Indonesian Journal of Electrical Engineering and Informatics (IJEEI)
Vol. 10, No. 1, March 2022, pp. 102~111
ISSN: 2089-3272, DOI: 10.52549/ijeei.v10i1.3481

Time Efficiency on Computational Performance of PCA, FA and TSVD on Ransomware Detection

Benni Purnama[1], Deris Stiawan[2], Darmawijoyo Hanapi[3], Mohd Yazid Bin Idris[4], Sharipuddin[5], Nurul Afifah[6], Rahmat Budiarto[7]
[1,5]Department of Computer Engineering, Universitas Dinamika Bangsa Jambi, Indonesia
[2,6]Department of Computer Engineering, Universitas Sriwijaya, Indonesia [3]Faculty of Mathematics and Natural Science, Universitas Sriwijaya, Indonesia [4]School of Computing, Universiti Teknologi Malaysia, Malaysia
[7]College of Computer Science and IT, Albaha University, Albaha, Saudi Arabia

Article Info

Article history:

Received Oct 29, 2021
Revised Dec 12, 2021
Accepted Feb 21, 2022

Keyword:

Ransomware
Dimensionality Reduction
PCA
FA
TSVD

ABSTRACT

Ransomware is able to attack and take over access of the targeted user's computer. Then the hackers demand a ransom to restore the user's access rights. Ransomware detection process especially in big data has problems in term of computational processing time or detection speed. Thus, it requires a dimensionality reduction method for computational process efficiency. This research work investigates the efficiency of three dimensionality reduction methods, i.e.: Principal Component Analysis (PCA), Factor Analysis (FA) and Truncated Singular Value Decomposition (TSVD). Experimental results on CICAndMal2017 dataset show that PCA is the fastest and most significant method in the computational process with average detection time of 34.33s. Furthermore, result of accuracy, precision and recall also show that the PCA is superior compared to FA and TSVD.

Corresponding Author:

Deris Stiawan,
Department of Computer Engineering,
Universitas Sriwijaya, Indralaya, Indonesia,
Email: deris@unsri.ac.id

1. INTRODUCTION

Ransomware is very dangerous because it can attack a computer without the user's knowledge [1]. The purpose of ransomware is to profit by illegal means [2]. Smart devices including smartphones with their operating systems are also being targeted by hackers. This popularity has led to the rise of ransomware attacks in recent years. Thus, early detection technique against ransomware is required in order to make users' private data is secure [3]. Many research works on securing users' data have been carried out. The researchers use a huge dataset such as CICAndMal2017 dataset with a capacity of 30 GB [4]. With very large size of data, computing processes take very long time [5]. Therefore, a dimension reduction method is required to shorten the computing time by shrinking the dimensionality of the data [6]. Commonly used dimension reduction methods include: PCA, FA and TSVD [5]. PCA selects certain groups of the origin features. The features are transformed into the lower dimension structure. The best principal components then are selected. FA does not only reduce the dimensionality of the data, it is a useful approach to find latent variables which are not directly measured in a single variable but rather inferred from other variables in the dataset [7]. TSVD performs linear dimensionality reduction by means of truncated singular value decomposition (SVD) [8]. It works well with sparse data in which many of the row values are zero.

A number of studies have been conducted in in the field of dimensionality reduction in Ransomware Detection. One of the researchers conducting research on ransomware detection was labib et al. [9], who presented a major component analysis approach by introducing threshold values for detecting subject interference. The results showed that the accuracy performance achieved the best performance which is 100%. The study also presented a graphical approach to interpreting the results obtained from bi-plot implementation.

Woodhams et al [10] tried to implement a point of view in applying Heart Optical Mapping. The study appliesPCA in terms of dimensional reduction. Evidently, PCA is effective in eliminating video noise levels by producing clear, albeit incomplete, video performance from original videos. PCAs de-noise much more complete video noise than Median Filtering or Gaussian Blur, with no loss of detail and resolution. Bruno et al [11] have been proposed the PCA for reducing the number of feature dimensions. Features that have important information are required to describe a large set of data, thus enabling processing by a support vector machine (SVM). In addition, PCA creates new features of native data functions without having to lose information from datasets. Datti et al. [12] proposed PCA and Linear Data Analysis (LDA). In this study PCAand LDA were very helpful in improving performance in terms of accuracy and training time. The results of carried out experiments show that the LDA and PCA dimension reduction methods are able to produce good accuracy. Aburomman and Reaz [13] have been revealed that the use of PCA and LDA can improve performance in detection systems. Literally, the ensemble feature extraction method shows excellentand optimal performance in intrusion detection rates. Taguchi and Murakami [14] also proposed PCA as a dimensional reduction method for the process of identifying biomarkers of miRNA in the blood. The results ofsimulation experiments showed that PCA was able to reduce the features of the dataset, so that the computational process became shorter. Wibawa et al [15] applied KNN and PCA algorithms for FNA data in breast tumor classification. The PCA implementation is able to reduce dimensions to smaller so that it helps the computing process faster. Research using the KNN PCA algorithm also provides the best performance in term of accuracy rate of 97.36%. These results suggest that dimension reduction can improve computational performance. PCA is able to eliminate very significant correlations. Paukkeri et al [16] proposedPCA, Independent Component Analysis (ICA) and FA as dimensionality reduction for Spectral–Temporal Data. The experimental results showed that in order to extract features from distorted Spectral–Temporal data, the FA and ICA methods obtained the best andsignificant performance. Ruangpaisarn and Jaiyen [17] implemented TSVD that is able to reduce computing time.

Referring to some works on implementations of PCA, FA and TSVD in different application domains as discussed above, this work attempts to compare the three dimensional reduction methods in ransomware detection. The main goal is to investigate the impact of dimensionality reduction process that may help in speeding up detection time without losing the characteristics of the dataset.

The rest of the paper is organized as follows. Section 2 reviews some related works while Section 3 discusses the proposed method. Section 4 presents the experimental results along with discussion. Finally, Section 5 draws the conclusions and future works.

2. RELATED WORKS

Shingchern and Ming-Jen [18] used three different methods in the dimensionality reduction method, i.e.: PCA, FA and ICA in extracting undistorted items. The experimental results showed that the FA and ICA dimensional reductionmethods obtained higher accuracy results than PCA. So the study concluded that the FA and ICA dimensionalreduction methods are better than PCA. Fan et al., [5] showed that PCA is the best method of reducingthe dimensions of features because the latent factor in PCA greatly impacts most of the variances. Therefore, latent factors and additional variables are important parameters for reducing feature dimensions. Giovanni and Giorgio [19] introduced a new viewpoint in providing statistical interpretation of the traditional latent semantic analysis paradigm (LSA) that applies the TSVD dimensional reduction method. The results of the study showed that TSVD appeared in good performance in reducing features dimension. It can be observed that in statistical estimators derived from the LSA event relationship matrix by mapping probability distributions on Riemannian manifolds, TSVD is able to reduce the dimensions of features without changing the core information of the data. Sharipuddin et al. [20] proposed PCA in minimizing dataset features while maintaining as much variation as possible in the dataset, so that the information contained in the data does not change. The use of PCA for feature extraction on IDS also aims to improve validation performance in detection systems.

Zhang [21] proposed several new approaches by applying artificial neural network (ANN)-based models combined with PCA as dimensional reduction method in malware detection system. The results of experiments showed that PCA was significantly able to reduce the redundancy of features and the learning time. The outcome also proved that PCA is able to maintain data attributes, so as to minimize the impact in terms of data loss on 105,000 PDF documents. Harikumar and Kumaris [22] also performed the PCA method as one of the dimension reduction methods and Sparse Representation Classifier (SRC) as a classifier for the epilepsy risk levels classification from ECG signals. Theresults of experiments indicated that the reduction of dimensions was able to make the effectiveness of investigation time. PCA and SRC are able to make the process of classifying the risk level of epilepsy from electroencephalography brief and constant at certain time intervals whereas at other times, there is a structural change in quality value.

3. DATASET AND METHODOLOGY

Dataset in this study is taken from the University of New Brunswick (UNB) Dataset, i.e.: CICAndMal2017. The entries in HCICAndMal2017 dataset are in the form of PCAP format and grouped into two categories of normal and ransomware. Total data used as much as 15 GB ransomware and 15 GB benign [4]. This study aims to objectively evaluate three dimensionality reduction methods, i.e.: PCA, FA and TSVD to identify "Benign" and "Ransomware" for the low-dimensional representation of UNB dataset. The proposed methodology is displayed in Figure 1. In orderto visualize the dataset plot, the feature space has been reduced using a part of the dimensionality reduction provided by [4], [6] and [7].

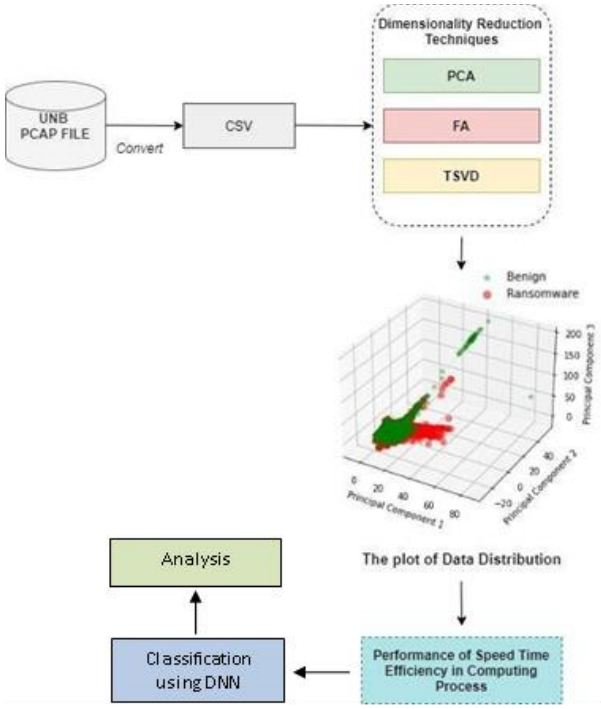


Figure 1. The Proposed Methodology

The log file is processed earlier into a feature vector text file that represents each PCAP file [23]. This text file is further converted into a CSV file. A representation of the amount of ransomware and benign data indicates the balance of data is shown in Figure 2 and Figure 3. Furthermore, the CSV data will be processed in the data preprocessing stage. There are many methods that can be used for dimension reduction such as feature selection method or that combine dimensions by calculating the average weight of correlated features, as depicted in Figure 4. Nevertheless, this paper will only focus on three methods, i.e.: PCA, FA and TSVD [24][8] [5].

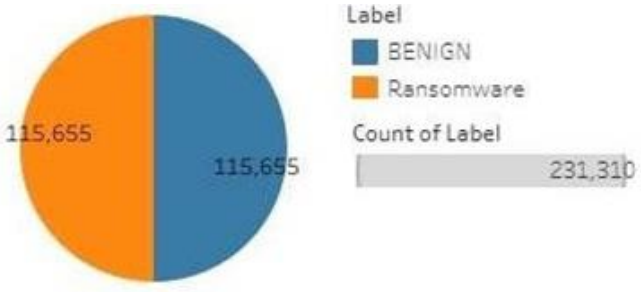


Figure 2. Distribution of Ransomware and Benign

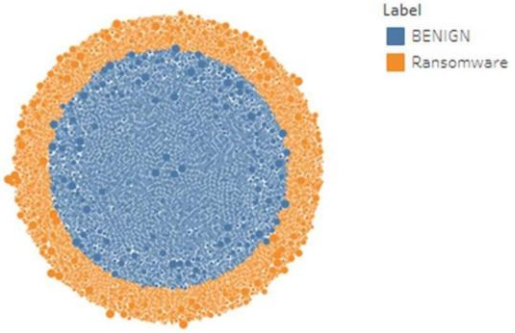


Figure 3. IP Source & Destination

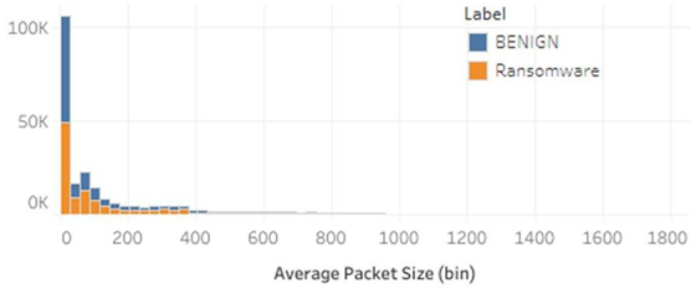


Figure 4. Average Packet Size

3.1. PCA

PCA is used to simplify data, by changing data linearly so that new data is more in optimal form. PCA is very useful in determining many minimum factors by taking into the maximum variance in the data for use in multivariate analysis. PCA converts high-dimensional datasets into lower data dimensions that still contain some data from high-dimensional datasets to the required features. PCA makes considerations with observation datasets A*1,* A*2,* A*3* ..., Ap with n dimensions that can be represented as a vector AT = [A*1,* A*2,…,* An] [25]. The PCA is capable to perform orthogonal transformations that projecting data into a subspace that minimizes the correlation of projection results. This subspace is referred to as the principal subspace. PCA is an unsupervised method where labels on training data are not used to conduct learning. In case of detection, PCA projection may not be optimal. PCA projection always tries to maintain all kinds of variations to the maximum regardless of the factors that give rise to these variations [26-29]. The stages of PCA algorithm are as follows. [6].

1. Variant-covariance matrix A is given by. (1)

$$\sum = [(A - \mu)(A - \mu)]T$$

2. Calculate the mean of each dimension of the dataset. (2)

$$Amean_{i} = \frac{1}{N}\sum_{j}^{N} = 1A_{(i,j)}$$

3. Calculate the covariance matrix of whole dataset (3)

$$C = \mathrm{X}\, x\mathrm{X}^{\mathrm{T}}$$

4. Compute eigenvectors and eigenvalues. (4)

$$CV = \lambda V$$

5. Parse eigenvectors by reducing eigenvalues and select the K eigenvectors with the largest eigenvalues to form a dimensional matrix (5)

$$PC = \mathrm{X}^{\mathrm{T}} * V$$

6. Convert samples to new sub-spaces using the Eigenvector matrix. (6)

$$PC_{\mathrm{S}} = \mathrm{X} * PC$$

3.2. FA

FA is a multivariate analysis code based on correlations between variables. Factor analysis is a statistical technique that can be used to provide a relatively simple description through the reduction of the number of variables called factors. FA is used to reducing data or summarize from old variables that are widely converted into a few new variables called factors, and still contain most of the information contained in the original variable [7]. The steps in determining the analysis of factors are to first formulate the problem and identify the original variable to which the factor will be analyzed. Then, the correlation matrix of variables anda factor analysis method are selected. Researchers determine the number of factors to be extracted from a lot of variables and rotation techniques will be used. Next, interpret the rotational resulting factor. The main purpose of FA is not only to reduce the dimensions of data, however, FA is also a useful approach to finding latent variables, which are not measured directly in one variable however, are inferred from other variables in the dataset. This latent variable is called the Factor Equation. The FA procedure is as follows.

1. Collect measurement data
2. Looking for correlation matrix

$$\mathrm{X} = \mu + LF + \mathrm{e} \tag{7}$$

3. Then, select the number of factors for inclusion
4. Extraction of a number of initial factors
5. Apply the rotation method to be used

$$C(\mathrm{X}) = LL' + \psi \tag{8}$$

6. Create a score factor for analysis

3.3. TSVD

Truncated Singular Value Decomposition (TSVD) is a reduction technique in linear dimensions. TSVD is a numerical computing technique that factorization of a zero matrix so that three zero matrices are obtained. One of the matrices obtained from the TSVD process will contain the singular values of the originalmatrix. In TSVD method can create a new matrix. After creating the new matrix, the next process can reconstruct the new matrix with smaller dimensions, but still similar to the original matrix. Once a new matrix is formed then apply the TSVD method to obtain a new representation matrix. The TSVD works well with sparse data where many line values are zero. In contrast, PCA works well with solid data. TSVD can also be used with solid data. Another difference is that factorization for TSVD is implemented on the data matrix while factorization for PCA is implemented on the covariance matrix [8]. Figure 5 illustrates the TVSD algorithm.

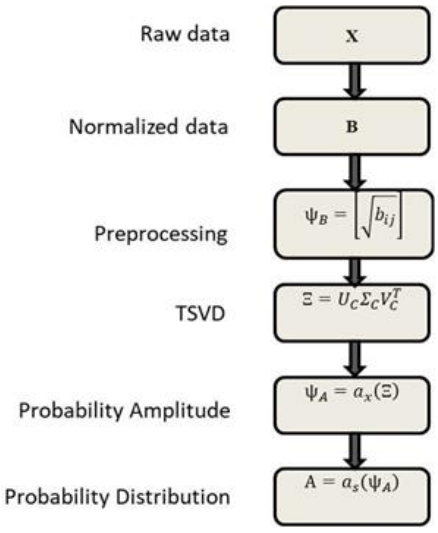


Figure 5. TSVD Algorithm

3.4. Classification

In this work, the results of feature extraction using PCA, FA, and TSVD will be evaluated using a Deep Neural Network. The aim is to determine the detection performance. Next, this work also measures the time processing for each dimensions reduction method on ransomware malware dataset. In addition, it uses 100 input delayer nodes, ReLU and Sigmoid for activation keys. Finally, the number of epochs for training is 100 epochs.

3.5. Experimental Setup

Table 1. Hardware and Software Specification

Category		Specification
Hardware	Processor	Intel Core i7, 3,1 GHz
	Operating System	Windows 10
	Memory	8GB RAM, 256 SSD
Software	Network Tools	Wireshark
	Python Programming	Python 3.2.7
	Data Analytics	Tableau
	DNN	Tensorflow
	Interface for Tensorflow	Keras

4. RESULTS AND DISCUSSION

This section presents dataset preparation, experimental details with PCA, FA, TSVD for feature selection, and last, experimental results and discussion.

4.1. Dataset Preparation

A high-dimensional dataset consisting in random vectors with low variances was generated to illustrate the algorithms performance. The quality of dimensionality reduction applied to the CICAndMal2017 UNB Dataset is exhibit by visualizing the plot of the data distribution using the PCA, FA and TSVD methods. The CICAndMal2017 UNB dataset consists of 85 features. One of the reasons is that the detection process is not optimal because of the large dimensions of the data. For more details of the features in the CICAnd Mal2017 dataset refer to Figure 6. The initial preparation is normalization, i.e.: removing features that have no value from the dataset and eliminating features that are not relevant to malware such as time, and IP address. The results of the preparation of this dataset will be used for feature extraction and detection processes.

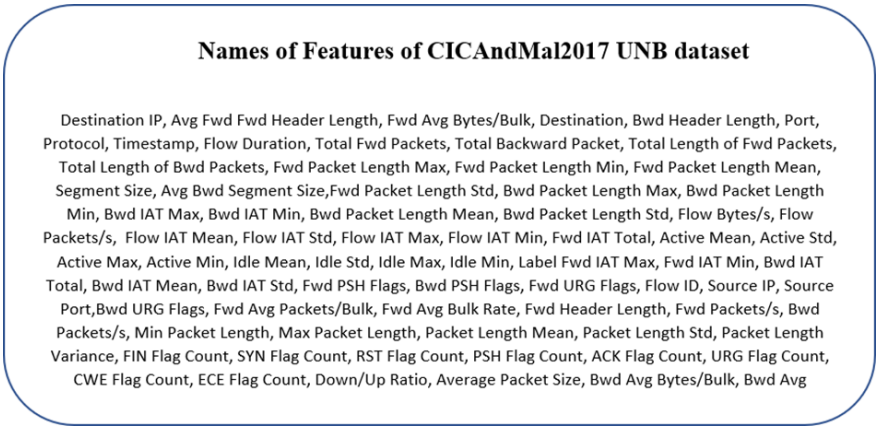


Figure 6. Features of CICAndMal2017 UNB dataset

4.2. Feature Extraction with PCA, FA and TSVD

At this stage, the dataset dimension is reduced into smaller degree. The smaller data dimension requires lesserprocessing time during the training, and in turn, improving the performance of the ransomware detection. Next, several scenarios are considered for the number of selected features, i.e.: 3, 6, 9, 12, 15, 18, 21, 24, 27, and 30. The results of this dimension reduction are values with a range of -1 to 1. Table 2 illustrates an example of the results of the dimension reduction using PCA, FA, and TSVD. The results from the features extraction will be used for malware detection using DNN.

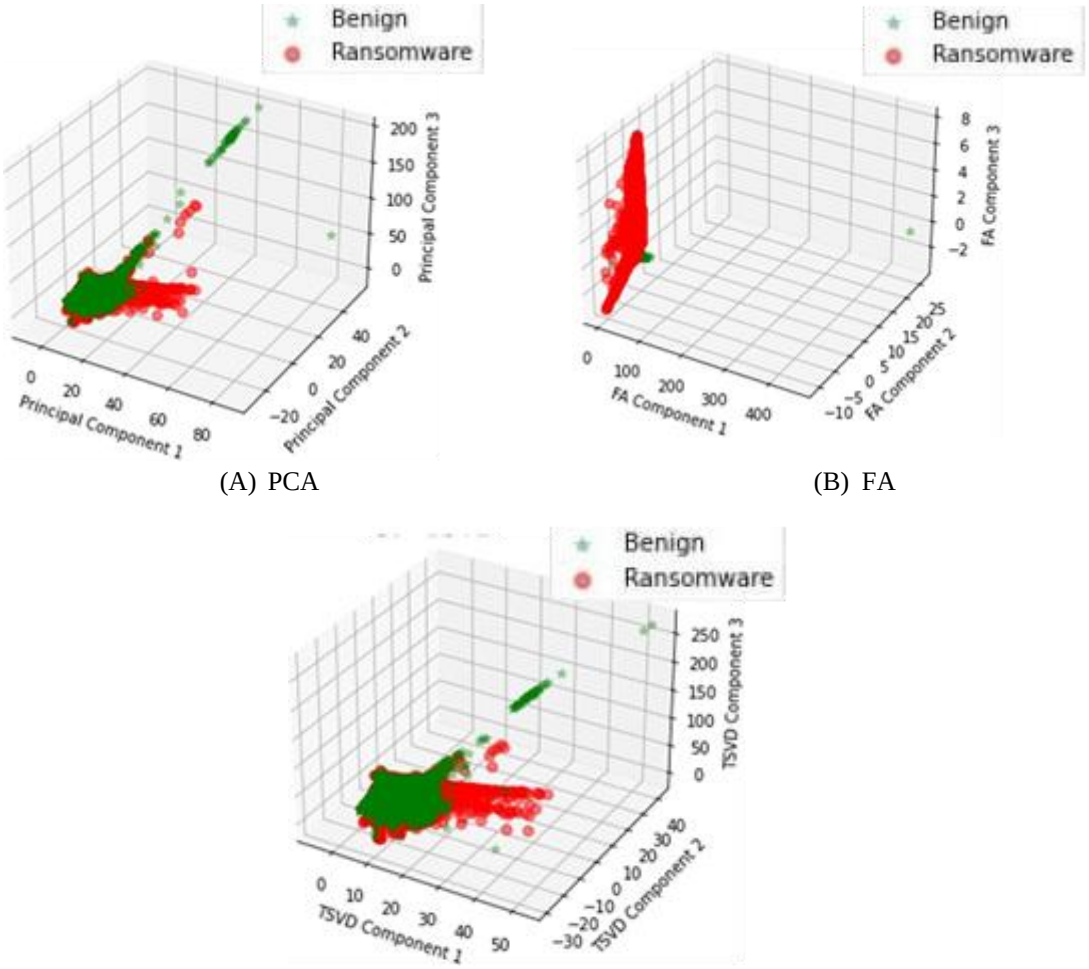
Table 2. Examples of Results for PCA, FA and TSVD

Features #	New Features subset
3	0.00306441, 0.03906215, -0.03321358,
6	-0.00189471, 0.06662967, 0.09061988, 0.10891438, 0.02286863, 0.01770335,
9	0.00306441, 0.03906215, -0.03321358, -0.01257658, -0.03450761, -0.02499266, 0.06350368, 0.05068012, -0.00189471,
12	0.03581673, 0.00306441, 0.03906215, -0.03321358, -0.01257658, -0.03450761, -0.02499266, 0.06350368, 0.05068012, -0.00189471, 0.06662967, 0.09061988,
15	0.06350368, 0.05068012, -0.00189471, 0.06662967, 0.09061988, 0.10891438, 0.02286863, 0.01770335 ,-0.03581673, 0.00306441, 0.03906215, -0.03321358, -0.01257658, -0.03450761, -0.02499266,
18	0.03906215, -0.03321358, -0.01257658, -0.03450761, -0.02499266, , -0.00189471, 0.06662967, 0.09061988, 0.10891438, 0.02286863, 0.03906215, -0.03321358, -0.01257658, -0.03450761, -0.02499266, 0.00189471, 0.06662967, 0.09061988,
21	-0.00189471, 0.06662967, 0.09061988, 0.10891438, 0.02286863, , 0.03906215, -0.03321358, -0.01257658, -0.03450761, 0.03906215, -0.03321358, -0.01257658, -0.03450761, -0.02499266, -0.02499266, 0.00189471, 0.06662967, 0.09061988, , -0.03321358, -0.01257658, , -0.03450761
24	0.03906215, -0.03321358, -0.01257658, -0.03450761, -0.02499266, 0.06350368, 0.05068012, -0.00189471, 0.06662967, 0.09061988, 0.03581673, 0.00306441, 0.03906215, -0.03321358, -0.01257658, -0.03450761, -0.02499266, 0.06350368, 0.05068012, -0.00189471, 0.06662967, 0.09061988, 0.03581673, 0.00306441
27	0.06662967, 0.09061988, 0.10891438, 0.02286863, 0.01770335 ,-0.03581673, 0.00306441, 0.03906215, -0.03321358, -0.01257658, -0.03450761, -0.02499266, 0.06350368, 0.05068012, -0.00189471, 0.06662967, 0.09061988, 0.10891438, 0.02286863, 0.01770335 ,-0.03581673, 0.00306441, 0.03906215, -0.03321358, -0.01257658, -0.03450761, -0.02499266
30	0.06350368, 0.05068012, -0.00189471, 0.06662967, 0.09061988, 0.10891438, 0.02286863, 0.01770335 ,-0.03581673, 0.00306441, 0.03906215, -0.03321358, -0.01257658, -0.03450761, -0.02499266, 0.06350368, 0.05068012, -0.00189471, 0.06662967, 0.09061988, 0.10891438, 0.02286863, 0.01770335 ,-0.03581673, 0.00306441, 0.03906215, -0.03321358, -0.01257658, -0.03450761, -0.02499266,

Figure 7 shows the scatter of the data resulting from dimension reduction using PCA, FA, and TSVD. Figure7A presents the scatter data for PCA, while Figure 7B shows the result of FA which has a smaller scatter of numerical results than the results of other methods. While the results of reduction with largest number of results shown in Figure 7C is for TSVD.

4.3. Experiment Results

To analyze the performance of ransomware detection using PCA, FA, and TSVD in improving classification performance, 3 (three) measurements are used, namely Precision, Recall, Accuracy. In the experiment, each featuresubset of PCA, FA, and TSVD was classified by DNN. Table 3 shows the results of the classification using DNN in term of accuracy, precision, and recall for each number of features of PCA, FA, and TSVD. The results show that the PCA method with DNN has higher accuracy results than other methods. Then the highest accuracy, precision, and recall are generated by using PCA with the feature numbers of 12 and 15.



(A) PCA (B) FA

(C) TSVD

Figure 7. Separation of Benign and Ransomware by features numbers of (A) PCA, (B) FA and (C) TSVD applied to their surrounding features

Table 3. The Result classification of the PCA, FA and TSVD using DNN

Features #	Acc	Prec	Rec	Acc	Prec	Rec	Acc	Prec	Rec
3	98.43	98.39	98.61	98.60	98.60	98.76	97.10	97.10	97.33
6	98.12	98.62	98.34	98.67	98.90	98.82	97.21	97.36	97.42
9	98.80	98.88	98.98	98.72	98.29	98.87	97.38	97.50	97.57
12	**98.93**	**98.96**	**98.91**	**98.79**	**98.65**	**98.68**	**97.85**	**97.86**	**97.89**
15	**98.83**	**98.90**	**98.96**	98.73	98.36	98.07	96.49	96.73	96.85
18	98.79	99.02	98.93	98.65	98.26	98.10	96.48	96.16	96.01
21	98.83	98.05	98.99	98.75	98.26	98.12	96.50	96.11	96.02
24	98.74	98.89	98.89	**98.77**	**98.42**	**98.17**	97.21	97.06	97.94
27	98.41	98.67	98.45	98.66	98.26	98.09	**97.63**	**97.20**	**97.05**
30	98.71	98.67	98.65	98.66	98.29	98.11	97.57	97.26	97.00

In the FA classification results, the highest results were for feature numbers of 12 and 24. The accuracy, precision, and recall results of FA were also better when compared to the TSVD method. While in TSVD, the best results were for feature numbers of 12 and 27. The experiments results of ransomware detection using the features selection resulted by PCA, FA, and TSVD are optimum. It is shown that the best performance results were achievedby PCA then FA, and then followed by the TSVD method.

Next is to measure the processing time for dimension reduction using PCA, FA, and TSVD on ransomwaremalware detection. The aim is to find out the best method of processing speed and detection accuracy results. Table 4 shows the time required to reduce the dimensions before being classified. Table 4 indicates the smaller the number of targeted features, the faster process is required to reconstruct the ransomware

dataset. This indication applies to all methods. The FA method is a method requires a little bit long time when compared to the PCA and TSVD methods.The PCA and TSVD methods have a fairly close time span between 26-40 seconds.

Table 4. Computing Process Time of the PCA, FA and TSVD for Ransomware Dataset

Features #	PCA	FA	TSVD
3	26.18s	102.25s	30.55s
6	31.36s	149.43s	33.83s
9	31.74s	189.37s	34.37s
12	34.38s	222.42s	34.55s
15	34.91s	151.22s	35.15s
18	35.04s	179.46s	36.84s
21	36.56s	336.40s	37.37s
24	36.68s	383.54s	38.95s
27	37.53s	223.52s	42.23s
30	38.92s	224.23s	40.43s
Average Reducing Time	**34.33s**	**216.18s**	**36.42s**

Figure 8 shows that PCA and TSVD get better performance in computing time on CICAndMal2017 ransomware dataset, i.e.: 34.33 seconds and 36.42 seconds, respectively, while FA is 216.18 seconds. It can be analyzed that PCA is the best linear dimension reduction method in this case. On the other hand, FA has very high computational time due to the factor of error variance.

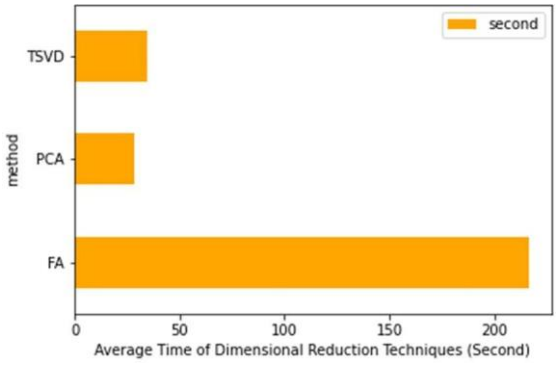


Figure 8. Comparison Graph of Dimensionality Reduction Results

5. CONCLUSION

In this research work the authors have investigated the efficiency of time in PCA, FA and TSVD dimensionality reduction methods implemented on ransomware dataset. The investigation result shows that PCA and TSVD methods get better performance in the computational processing time with 34.33 seconds and 36.42 seconds. On the other hand, FA are considered as slow with 216.18 seconds processing time, respectively. Furthermore, PCA and TSVD use total variance so that there is no information loss. As for the ransomware detection accuracy of PCA is superior to FA and TSVD. In future work, PCA or TSVD can be implemented in the process of classification and automatic detection of ransomware in real time fashion, or it can also create a tool to prevent ransomware.

ACKNOWLEDGMENTS

The authors thank Universitas Dinamika Bangsa for the research grant support under human resource development project with collaboration with COMNETS Lab at Faculty Computer Science, Universitas Sriwijaya.

REFERENCES

[1] O. T. Suryati and A. Budiono, “Impact Analysis of Malware Based on Call Network API With Heuristic Detection Method,” Int. J. Adv. Data Inf. Syst., vol. 1, no. 1, pp. 1–8, 2020, doi: 10.25008/ijadis.v1i1.176.

[2] M. K. Alzaylaee, S. Y. Yerima, and S. Sezer, "DL-Droid: Deep learning based android malware detection using real devices," Comput. Secur., vol. 89, no. November, 2020, doi: 10.1016/j.cose.2019.101663.
[3] S. I. Popoola, F. Sweetwilliams, S. N. John, and N. D. Academy, "Ransomware: Current Trend, Challenges, and Research Directions," no. October, 2017.
[4] A. H. Lashkari, A. F. Akadir, H. Gonzalez, K. F. Mbah, and A. A. Ghorbani, "Towards a network-based framework for android malware detection and characterization," Proc. - 2017 15th Annu. Conf. Privacy, Secur. Trust. PST 2017, no. December 2019, pp. 233–242, 2018, doi: 10.1109/PST.2017.00035.
[5] J. Fan, Q. Sun, W. Zhou, and Z. Zhu, "Principal Component Analysis for Big Data," Wiley StatsRef Stat. Ref. Online, pp. 1–13, 2018, doi: 10.1002/9781118445112.stat08122.
[6] I. T. Jollife and J. Cadima, "Principal component analysis: A review and recent developments," Philos. Trans. R. Soc. A Math. Phys. Eng. Sci., vol. 374, no. 2065, 2016, doi: 10.1098/rsta.2015.0202.
[7] R. W. Emerson, "Exploratory factor analysis," J. Vis. Impair. Blind., vol. 111, no. 3, pp. 301–302, 2017, doi: 10.1177/0145482x1711100313.
[8] M. Gavish, R. Schweiger, E. Rahmani, and E. Halperin, "ReFACTor: Practical Low-Rank Matrix Estimation Under Column- Sparsity," pp. 1–14, 2017, [Online]. Available: http://arxiv.org/abs/1705.07654.
[9] K. Labib and V. R. Vemuri, "An application of principal component analysis to the detection and visualization of computer network attacks," Ann. des Telecommun. Telecommun., vol. 61, no. 1–2, pp. 218–234, 2006, doi: 10.1007/BF03219975.
[10] L. Woodhams, D. Peters, and R. Pless, "Application of PCA to Cardiac Optical Mapping," 2017.
[11] B. L. Dalmazo, J. P. Vilela, P. Simoes, and M. Curado, "Expedite feature extraction for enhanced cloud anomaly detection," Proc. NOMS 2016 - 2016 IEEE/IFIP Netw. Oper. Manag. Symp., pp. 1215–1220, 2016, doi: 10.1109/NOMS.2016.7502990.
[12] R. Datti and S. Lakhina, "Performance Comparison of Features Reduction Techniques for Intrusion Detection System," vol. 8491, pp. 332–335, 2012.
[13] A. A. Aburomman and M. B. I. Reaz, "Ensemble of binary SVM classifiers based on PCA and LDA feature extraction for intrusion detection," Proc. 2016 IEEE Adv. Inf. Manag. Commun. Electron. Autom. Control Conf. IMCEC 2016, pp. 636–640, 2017, doi: 10.1109/IMCEC.2016.7867287.
[14] Y. H. Taguchi and Y. Murakami, "Principal Component Analysis Based Feature Extraction Approach to Identify Circulating microRNA Biomarkers," PLoS One, vol. 8, no. 6, 2013, doi: 10.1371/journal.pone.0066714.
[15] M. S. Wibawa and K. D. P. Novianti, "Reduksi fitur untuk optimalisasi klasifikasi tumor payudara berdasarkan data citra FNA," Konf. Nas. Sist. Inform., pp. 73–78, 2017.
[16] M. S. Paukkeri, I. Kivimäki, S. Tirunagari, E. Oja, and T. Honkela, "Effect of dimensionality reduction on different distance measures in document clustering," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 7064 LNCS, no. PART 3, pp. 167–176, 2011, doi: 10.1007/978-3-642-24965-5_19.
[17] Y. Ruangpaisarn and S. Jaiyen, "SEMG signal classification using SMO algorithm and singular value decomposition," Proc. - 2015 7th Int. Conf. Inf. Technol. Electr. Eng. Envisioning Trend Comput. Inf. Eng. ICITEE 2015, pp. 46–50, 2015, doi: 10.1109/ICITEED.2015.7408910.
[18] S. D. You and M. J. Hung, "Comparative study of dimensionality reduction techniques for spectral–temporal data," Inf., vol. 12, no. 1, pp. 1–12, 2021, doi: 10.3390/info12010001.
[19] G. Pilato and G. Vassallo, "TSVD as a Statistical Estimator in the Latent Semantic Analysis Paradigm," IEEE Trans. Emerg. Top. Comput., vol. 3, no. 2, pp. 185–192, 2015, doi: 10.1109/TETC.2014.2385594.
[20] Sharipuddin et al., "Features extraction on iot intrusion detection system using principal components analysis (Pca)," Int. Conf. Electr. Eng. Comput. Sci. Informatics, vol. 2020-October, pp. 114–118, 2020, doi: 10.23919/EECSI50503.2020.9251292.
[21] J. Zhang, "Machine Learning With Feature Selection Using Principal Component Analysis for Malware Detection: A Case Study," 2019, [Online]. Available: http://arxiv.org/abs/1902.03639.
[22] R. Harikumar and P. Sunil Kumar, "Analysis of singular value decomposition as a dimensionality reduction technique and sparse representation classifier as a post classifier for the classification of epilepsy risk levels from EEG signals," J. Chem. Pharm. Sci., vol. 8, no. 2, pp. 191–194, 2015.
[23] S. Folek, "Ransomware Analysis and Detection Systems," Edinburgh Napier Univ. Degree BEng Comput. Secur. Forensics, no. November, 2016.
[24] S. Sehgal, H. Singh, M. Agarwal, V. Bhasker, and C. Engineering, Data Analysis Using Principal Component Analysis. .
[25] S. B. Mohammed and R. G. M. Helali, "Usage of Principal Component Analysis (PCA) in AI Applications," Int. J. Eng. Res. Technol., vol. 5, no. 12, pp. 372–375, 2016.
[26] Y. Hamid, M. Sugumaran, and L. Journaux, "A fusion of feature extraction and feature selection technique for network intrusion detection," Int. J. Secur. its Appl., vol. 10, no. 8, pp. 151–158, 2016, doi: 10.14257/ijsia.2016.10.8.13.
[27] [1] S. Sharipuddin et al., "Features Extraction on IoT Intrusion Detection System Using Principal Components Analysis (PCA)," Proc. EECSI 2020 - 1-2 Oct. 2020, pp. 114–118, 2020.
[28] [2] S. Sharipuddin et al., "Intrusion detection with deep learning on internet of things heterogeneous network," IAES Int. J. Artif. Intell., vol. 10, no. 3, p. 735, 2021, doi: 10.11591/ijai.v10.i3.pp735-742.
[29] [3] S. Sharipuddin et al., "Enhanced Deep Learning Intrusion Detection in IoT Heterogeneous Network with Feature Extraction," Int. J. Electr. Eng. Informatics, vol. 9, no. 3, pp. 747–755, 2021, doi: 10.52549/ijeei.v9i3.3134.