LAUTECH Journal of Engineering and Technology 18 (4) 2024: 23-33

10.36108/laujet/4202.81.0430

Comparative Analysis of Dimensionality Reduction on Ransomware Detection Using Machine Learning Techniques

[1]Akinola Olaoluwa, [*2]Amusan Elizabeth and [3]Adeosun Olajide

[1]Department of Computer Science, Ladoke Akintola University of Technology, Ogbomoso; sotm13@gmail.com
2Department of Cyber Security Science, Ladoke Akintola University of Technology, Ogbomoso; eaadewusi@lautech.edu.ng
[3]Department of Computer Science, Ladoke Akinola University of Technology, Ogbomoso: ooadeosun@lautech.edu.ng

Article Info

Article history:

Received: Sept. 17, 2024
Revised: Oct. 25, 2024
Accepted: Nov. 3, 2024

Keywords:

Dimensionality Reduction,
Features,
Machine Learning,
Ransomware Detection
.

Corresponding Author:
eaadewusi@lautech.edu.ng

ABSTRACT

Ransomware attacks continue to evolve as a pervasive threat to cybersecurity such as data loss, financial losses, and potential disruption of critical services which have prompted the need for robust detection mechanisms. Leveraging on machine learning techniques for ransomware detection has gained recognition; however, the high-dimensional nature of feature spaces has posed some challenges in model efficiency and effectiveness. This research therefore explores the impact of two well-known dimensionality reduction methods that may enhance ransomware detection using five popularly used machine learning algorithms which are K-Nearest Neighbor (KNN), Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM) and Naive Bayes (NB). Through comprehensive analysis and experimentation, two well-known dimensionality reduction techniques, Linear Discriminant Analysis (LDA) and Principal Component Analysis (PCA) were examined on the selected machine learning algorithms using a Ransomware Portable Executable Header Feature Dataset publicly available on an online data repository with URL https://data.mendeley.com/datasets/p3v94dft2y/2 with 1028 features. Metrics such as Accuracy, Recall, Precision and F1-Score were used to evaluate the classifiers. The comparative analysis of LDA and PCA revealed a discernible preference for one classifier over another. From the results, it was observed that the performance of classifiers with PCA was better than that of with LDA. Also, Decision Tree and Random Forest classifiers outperform the other three algorithms without using dimensionality reduction as well as with both PCA and LDA.

INTRODUCTION

Ransomware is a type of malware from cryptovirology that threatens to publish the victim's data or perpetually block access to it unless a ransom is paid. Ransomware's main objective is extortion by imposing some form of denial of service to either the system or system resources such as files until a ransom is paid. This makes ransomware different from conventional malware that seeks to replicate, delete files, exhilarate data or extensively consume system resources **(**Urooj *et al.*, 2021**).** While some simple ransomware may lock the system so that it is not difficult for a knowledgeable person to reverse it, more advanced malware uses a technique called cryptoviral extortion (Alraizza and Algarn (2023). In a properly implemented cryptoviral extortion attack, recovering the files without the decryption key is an intractable problem. Digital currencies

such as Paysafecard or bitcoin and other cryptocurrencies are used for the ransoms. This makes tracing and prosecuting the perpetrators difficult.

In the past few decades, numerous dimensionality reduction techniques have been used for filtering the data samples of the considered dataset. Reduction of dimensionality requires mapping of inputs that are of high dimensionality to a lesser dimensionality so that similar points in the input space are mapped to neighboring points on the manifold. (Reddy *et al.*, 2020). Dimensionality reduction techniques can tremendously reduce the time complexity of the training phase of Machine Learning algorithms hence reducing the burden of the machine learning algorithms. As a result, dimensionality reduction facilitates, among others, classification, visualization, and compression of high-dimensional data. With the aforementioned merit of dimensionality reduction, this study analyzed dimensionality reduction and machine learning classification model for Ransomware.

Machine learning is considered ideal for analyzing the behavior of processes or applications because it can effectively learn patterns and anomalies in large datasets, which can be difficult for humans to detect. In the context of ransomware detection, machine learning algorithms can be trained on large datasets of both benign and malicious software to learn the behavioral characteristics that distinguish ransomware from legitimate software. This training can be used to identify new and previously unseen variants of ransomware, including zero-day attacks, based on their behavioral patterns (Khammas, 2022).

Ransomware attacks pose a significant threat to individuals and organizations, leading to data loss, financial losses, and potential disruption of critical services. While machine learning techniques have shown promise in detecting ransomware, the high-dimensional nature of feature spaces often presents challenges in model training and deployment (Alsaidi *et al*, 2022).

In this paper, the performance of the following Machine Learning (ML) classification algorithms: K-Nearest Neighbor (KNN), Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM) and Naive Bayes (NB) were examined concerning the effects of two well-known dimensionality reduction techniques: Linear Discriminant Analysis (LDA) and Principal Component Analysis (PCA) on a Ransomware PE Header Feature Dataset.

The aim of this paper was therefore to analyze the effect of dimensionality reduction on the performance of the selected five (5) machine learning algorithms in detecting ransomware. To achieve this, the research worked through the following objectives: the first was to acquire ransomware data and preprocess it. Secondly, dimensionality reduction was performed on the acquired preprocessed dataset using PCA and LDA. The third objective was to investigate the performance of dimensionality reduction on selected machine learning algorithms using metrics such as Accuracy, Recall, Precision and F1-Score. Finally, a comprehensive analysis of the performance of machine learning algorithms with and without dimensionality reduction was carried out

LITERATURE REVIEW

This section highlights some relevant concepts followed by a brief review of literature.

Ransomware Detection

There are three primary ways to detect ransomware. First is using the Signature-Based Detection Technique which as explained in Urooj *et al*, (2022)

is such that the patterns or codes of already available threats are compared with the examined code. One of the limitations of this approach is the outdated samples. New variants are developing with time which requires a periodic update to the available ransomware collection. Another method is the Behavior-Based Detection Technique where the behavior of the program is observed to determine whether it is malware or benign (Aslan and Samet, 2020). It looks for program behavior, not program code or code sequence. Even though the program codes are changed, the behavior of the program will be the same or similar; therefore, it can still be detected with this method. The third method is the heuristic technique which is based on experiences that rely on machine learning techniques to detect certain characteristic features of the relevant program. (Shah and Farik, 2017). The heuristic method is trained using several samples before the detector, then the trained system is run for new program samples to catch malware (Gorment *et al*, 2023).

Dimensionality Reduction Techniques

The two popular dimensionality reduction techniques, Principal Component Analysis and Linear Discriminant Analysis are discussed.

Principal Component Analysis (PCA)

Principal Component Analysis (PCA) is a statistical method that uses an orthogonal transformation. It converts a group of correlated variables to a group of uncorrelated variables. Also, PCA can be used for the examination of the relationships among a group of variables which then makes it fit for dimensionality reduction (Reddy et al., 2020). With the assumption that a dataset x^1, x^2, ….., x^m has n dimension inputs. N-dimension data must be reduced to k-dimension (k < n) using PCA. First, PCA involves standardization of the raw data and proceeds to calculate the covariance matrix of the same. Furthermore, the eigenvector and eigenvalue of the matrix is calculated as given in Equation 1.

$$u^t \sum = \lambda\mu$$

$$U = \begin{bmatrix} | & | & | \\ u_1 & u_2 & u_n \\ | & | & | \end{bmatrix},\ u_{i \in R^n} \quad (1)$$

The feature vector is formed using the eigenvectors of the covariance matrix, to reorient the data from the original axes to the ones represented by the principal components (hence the name Principal Components Analysis). This can be done by multiplying the transpose of the original data set by the transpose of the feature vector as shown in Equation 2.

$$Final\ DataSet = FeatureVector^T * StandardizedOriginalDataSet^T \quad (2)$$

Linear Discriminant Analysis (LDA)

Linear Discriminant Analysis **(**LDA) is another popular dimensionality reduction approach for pre-processing steps in data mining and machine learning applications. The main aim of LDA is to project a dataset with a high number of features onto a less-dimensional space with good class separability. This will reduce computational costs (Reddy *et al*., 2020). Both techniques employ linear transformation but while PCA is an unsupervised algorithm, LDA is a supervised algorithm.

Related Works

Carlin *et al.* (2018), developed a dynamical analysis with a detecting cryptomining technique. The dataset consists of 490 samples and is collected from VirusShark. Of a total of 490 samples, 194 are benign and Cryptomining has 296 HTML files or malicious samples. The Random Forest classifier is used and implemented in WEKA version 3.9. The data used 10-fold cross-validation. The best accuracy of Random Forest is 99.05%. The FPR is 99.7%, and FNR is 98.6 %

Also, Jerlin and Marimuthu. (2018), proposed an efficient Rate-based MDNBS (Multi-Dimensional Naïve Bayes Classification technique for malware classification using API call sequences. MDNBS is used to classify types of malware as worms, viruses, Trojans, or normal. Compared to other existing machine learning algorithms, their experimental results showed that the proposed technique has higher accuracy.

Similarly, Alrimy *et al.* (2018), presented a detection framework integrating two variants of support vector machine (SVM) classifiers. The author uses ordinary SVM for behavioural detection and one-class SVM (OCSVM) for anomaly detection. The studies gave a theoretical concept of pre-encryption of ransomware across ransomware families to build an early detection mode.

Furthermore, Alhawi *et al.* (2018), presented a machine learning- (ML) based solution for the detection of ransomware. The dataset was collected from VirusTotal which was a combination of both malicious and benign and contains 264 records having 9 ransomware families and 3 types of benign. Using dataset network traffic features, a true positive detection rate of 97.1 percent was obtained, and using a decision tree classifier, a zero false positive rate (FPR) and true positive rate (TPR) of 96.3 percent was achieved.

Kok, *et al*. (2019), proposed a Pre-Encryption Detection Algorithm (PEDA) that consists of two phases, PEDA phase one is a Windows application program interphase (API) generated by a suspicious program which is captured and analyzed using the Learning Algorithm (LA). The learning algorithm (LA) now further determines whether the suspicious program is cryptographic ransomware or not, through API pattern recognition. If the prediction was a crypto-ransomware, PEDA would generate a signature of the suspicious program, and store it in the signature repository, which is in Phase II. In PEDA-Phase-II, the signature repository allows the detection of crypto-ransomware at a much earlier stage, which was at the pre-execution stage through the signature matching method. This method can only detect known crypto-ransomware, and although very rigid, it was accurate and fast.

Almashhadani *et al.* (2019), described a detection system that extracted features exclusively from network traffic. The authors obtained twenty features from the characteristics of TCP, HTTP and DNS traffic. Almashhadani *et al.* (2019), built a decision-making module based on two classifiers. One classifier was based on per-packet features, while the second one was based on flow-based features. It did not require that both classifiers detect the ransomware. The classifiers evaluated were Random Forests, Bayes Networks, Support Vector Machines and Random Trees. The algorithm that provided the highest accuracy was selected for each classifier.

The work of (Cusack *et al.* 2018) proposed a ransomware detection model based on machine learning methods using network traffic data. The researchers monitored the network communication between the victim's machine and the command and control (C&C) to detect and prevent the delivery of the encryption key needed to encrypt the victim's files without which the encryption process did not start. The authors used dimensionality reduction techniques to find the eight attributes that most contribute to the detection of ransomware in network traffic. However, the solution suffers from having a 12.5% false positive rate, which can generate many false alarms.

Umme *et al.*, (2020), proposed a model that extracts the novel features from the ransomware dataset and performs classification of the ransomware and benign files. The proposed model can detect a large

number of ransomware from various families at runtime and scan the network, registry activities, and file system throughout the execution. API-call series was reutilized to represent the behavior-based features of ransomware. To predict the ransomware, the method uses online machine learning algorithms to analyze the fourteen-feature vector that is extracted at runtime. 78,550 recent ransomware datasets, both benign and malicious, were tested and compared with Ada Boost and random forest to confirm the efficacy and scalability. The testing accuracy was increased to 99.56 %.

Borah *et al. (*2021) performed a classification called ERAND (Ensemble Ransomware Defense) for defense against ransomware. The authors used the NSGAII to calculate the weights of five classifiers (ExtraTree, Gradient Boosting, AdaBoost, XGBoost and Random Forest) and achieved high accuracy, finding accuracies for each family above 95%. However, the methodology used by the authors became quite obscure and caused several different interpretations.

Ransomware detection studies carried out static, dynamic, and hybrid analysis to classify malware as ransomware or benign. Reddy *et al.* (2020) investigated four popular Machine Learning (ML) algorithms, Decision Tree Induction, Support Vector Machine (SVM), Naive Bayes Classifier and Random Forest Classifier using two of the prominent dimensionality reduction techniques, Linear Discriminant Analysis (LDA) and Principal Component Analysis (PCA) on Cardiotocography (CTG) dataset which was gotten from University of California and Irvine Machine Learning Repository. However, this work analyses the effect of dimensionality reduction on a Ransomware Portable Executable Header Feature Dataset obtained from an online data repository with URL https://data.mendeley.com/datasets/p3v94dft2y/2 with 1028 features using five Machine Learning (ML) classification algorithms: K-Nearest Neighbor (KNN), Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM) and Naive Bayes (NB) were examined with the effects of two well-known dimensionality reduction techniques which are Linear Discriminant Analysis (LDA) and Principal Component Analysis (PCA)

METHODOLOGY

Research Steps

The steps employed in analysing and evaluating the impact of dimensionality reduction on the efficiency of the selected machine learning algorithms' performance architecture are outlined below:

1) In step 1, the min-max standard scaler normalization method is applied to scale the features of the dataset to a range between 0 and 1. This step ensures that all features have the same scale, preventing certain features from dominating others during dimensionality reduction. Conversion of Categorical data within the ransomware dataset, such as labels or types, are converted into the numerical format. This conversion enables the algorithms to interpret and analyze the data effectively.

2) In step 2 feature engineering using Linear Discriminant Analysis (LDA) and Principal Component Analysis (PCA) were applied on normalized datasets to perform dimensionality reduction.

3) In step 3 the resultant dataset then be fed into the Machine Learning algorithms Decision Tree, Naive Bayes, Random Forest and SVM. The performance of these classifiers was then evaluated on the metrics; Accuracy, Precision, Recall, and F1-Score.

4) In step 4 the results obtained by the Machine Learning algorithms with and without

dimensionality reduction were analyzed to determine the effect of dimensionality reduction on the performance of the respective algorithms.

5) Steps 1 to 4 were repeated on Ransomware Datasets to analyze the performance of PCA and LDA.

A diagrammatic representation of these steps is shown in Figure 1 which is the block diagram on PCA and LDA dimensionality reduction techniques.

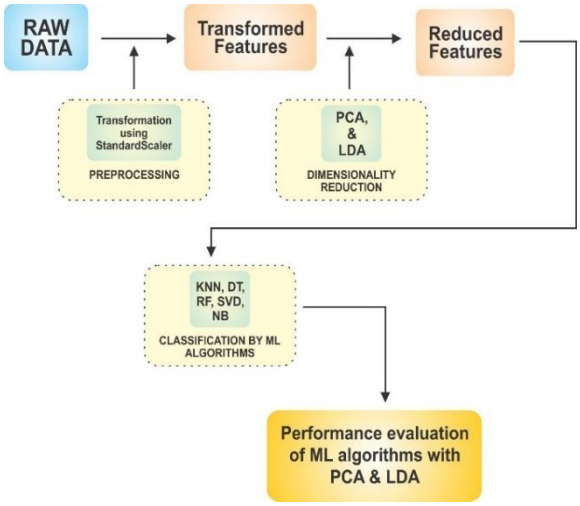

Figure 1: Block diagram on PCA and LDA dimensionality reduction techniques.

RESULTS AND DISCUSSION

This experimentation was performed on a Ransomware Portable Executable Header Feature Dataset which is available on an online data repository with the URL https://data.mendeley.com/datasets/p3v94dft2y/2. Implementation was done using Python programming language on a machine with a 2.60GHz Intel Core i7 processor, 16GB RAM and Debian Stretch (Linux) operating system. This section presents the results of the analysis with and without dimensionality reduction.

Dataset Description

The acquired dataset from an online data depository with URL https://data.mendeley.com/datasets/p3v94dft2y/2 contains headers of 2157 binary executable samples comprising 1134 legitimate software (goodware) and 1023 ransomware, grouped into 25 ransomware families. The dataset was retrieved by extracting raw information of the Portable Executable header. The CSV file columns are sample ID, filename, target class (GR), family ID, and numerical columns from 0 to 1023, as shown in Table 1.

Table 1: Ransomware PE Header Dataset

	ID	filename	GR	family	0 – 1023
Goodware	10000 to 11133	Their name.exe	0	0	Numerical features ranging from 0 to 255
Ransomware	20000 to 21022	Their SHA-256 hash	1	25 family IDs	Numerical features ranging from 0 to 255

Feature Extraction

Delving into how the features are extracted, is a crucial step in this project to reduce the dimensionality of ransomware dataset using PCA and LDA. Feature extraction involves transforming the original high-dimensional dataset into a lower-dimensional one while retaining important information. Explained variance ratio, feature selection, and why we chose certain features explained the importance of the selected features.

Explained variance ratio

In PCA, each principal component captures a portion of the dataset's variance. The explained

variance ratio of a component shows how much variance it covers. A higher ratio means more important information is preserved.

Feature selection

To select features, principal components were sorted by their explained variance ratios in PCA. Looking at the cumulative explained variance ratio plot (Figure 2) to decide how many components to keep. 25 components were selected to balance keeping enough information and reducing dimensionality.

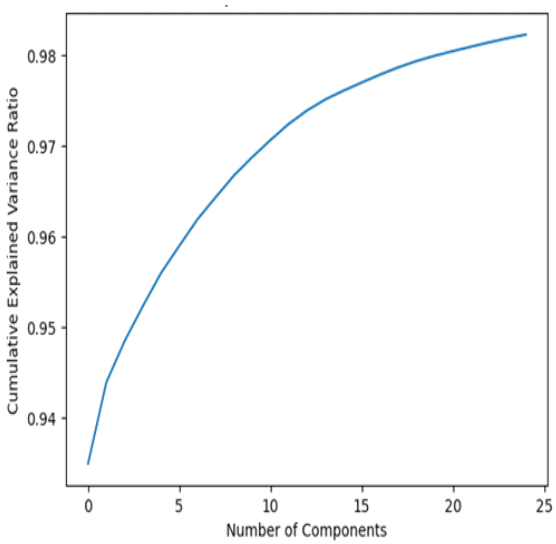

Figure 2: Cumulative Explained Variance Ratio Plot for PCA

In LDA, linear discriminants with higher eigenvalues was picked, which represent more separation between classes. This helps distinguish classes better for tasks like classification as shown in Figure 3

Rationale for Feature Selection

25 components and discriminants were selected to balance keeping important information while simplifying analysis. This way, yet maintain most of the original variance while making computations faster and easier.

In short, the features extraction approach through PCA and LDA gave a concise outcome and informative dataset, which was used for various analyses in this project.

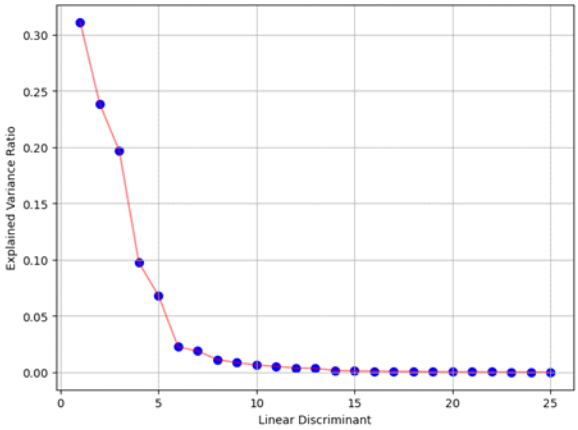

Figure 3: Explained Variance Ratio per Linear Discriminant

Metrics for Evaluation of the Models

To evaluate the performance of the selected machine learning algorithms for ransomware classification, different measures were considered for effectiveness. Therefore, the performance evaluation was based on the following classification evaluation parameters:

Accuracy: expressed as an overall view based on the number of predictions being classified correctly. In a high-level scale, it is easy to represent accuracy in the form of the number of true positive and negative events over the total in 100%. The equation for accuracy is shown in the Equation 3

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} x100 \quad (3)$$

Recall: this is the measure of the number of true positive predictions over the total number of true positive and false negative events as denoted in Equation 4. This metric is used to show the percentage of the true positive rate that can tell us the ratio of events that truly have ransomware.

$$Sensitivity = \frac{TP}{TP+FN}\ x\ 100. \quad (4)$$

Precision: Precision also known as specificity, is the measure of the number of true positive events over the total amount of true positive and false positive events as shown in Equation 5. This metric tells us the portion of correct positive classifications of ransomware from cases that are predicted as positive.

$$\text{Precision} = \frac{TN}{TN+FP} x100. \quad (5)$$

F1 score: The harmonic means of Precision and Recall. $F1\ score$ is a better performance metric than the accuracy metric for imbalanced data as shown in Figure 6.

$$F1 = 2x\frac{Precision\ x\ Recall}{Precision+Recall} \quad (6)$$

Performance Evaluation of Classifiers Without Dimensionality Reduction

The results of experimentation are discussed in this section. First the dataset, without dimensionality reduction is experimented with using the following machine learning algorithms: K-Nearest Neighbor (KNN), Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM) and Naive Bayes (NB).

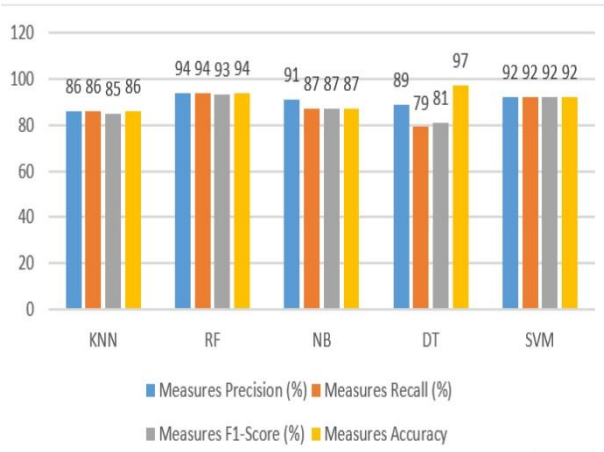

Figure 4: Classifiers Scores without Dimensionality Reduction

Figure 4 shows the performance of aforementioned algorithms on the dataset based on accuracy, sensitivity, specificity and F1-Score, it shows that all algorithms performed almost equally well on all measures. Accuracy of KNN, DT, NB, RF and SVM are 86%, 97%, 87%, 94%, and 92%, respectively. The sensitivity achieved by these algorithms is 72%, 84%, 75%, 88% and 83% respectively. Specificity achieved is 79%, 93%, 84%, 73% and 85% respectively. F1-Score achieved is 73%, 87%, 76%, 66% and 83% respectively. The above results show that random forest and support vector machine outperformed the other three algorithms in terms of specificity, sensitivity and F1-score where RF, DT and SVM performed better than other algorithms in terms of accuracy.

Performance Evaluation of Classifiers with PCA and LDA

PCA (Principal Component Analysis)

Applying PCA (Principal Component Analysis) as a dimensionality reduction technique to reduce the number of features in the dataset from 1028 features to 25 features. Then the dataset with reduced features is evaluated using K-Nearest Neighbor (KNN), Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM) and Naive Bayes (NB) classifiers. Figure 5 shows the performance of these classifiers on a reduced dataset in terms of accuracy, sensitivity, specificity and F1-Score measures. Accuracy of KNN, DT, NB, RF and SVM are 89%, 95%, 95%, 95%, and 91% respectively. The sensitivity achieved by these algorithms is 84%, 94%, 95%, 95% and 91% respectively.

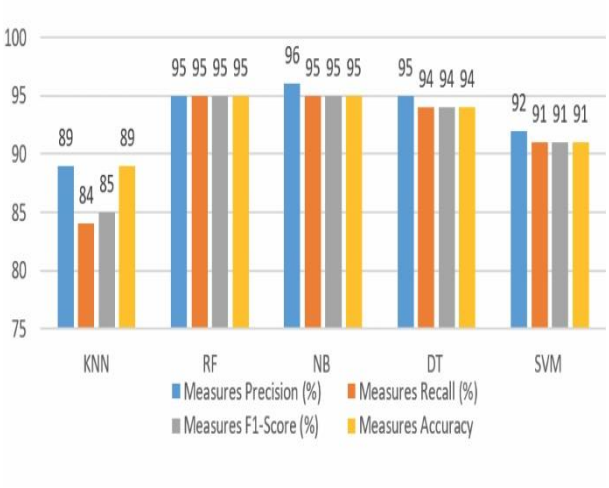

Figure 5: Classifiers Scores with PCA

Specificity achieved is 89%, 95%, 96%, 95% and 92% respectively. F1-Score achieved is 85%, 94%, 95%, 95% and 91% respectively. The above results show that NB, RF and DT outperformed the other two algorithms in terms of specificity, sensitivity and F1-score where RF, DT and NB also performed better than other algorithms in terms of accuracy.

LDA (Linear Discriminant Analysis)

Applying LDA (Linear Discriminant Analysis) as a dimensionality reduction technique to reduce the number of features in the dataset from 1028 features to 25 features. Then the dataset with reduced features is evaluated using K-Nearest Neighbor (KNN), Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM) and Naive Bayes (NB) classifiers.

Figure 4 shows the performance of these classifiers on a reduced dataset in terms of accuracy, sensitivity, specificity and F1-Score measures. Accuracy of KNN, DT, NB, RF and SVM are 84%, 79%, 85%, 85%, and 83% respectively. The sensitivity achieved by these algorithms is 89%, 97%, 87%, 85% and 83% respectively. Specificity achieved is 89%, 98%, 91%, 89% and 89% respectively. F1-Score achieved is 88%, 97%, 87%, 85% and 85% respectively. The results here show that DT outperformed the other four algorithms in terms of specificity, sensitivity and F1-score where RF and NB also performed better than other algorithms in terms of accuracy.

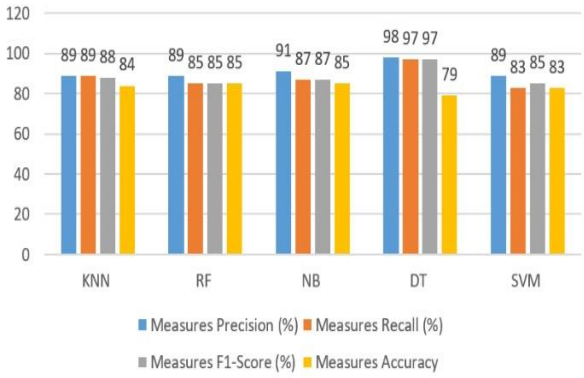

Figure 6: Classifiers Scores with LDA

Comparison between Classifiers' Accuracy Scores

The comparative analysis of LDA and PCA reveals a discernible preference for one classifier over another. The Decision Tree (DT) classifier, for instance, demonstrates superior performance with an accuracy rate of 97% when applied without dimensionality reduction on the ransomware samples, which consist of 1028 features. Conversely, its performance diminishes notably to an accuracy score of 79% when subjected to LDA with a dimensionality reduction to 25 features. However, the classifier exhibits improved performance with PCA, achieving an accuracy score of 94% compared to LDA. This underscores the intrinsic characteristics of the dataset, indicating minimal or absent linearity. Furthermore, the reduction in dimensions impairs the classifier's learning capacity.

In contrast, the classifiers exhibit suboptimal performance under LDA, with NB and RF attaining their highest accuracy scores at 85% and DT at its lowest at 79%. This is in stark contrast to the classifiers' performance under PCA, where NB and

RF achieve a high accuracy score of 95% and KNN records the lowest accuracy score of 89%.

Table 2: Summary of Results for Ransomware Dataset

	Macro Average			Weighted Average			
	Precision (%)	Recall (%)	F1-Score (%)	Precision (%)	Recall (%)	F1-Score (%)	Accuracy
KNN	79	72	73	86	86	85	86
RF	93	84	87	94	94	93	94
NB	84	75	76	91	87	87	87
DT	73	68	66	89	79	81	97
SVM	85	83	83	92	92	92	92
KNN + LDA	79	78	75	89	89	88	84
RF + LDA	83	76	74	89	85	85	85
NB + LDA	89	75	77	91	87	87	85
DT + LDA	94	95	94	98	97	97	79
SVM + LDA	79	77	75	89	83	85	83
KNN + PCA	81	74	75	89	84	85	89
RF + PCA	92	86	87	95	95	95	95
NB + PCA	91	88	87	96	95	95	95
DT + PCA	88	87	87	95	94	94	94
SVM + PCA	82	81	80	92	91	91	91

CONCLUSION

In conclusion, this work analyzed the effect of two pioneer dimensionality reduction techniques, namely Principal Component Analysis and Linear Discriminant Analysis on five (5) ML algorithms; KNN, SVM, DT, FT and NB. These dimensionality reduction techniques were applied on Ransomware Portable Executable Header Feature Dataset which is available on an online data repository with URL https://data.mendeley.com/datasets/p3v94dft2y/2. This dataset contains headers of 2157 binary executable samples comprising 1134 legitimate software (goodware) and 1023 ransomware, grouped into 25 ransomware families. The dataset was retrieved by extracting raw information of the PE header. By choosing to retain 25% of the components using PCA, the number of dependent features has been reduced to 25, whereas LDA reduced the dependent features also to 25. This reduced dataset was trained using five popular classifiers, Decision Tree classifier, Naive Bayes classifier, Random Forest classifier, K-Nearest Neighbor Classifier and SVM. From the results, it is observed that the performance of classifiers with PCA is better than that of with LDA. Also, Decision Tree and Random Forest classifiers outperform the other three algorithms without using dimensionality reduction as well as with both PCA and LDA.

REFERENCES

Alhawi O. M., Baldwin J, Dehghantanha A. (2018) Leveraging machine learning techniques for Windows ransomware network traffic detection. Cyber Threat Intell 70:93–106.

Almashhadani A., Kaiiali M., Sezer S., O'Kane P. (2019) A multi-classifier network-based crypto ransomware detection system: a case study of locky ransomware. *IEEE Access.* 7:47053–47067.

Alraizza, A.; Algarni, A.(2023): Ransomware Detection Using Machine Learning: A Survey. Big Data Cogn. Comput. 7, 143. https://doi.org/10.3390/bdcc7030143.

Al-rimy BAS, Maarof MA, Shaid SZM (2018) Ransomware threat success factors, taxonomy, and countermeasures: a survey and research directions. Comput Secur 74:144–166.

Alsaidi A. M., Wael M.S. Yafooz, Hashem Alolofi, Ghilan Al-Madhagy Taufiq-Hail, Abdel-Hamid M. Emara, Ahmed Abdel-Wahab (2022): Ransomware Detection using Machine and Deep Learning Approaches. International Journal of Advanced Computer Science and Applications, (IJACSA). Vol. 13, No. 11, 112-119.

Aslan, O., & Samet, R. (2020). A Comprehensive Review on Malware Detection Approaches. *IEEE Access*, *8*(January), 6249–6271. https://doi.org/10.1109/ACCESS.2019.2963724.

Borah M., T. Quertier, and S. Morucci, (2022) "AI-based malware and ransomware detection models.

Carlin D., O'kane, P. Sezer S., and Burgess J., (2018) "Detecting cryptomining using dynamic analysis," in *Proceedings of the 2018 16th Annual Conference on Privacy, Security and Trust (PST)*, pp. 1–6, Belfast, UK, August.

Cusack G, Michel O, Keller E (2018) Machine learning-based detection of ransomware using SDN, pp 1–6. https://doi.org/10.1145/3180465.3180467.

Gorment N.Z, Selamat A., Cheng L.K., and Krejcar O. (2023): Machine Learning Algorithm for Malware Detection: Taxonomy, Current Challenges, and Future Directions. IEEE Access. Vol. 11, pp. 141045-141089. Digital Object Identifier 10.1109/ACCESS.2023.3256979.

Jerlin, M. A., & Marimuthu, K. (2018). A New Malware Detection System Using Machine Learning Techniques for API Call Sequences. Journal of Applied Security Research, 13(1), 45–62. https://doi.org/10.1080/19361610.2018.1387734.

Khammas B.M (2022): Comparative analysis of various machine learning algorithms for ransomware detection. TELKOMNIKA Telecommunication Computing Electronics and Control Vol. 20, No. 1, February 2022, pp. 43~51. ISSN: 1693-6930, DOI: 10.12928/TELKOMNIKA.v20i1.18812.

Kok, S. H., Abdullah, A., & Jhanjhi, N. Z. (2019). Prevention of Crypto-Ransomware Using a Pre-Encryption Detection Algorithm. 1–15.

Reddy, G. T., Reddy, M. P. K., Lakshmanna, K., Kaluri, R., Rajput, D. S., Srivastava, G., & Baker, T. (2020). Analysis of Dimensionality Reduction Techniques on Big Data. *IEEE Access*, *8*, 54776–54788. https://doi.org/10.1109/ACCESS.2020.2980942.

Shah, N., & Farik, M. (2017). Ransomware - Threats Vulnerabilities And Recommendations. *International Journal of Scientific & Technology Research*, *6*(6), 307–309.

Umme Zahoora, Khan, A., Rajarajan, M. *et al.* (2022). Ransomware detection using deep learning based unsupervised feature extraction and a cost sensitive Pareto Ensemble classifier. *Sci Rep* **12**, 15647 https://doi.org/10.1038/s41598-022-19443-7

Urooj, U.; Al-rimy, B.A.S.; Zainal, A.; Ghaleb, F.A.; Rassam, M.A. (2022) Ransomware Detection Using the Dynamic Analysis and Machine Learning: A Survey and Research Directions. *Appl. Sci.*, *12*, 172. https://doi.org/10.3390/app12010172.