AdaBoostM1seems to have better accuracy than the Convolutional Neural Networks-Long Short-Term.

Group Statistics

	Group	N	Mean	Std.Deviation	Std.Error Mean
Accuracy	Novel AdaBoostM1	10	93.8500	2.10145	0.66454
	Convolutional Neural Networks-Long Short-Term(CNN-LSTM)	10	76.3100	2.38290	0.75354

Table 2 represents the independent sample test with f1 score, level of significance as 0.05, and 95% confidence interval differences for CNN-LSTM and Novel AdaBoostM1. It found that Novel AdaBoostM1 significantly different from CNN-LSTM with 0.089 (p>0.05)

	Levene's Test for Equality of Variances		T-test of Equality of Means					95% of the confidence interval of the Difference	
Accuracy	**F**	**Sig.**	**t**	**df**	**Sig (2-tailed)**	**Mean Difference**	**Std Error Difference**	**Lower**	**Upper**

Equal Variance Assumed	0.019	0.893	17.935	18.0	0.000	18.41000	1.02651	16.25338	20.56662
Equal Variance Not Assumed	-	-	17.935	17.789	0.000	18.41000	1.02651	16.25245	20.56755

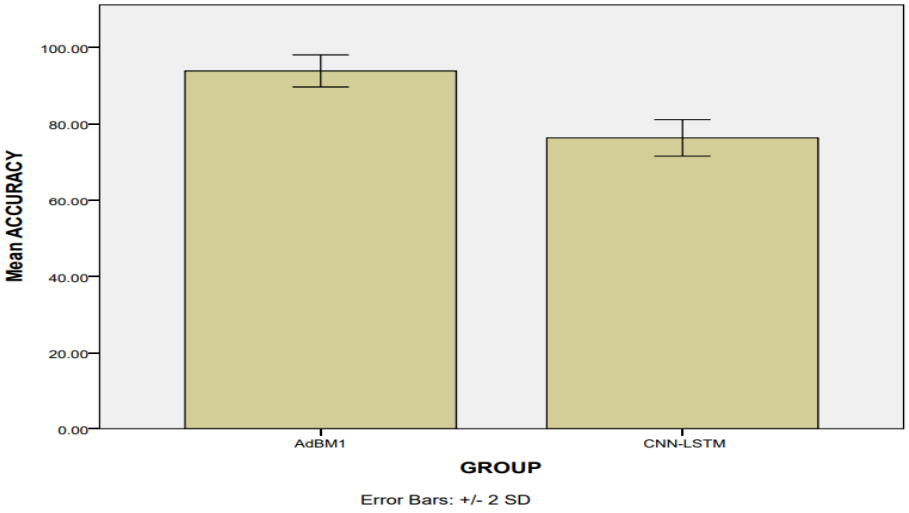

Fig. 1 Barcharts represent the comparison of mean accuracy and Standard Errors for CNN-LSTM and Novel AdaBoostM1. Novel AdaBoostM1 is better than CNN-LSTM in terms of Mean Accuracy and Standard Deviation. X-Axis: CNN-LSTM vs AdaBoostM1(AdBM1) Y-Axis: Mean Accuracy of detection ± 2 SD.