A Critical Review of Usha, G. et al, 2021, Enhanced Ransomware Detection Techniques using Machine Learning Algorithms, 4th International Conference on Computing and Communication Technologies (ICCCT), pp 52-58 and Masum, M. et al, 2022, Ransomware Classification and Detection with Machine Learning Algorithms, 12th International Conference on Computing and Communications Technologies (ICCCT), pp 316-322

**Assessment number: J107790**

**INTRODUCTION**

The aims of this paper is to identify and summarise the key findings of two research papers and to make a comparison by critically evaluating the strengths and weaknesses of the two papers. The first paper is by Usha, G. et al, titled “Enhanced Ransomware Detection Techniques using Machine Learning Algorithms”, published in 2021 which has three primary aims. The first aim is to discuss existing research in ransomware detection techniques and their weaknesses. The second aim is to compare various machine learning algorithms which detect ransomware attacks to ransomware datasets. The final aim of the paper is to collect a variety of behavioural data for each ransomware and to compare the results and attributes to understand the behaviour of these attacks. To complete the first aim, Usha, G. et al analysed a vast number of research papers surrounding current literature on ransomware in order to create an insight into the types of ransomware attacks and a detailed working of how the attacks function. Different detection methods were studied with advantages and disadvantages found for each one, as well as a study into the recovery of a system following an attack. The second aim was completed by identifying the stages of ransomware attacks, which are distribution, infection, staging, scanning, encryption and payday. By observing the different stages, it was found countermeasures would need to be in place before stages 3 and 4 to prevent major damage to the system. Four machine-learning algorithms used for ransomware detection called “K-nearest neighbor” (KNN for short), “Decision Tree”, “Random Forest” and “Gaussian Naïve Bayes” were studied and used to create a system architecture used for the detection of ransomware attacks. The final aim compared the accuracy of detection for each algorithm. The results found that the Random Forest algorithm had the highest percentage of accuracy of 96%, followed by KNN with 94.7%, Decision Tree with 91.1% and Naïve Bayes with the lowest accuracy of 90.2%.

The second paper is by Masum, M. et al, titled “Ransomware Classification and Detection with Machine Learning Algorithms”, published in 2022 and has two primary focuses. The first aim of this paper is to carry out a detailed investigation on the classification of the types of ransomwares and use this information to propose a feature selection-based framework using features adopted from other classifiers and architectures. The second aim is to collect datasets of each classification of the types of ransomwares and use these datasets to compare the machine learning algorithms and their effectiveness to detect ransomware attacks. The paper focuses on five types of detection algorithms which are “Decision Tree”, “Random Forest”, “Naïve Bayes”, “Logistic Regression” and “Neural Network-based classifiers”. The first aim of the paper was completed by analysing a variety of research papers which study conventional detection techniques on malware. Masum, M. et al found most ransomware families shared common traits including payload persistence, stealth techniques and network traffic. After conducting research on the classification of ransomware, Masum, M. et al found conventional classification techniques limited and proposed using modern machine learning algorithms to improve the detection and prevention of ransomware attacks. The second aim compared different datasets used on each classifier and gave results on their accuracy, F-beta, recall and precision. The results showed that Random Forest has the overall best scores for accuracy, F-beta and precision, with the second-best result in recall, making it the most reliable of the five classifiers.

Both papers by Usha, G. et al and Masum, M. et al had proven to show in-depth analysis of a variety of algorithms with both research and experimental data. Masum, M. et al showed a more in-depth analysis compared to Usha, G. et al but was not written as reader friendly due to the overuse of statistics and terminology without clear clarification. Both papers found RF to be the best algorithm used as a classifier with strong data to support their claims but showed weakness in their evaluation and conclusion of their results due to their lack of discussion of the other algorithms.

**SUMMARY**

The paper published by Usha, G. et al has three primary focuses, starting with researching current methods of detecting ransomware attacks to evaluate their strengths and weaknesses. Usha, G. et al carries out in-depth analysis of three research papers and combines the information from four others, collecting data from seven papers overall. The authors use the information gathered from these papers to determine the best machine learning algorithms found in those papers as well as their advantages and disadvantages as a basis for their own research and to map the stages of ransomware attacks. Using the data obtained from these papers, Usha, G. et al focused on a few specific algorithms and found the non-parametric method K-Nearest Neighbor (KNN) to be one of the easiest algorithms to use when identifying ransomware classification and regression. The next algorithm the authors look at in-depth is the Decision Tree (DT) algorithm which was broken down into a number of stages for the detection and regression of ransomware attacks. The next algorithm is Random Forest (RF) which the authors found to be a simple method to use which gives consistently reliable results and one of the best to use for classification and regression. The last algorithm Usha, G. et al looked at in-depth was Gaussian Naïve Bayes (NB) which the authors determined was an easy method for classification with high functionality.

The second focus from Usha, G. et al was to identify the stages of ransomware attacks, which were split into six defined stages. The first stage, “Distribution”, is where the malware is created and distributed through the internet. The second stage, “Infection”, is where the malware enters the system. This can be through social engineering attacks or using other malware. The third stage, “Staging”, is where the ransomware creates a working environment within the system by infecting multiple folders. The fourth stage, “Scanning”, is where the ransomware scans for folders with important documents to use as blackmail against the owner. The fifth stage, “Encryption”, is where the ransomware encrypts the important folders. The final stage, “Payday”, is where the ransomware creates a ransom note which is displayed to the user, most commonly as text or on the wallpaper.

The final focus from Usha, G. et al was to look at four machine learning algorithms to determine the accuracy each algorithm has at detecting ransomware attacks and using this data as well as the information gathered from other research papers to determine the best method of detection and regression. The four algorithms analysed in this paper were KNN, DT, RF and NB. The algorithms were used on ISOT Ransomware Detection Dataset which was obtained from IST research lab to complete the experimental results of the paper. This dataset is a collection of 669 benign samples from families most widely used. Usha, G. et al determined RF to be the best method as it has the highest accuracy of the four for detecting ransomware attack at 96% whilst also being able to distinguish between benign and infected files. The drawback to RF is the costly use of estimation trees on the system, but this does give the method a better and more consistent accuracy. The authors state the best alternative to RF would be KNN as it has a high accuracy of detention with 94.7% but has the drawback of being best utilized on smaller datasets due to high real execution time cost.

The paper published by Masum, M. et al has two primary focuses, starting with carrying out a detailed investigation of other research papers to classify the types of ransomwares and using the information and data obtained to propose a feature selection-based framework. Masum, M. et al found that most ransomware families share common behavioural traits which include, but not limited to, payload persistence, stealth techniques and network traffic. After conducting research on the classification of ransomware, Masum, M. et al found conventional classification techniques limited and proposed using modern machine learning algorithms to improve the detection and prevention of ransomware attacks. Using this data, Masum, M. et al proposed the framework seen in figure 1 (Masum, M. et al, 2022).

A diagram of a software algorithm

Description automatically generated

The second focus from Masum, M. et al was to collect datasets of each classification of the types of ransomwares to use on five machine learning algorithms to evaluate their effectiveness of detecting ransomware attacks. The algorithms analysed by Masum, M. et al were Decision Tree (DT), Random Forest (RF), Naïve Bayes (NB), Logistic Regression (LR) and Neural Network (NN)-based classifiers. The authors collected four types of data on each algorithm: accuracy, F-beta, Recall and Precision, each with standard deviation. Each algorithm also had an ROC curve produced with the area under curve (AUC) calculated. The results obtained showed RF to have the highest accuracy of 0.99 standard deviation of only 0.01, the highest F-beta of 0.97 with a standard deviation of 0.03 and the highest degree of precision of 0.99 with a standard deviation of 0.00. RF had the second-best recall score of 0.97 with a standard deviation of 0.03, second to NB with a score of 0.99 and a standard deviation of 0.00. These results showed RF to be the best classifier overall with DT and NN showing a comparable performance. The ROC curve data produced showed RF, LR and NN to have a consistent AUC score of 0.99 making them equally consistent, with NB scoring 0.73 making it very inconsistent.

**CRITIQUE**

Both papers published by Usha, G. et al and Masum, M. et al focused on the use of machine learning algorithms to detect and classify ransomware attacks. According to California State University, a simple guide to the structure of a research paper is to state the author and author’s professional affiliation, write an introduction, literature review, methodology, results, discussion and conclusions. Schwartz, B. et al provided a more detailed structure of an APA style research paper which both Usha, G. et al and Masum, M. et al have structured similarly (Schwartz, B. et al, 2020). Usha, G. et al have written a strong introduction showing a detailed level of research to achieve their first aim. The purpose of their research was stated clearly by giving good examples of the importance of ransomware detection techniques and a description of malware and ransomware. Usha, G. et al listed reasons for the importance of protection against malware, such as poor cloud but not include any data to support such claims and how this links to malware and ransomware. Usha, G. et al could reference Singh, A. & Chatterjee, k. or Sethi, S. & Sruti, S. where they state many issues with cloud security including the use of third parties which cause an increase in threat (Singh, A. & Chatterjee, k., 2017). Despite the lack of citations to support a number of claims, the literature is written in an easy way for the reader to follow. Similarly, Masum, M. et al have structured the introduction well with a high level of research done prior. Masum, M. et al had shown more depth to their research comparing much more data, but the structure of their literature makes it more difficult to follow and comprehend with their frequent use of data, abbreviations and terminology.

Both Usha, G. et al and Masum, M. et al studied the effectiveness of DT, NB and RF. Usha, G. et al also studied KNN where Masum, M. et al studied LR and NN. Both papers provided detailed statistics for each algorithm giving key strengths and weaknesses. Usha, G. et al provided a detailed conclusion of the results for RF and KNN but did not critique the results of the other algorithms in the conclusion. To improve upon the conclusion, Usha, G. et al could mention more of the limitations of each algorithm, such as RF has the limitations of the choice of parameters and interference of noises (Ren, Q. et al, 2017), or KNN having a low efficiency in classification applications (Zhao, W. et al, 2012). Masum, M. et al provided a much less detailed conclusion, only stating RF to be the best classifier based upon their results. To improve upon this, Masum, M. et al could discuss the advantages and disadvantages of each algorithm based on their results and research, such as the limitations stated by Ren, Q. et al or Zhao, W. et al stated above. Both Usha, G. et al and Masum, M. et al proposed RF to be the best classifier based on their results with strong data to support. Usha, G. et al went on to describe the future scope of this area of research by discussing how the methods can be improved, whereas this is an area Masum, M. et al did not discuss and is an area for improvement. Masum, M. et al could discuss future scope especially with the promise shown from NN classifiers, with a vast number being developed and improved upon for different purposes, such as Perceptron, Feed Forward, Multilayer, Convolutional and many others (Team, G. L., 2020).

**CONCLUSION**

Both papers from Usha, G. et al and Masum, M. et al were very detailed and gave in-depth insight into the use of machine learning algorithms for ransomware detection and classification. Whilst Usha, G. et al did not provide as much statistical data as Masum, M. et al, the literature was easier to understand overall making it more reader friendly. The research and experimental analysis of both papers was very detailed with significant data being produced for each algorithm. The weakness for both papers lay in their conclusions. Usha, G. et al did not meet one of their aims in

identifying the shortcomings of each algorithm and needed to go deeper into this aspect. Despite this, Usha, G. et al met the rest of the criteria and proposed further areas of study. Whilst Masum, M. et al met both of their aims in the paper by proposing a detailed framework for ransomware detection and carrying out in depth analysis of the machine learning algorithms, there is much the authors could do to improve on the conclusion due to its lack of depth and analysis. Both papers could be improved by discussing the advantages and disadvantages found for each algorithm and using this information to list from best to worst.

**REFERENCES**

1. Masum, M., Hossain Faruk, M. J., Shahriar, H., Qian, K., Lo, D., & Adnan, M. I. (2022, January 1). Ransomware Classification and Detection With Machine Learning Algorithms. IEEE Xplore. https://doi.org/10.1109/CCWC54503.2022.9720869
2. Ren, Q., Cheng, H., & Han, H. (2017). Research on machine learning framework based on random forest algorithm. https://doi.org/10.1063/1.4977376
3. Schwartz, B. M., Landrum, R. E., & Gurung, R. A. R. (2020). An EasyGuide to APA Style. In Google Books. SAGE Publications. https://books.google.co.uk/books?hl=en&lr=&id=GyHmDwAAQBAJ&oi=fnd&pg=PT12&dq=How+to+write+an+APA-style+research+report&ots=DI\_5VWjGmv&sig=c3XCOlRP6JaEQxFl7k3hlJUqzGM&redir\_esc=y#v=onepage&q=How%20to%20write%20an%20APA-style%20research%20report&f=false
4. Sethi, S., & Sruti, S. (2017). Cloud Security Issues and Challenges. Www.igi-Global.com; IGI Global. https://www.igi-global.com/chapter/cloud-security-issues-and-challenges/171349
5. Singh, A., & Chatterjee, K. (2017). Cloud security issues and challenges: A survey. Journal of Network and Computer Applications, 79, 88–115. https://doi.org/10.1016/j.jnca.2016.11.027
6. Structure of Typical Research Article | California State University Monterey Bay. (n.d.). Csumb.edu. https://csumb.edu/library/library-instruction/structure-typical-research-article/
7. Team, G. L. (2020, April 29). Types of Neural Networks and Definition of Neural Network. GreatLearning. https://www.mygreatlearning.com/blog/types-of-neural-networks/
8. Usha, G., Madhavan, P., Vimal Cruz, M., Vinoth, N. A. S., Veena, & Nancy, M. (2021). Enhanced Ransomware Detection Techniques using Machine Learning Algorithms. 2021 4th International Conference on Computing and Communications Technologies (ICCCT). https://doi.org/10.1109/iccct53315.2021.9711906
9. Zhao, W., Tang, S., & Dai, W. (2012). An Improved kNN Algorithm based on Essential Vector. Electronics and Electrical Engineering, 123(7). https://doi.org/10.5755/j01.eee.123.7.2389