T.C.

ALTINBAŞ UNIVERSITY

Electrical and Computer Engineering

AUTOMATIC MALWARE DETECTION USING DATA MINING TECHNIQUES BASED ON POWER SPECTRAL DENSITY (PSD)

YASEEN AHMED ALSUMAIDAEE

Master's Thesis

Supervisor

Dr. Sefer KURNAZ

Istanbul, (2019)

AUTOMATIC MALWARE DETECTION USING DATA MINING TECHNIQUES BASED ON POWER SPECTRAL DENSITY (PSD)

by

Yaseen AL Sumaidaee

Electrical and Computer Engineering

Submitted to the Graduate School of Science and Engineering

in partial fulfillment of the requirements for the degree of

Master of Science

ALTINBAŞ ÜNİVERSİTESİ

[2019]

Bu çalışma tarafımızca incelenmiş olup, kapsam ve kalite açısından Yüksek Lisans tezi olmaya yeterli bulunmuştur..

__________________________________	__________________________________
Akademik Ünvan Ad SOYAD	Akademik Ünvan Ad SOYAD
Eş Danışman	Danışman

İnceleme Komitesi Üyeleri (İlk isim jüri başkanına, ikinci isim tez danışmanına aittir.)

Akademik Ünvan Ad SOYAD	Fakülte, Üniversite	___________________
Akademik Ünvan Ad SOYAD	Fakülte, Üniversite	___________________
Akademik Ünvan Ad SOYAD	Fakülte, Üniversite	___________________
Akademik Ünvan Ad SOYAD	Fakülte, Üniversite	___________________
Akademik Ünvan Ad SOYAD	Fakülte, Üniversite	___________________

Bu çalışma bir .... tezinin tüm gerekli şartlarını taşımaktadır..

____

Akademik Ünvan Ad SOYAD

Bölüm Başkanı

____

Akademik Ünvan Ad SOYAD

Enstitü Müdürü

Fen Bilimleri Enstitüsü onayı: ____/____/____

DEDICATION

Give my humble effort to:

Which he has never forgotten, and whose age he has enriched for our upbringing and My beloved father may Allah have mercy on him Who supported me and accompanied me with my dear mother, to prolong her life and save her My brothers and sisters enlighten my way

All my friends who stood with me

ACKNOWLEDGEMENTS

I would like to thank

Asst.Prof Dr. Safar kurnaz supervisor for his help and assistance throughout my work and My friend and colleague also to Dr. Ahmed Karim that helped me, All of me have been supported by my close friends To them all thanks and appreciation.

ÖZET

GÜÇ SPEKTRAL YOĞUNLUĞUNA (PSD) DAYALI VERI MADENCILIĞI TEKNIKLERI KULLANARAK OTOMATIK KÖTÜ AMAÇLI YAZILIM TESPITI

[Alsumaidaee, Yaseen]

Yüksek Lisans, Elektrik ve Bilgisayar Mühendisliği, Altınbaş Üniversitesi,

Danışman: Sefer KORNAZ

Tarih: Mart, 2019

Sayfa Sayısı: 60

Kötü amaçlı bir yazılım, orijinal bir yazılımın görünümünün altındaki sürecine titizlikle ulaşan bir yazılımdır. Klasik yöntemler, bu yazılımın, imzaları sunulmayan yeni ve gizli örneklere küçük riskler gösterdiğini göstermek için imzalar uygulamaktadır. Kötü amaçlı yazılım incelemesinin vurgusu, imza tasarımları uygulamasından, bu kötü amaçlı yazılımların gösterdiği kötü niyetli davranışı sınıflandırmaya kadar kararsızdır. Kötü amaçlı yazılımları mekanik olarak etkili bir şekilde fark etmek için sayısız veri madenciliği yöntemi önerildi. Bu tezde, çeşitli veri madenciliği yöntemlerini uygulayarak kötü amaçlı yazılım veri kümesinin özelliklerini ve PSD'nin gizli verimini çıkarmak için Güç Spektral Yoğunluğu (PSD) uygulanmıştır: Destek Vektör Makinesi (SVM), Radyal Temel Ağ (RBF) ve çok katmanlı perceptron (MLP). Bu teknikler, bu alandaki yaygın araştırmalarla karşılaştırıldığında kayda değer sonuçlar vermiştir.

Anahtar kelimeler: Kötü amaçlı yazılım, Veri madenciliği, Güç spektral yoğunluğu, bilgisayar güvenliği.

ABSTRACT

AUTOMATIC MALWARE DETECTION USING DATA MINING TECHNIQUES BASED ON POWER SPECTRAL DENSITY

(PSD)

[Alsumaidaee, Yaseen]

MS, Electrical and Computer Engineering, Istanbul Altinbas University,

Supervisor: Sefer KORNAZ

Co-Supervisor:

Date: March 2019

Pages: 60

A malware is a software that furtively achieves its process below the appearance of a genuine software. classic methods apply signatures to distinguish these software’s denote tiny risk to new and hidden examples whose signatures are not offered. The emphasis of malware investigation is instable from applying signature designs to classifying the malicious conduct showed by these malwares. Numerous data mining methods proposed to notice malware mechanically in effectual face. In this thesis, Power Spectral Density (PSD) applied to extract the features of malware dataset and the yield of PSD confidential applying several of data mining methods: Support Vector Machine (SVM), Radial Basis Network (RBF) and multi-layer perceptron (MLP). These techniques presented remarkable results when compared with common researches in this field.

Keywords: Malware, Data mining, Power spectral density, computer security.

TABLE OF CONTENT

PAGES

TABLE LIST

PAGES

LIST OF FIGURES

PAGES

LIST OF ABBREVIATIONS

Firstly:Support Vector Machine : SVM

Secondly:Neural Network : NN

Thirdly:Radial Basis Function : RBF

1. INTRODUCTION

Over the previous couple of years, cyber danger scenery has altered melodramatically and removed from economically interested attacks to targeted attacks, especially Advanced Persistent Threats. Starting from the year 2010 with Operation Aurora, we have witnessed increasing number of such targeted attacks including Stuxnet, Duqu, Flame, Red October, Snake, etc. [1]. This new class of attacks become top priority cyber risks for governments and commercial entities because of its sophistication in terms of tools and techniques employed and well-funded and skilled threat actors. In targeted attacks, efficient workers goal exact objects obstinately with tall motivation, evades security fortifications in home, employs progressive gears and strategies, maintains long-time attendance in board setting and operates sluggish and furtive to evade discovery [2].

Malware theatres an energetic part in achievement of a beleaguered bout and is working nearly in each stage of a bout lifespan pending the operator's goal is realized. It carries out extensive variety of errands counting cooperating model, mounting freedoms, upholding attendance, exfiltrating data, interactive with the workers over knowledge and switch waiters, loud out instructions, etc. Smooth however these errands are not odd to beleaguered malware only and also traditional malware could carry out most of these tasks throughout its execution, beleaguered malware is still predictable to entertainment different than the classical malware because of its stealthy nature.

Informal but too actual effectual way to disclose malwares performance when it contaminates an organization is running it in a controlled atmosphere and detention all the vicissitudes on the system and net throughout the examination procedure. This analysis method is called dynamic analysis and quick insight into malware can be gained by running it in a dynamic analysis sandbox for a very short time (usually 3 minutes). However, it does not work well for all kinds of malware, because there are some malwares that could detect it is running in a sandbox and stop or delay its execution, or could not complete all its tasks within the analysis period, or try to deceive analyst by doing nothing malicious or suspicious. Even with its all limitations dynamic analysis is extensively used in malware analysis field because huge number of new malwares are discovered every day and they are needed to be analyzed in a wild and automatic method. In our thesis work, we used a modified version of a popular open-source automatic malware examination organization, Cuckoo Sandbox, for capturing malware behavior [3].

1.1 PROBLEM STATEMENT

This study solves and explain malware attacks problem which analysis the previous studies and determine the weakness of these studies and try to fixed and presented new approaches in investigating this problem, there is number of points that will investigate in this research:

- How to select the best classification method for malware attacks problem?
- what is the best feature extraction method that applied to malware problem and presented best features that let us to obtain remarkable results?
- which data mining techniques presented best results when combined with feature extraction methods.

1.2 THESIS CONTRIBUTIONS

In this thesis, number of contributions are presented which they not presented and explained in previous studies as illustrated in below points:

- Presented power spectral density as feature extractor in malware classification problem.
- Combine power spectral density with several data mining techniques which mean new techniques for solving malware problem.
- Remarkable results are obtained and compared with common studies and researches presented in this field.

2. OVERVIEW

2.1 VICTIMS AND AIMS OF MALWARE ATTACKS

3 most mutual goal of DDoS can be recorded as below [4];

- E-commerce: Internet is the key component of many trades. Millions of clients are buying goods/amenities each day by their mobile devices or computers. Agreeing to JP Morgan, 2011 e-commerce income is 680 billion dollars. Customers identical shopping online since e-commerce websites are secure, responsive and always available. This big market attracted criminals to see the big opportunities. Online shopping companies are victimized mass attacks. Hackers obligate facts stealing, coercion, individuality robbery and deception. Malware attacks can cause millions of dollars loose, if service is gone or slowed. Even though you can get your services up in a short time, your customers think that your website is not secure. Rendering to review by the Ponemon Institute, the average entire rate of single information opening was extra than 7.2 million dollars in 2011.

- Financial Services: Nowadays, a person can use online banking systems, global money transfers or payment processes over internet. People can make these activities in anywhere, anytime and any devices. They expect their information to be secure and the service is reliable and quick. Malware attacks can have very big loose of money, if these services got slow or interrupted. Especially the attacks whose target's transaction processes or big trade systems can cause catastrophic affects. Many banks and stock exchanges, including Bank of America, New York stock exchanges, reported that they have been under Malware attack.

- Online Gaming: It is very big area with millions of players who can play so many games such as gambling and video games. People can play these games on different platforms; PCs, Xbox, PlayStation. Performance and availability are the key elements of online gaming. The activist hacker group Anonymous directed Malware attack to Sony PlayStation network in 2010.

2.2 STATIC MALWARE ANALYSIS

Static analysis is the procedure of investigative a binary file exclusive of executing it in order to control if it is malicious or not. Mining static topographies from a binary file to achieve malware recognition is approved out by static analysis gears or methods.

[6] were amongst the primary who presented static topographies for malware discovery by disbursing numerous dissimilar classifiers. They showed experimentations founded on 3 dissimilar kinds of static topographies counting Portable Executable (PE), cords and byte order n-grams. After PE shot of a binary, they removed active connection collection info and built 3 dissimilar topographies courses on behalf of if a DLL was applied, if an exact function confidential a DLL was named, and total of sole purpose noises secret all DLL. Prearranged cords confidential a two was removed and applied as a feature. They too rehabilitated binary files into hexadecimal cyphers and applied byte orders of cyphers as n-gram topographies. depend on these 3 topographies, dissimilar classifiers are working to categorize new binaries as hateful or not. Though cord topographies produced the uppermost correctness, finest discovery degree was attained by applying byte order n-grams.

Their exertion enthused and fortified others to stab alike methods for malware recognition. [7] accepted and improved the byte order n-grams method. They attained healthier consequences in distinguishing hateful binaries via classifiers counting SVM, DT and NN of them. [8] removed API noises from PE shot of a binary file similar what [6] did in their effort and applied them as topographies for categorizing a binary as malicious or clear.

Additional progressive static analysis and opposite engineering job to excerpt topographies from binaries was approved out in [9] likened to additional the whole thing mentioned here. Two was primary undid and then distance and frequency of purpose designations were removed. Created on the function designation distance features, they perform malware classification between different malware families and their results suggested that function name length is significant as a feature in distinguishing malware families.

Although static features extensively-applied to distinguish or recognize malware via data mining, there be some limits. researchers shoulder that the malware is unloaded, or not encoded and static topographies can be mined correct absent from the binary. Though, it is actual shared for a malware to be crowded or encoded and in approximately cases, it is not likely to completely unload or decrypt malware. There too be wide diversity of mystification methods that can frustrate the entire procedure [10].

2.3 DYNAMIC MALWARE ANALYSIS

Basic Dynamic Analysis suggests and proposals the chance of challenging, performing and consecutively hateful code and investigative on the scheme to checkered its conduct, procedures and possibly remove the contagion. Though, it must be specified that it is vital to system a computer-generated setting or computer-generated workroom that will let a investigator to education the performed malware deprived of harmful the real info scheme or network. Smooth however characteristic active analysis is a share of the malware analysis, it consumes approximately disadvantages, so the progressive technique of dynamic analysis is obligatory.

Advanced Dynamic Analysis employments a disassembler to inspect the interior disorder and the events of a performed malicious file. This technique delivers a method to obtain additional full information from a hateful package [11],[12]. Alike approaches are valued for procurement info. There are 2 approaches for dynamic malware examination that can be presented:

- Investigative the difference among stated states: On this time, there are 2 conditions. The primary national is the contagion of the malware and the national afterward the contagion. It is vital to distinguish how the info scheme was in the primary national to be talented to excerpt data about the malware though it is consecutively. Lastly, an account of the conditions likening apiece additional of the performances of the malware are obtainable.

- Monitoring consecutively doings and facilities: Where, hateful programs performed are experiential. Additional particulars what an investigator inspects, and discoveries finished a dynamic analysis process.

2.5. Performance Analysis of Detection Mechanism

To evaluate performance of malware attack detection mechanisms, true positive rate (TPR) and false positive rate (FPR) are used.

$$\text{TPR} = \frac{TP}{(TP + FN)} \tag{2.1}$$

TP is the number of samples those correctly defined as attack. FN is the number of normal traffic those defined as attacks.

$$\text{FPR} = \frac{FP}{(FP + TN)} \tag{2.2}$$

FP is the number of attack samples those detected as attacks. TN is the number of samples those correctly defined as legitimate traffic. These parameters used to evaluate the performance of data mining techniques that presented in section 3.

2.4 DATA MINING

Data mining, also termed knowledge discovery in databases, in computer science, the procedure of learning stimulating and valuable forms and relations in huge dimensions of data. The arena associations tools from statistics and machine learning with database supervision to investigate huge numerical groups, recognized as data sets. Data mining is extensively applied in commercial, science study, and government safety (detection of network attacks and terrorists).

The term "data mining" is in detail a contradiction, because the objective is the mining of patterns and information from huge quantities of facts, not the selection of feature itself. [13] It also is a slogan [14] and is regularly used to any procedure of important feature or data processing, selection, warehousing, examination, and indicators) as well as any submission of computer decision provision method, counting machine learning and commercial intelligence.

The real data mining job is the semi-automatic or automatic investigation of huge amounts of feature to select earlier unidentified, stimulating forms like sets of data records, rare records, and dependences. This typically includes applying record methods like spatial indices. These forms can then be seen as a type of rapid of the input feature, and may be applied in more investigation or, for instance, in data mining and prognostic analytics. For instance, the data mining stage force recognize several sets in the feature, which can then be applied to get extra true prediction outcomes by a decision support method. Neither the statistics collection, data groundwork, nor outcome clarification and recording is portion of the data mining stage, but do go to the general KDD procedure as extra stages. Generally the data mining techniques divided into 3 parts: The techniques are classified as a supervised, semi-supervised and unsupervised learning methods. These techniques are organized in Figure 2.1.

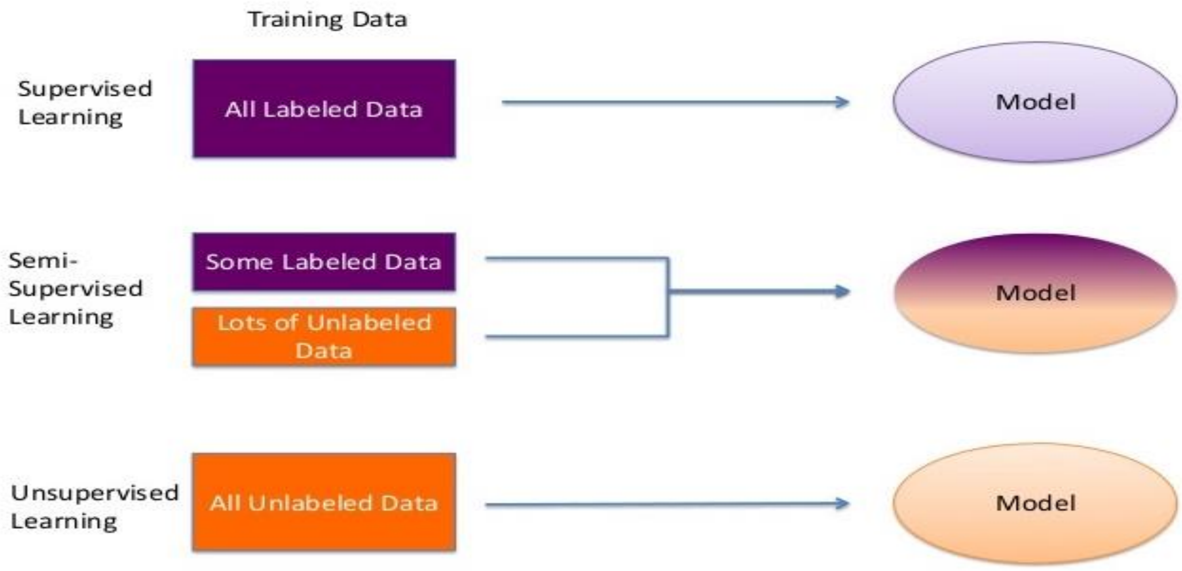


Figure 2.1: Data mining techniques categories

2.4.1 Supervised Learning

Supervised learning is naturally prepared in the background of cataloguing, when we poverty to map input to yield labels, or regression, when we need to map input to a continuous yield. Mutual techniques in supervised training contain logistic regression, NB, SVM, ANN, and RF. In mutually regression and classification, the objective is to discover exact relations or construction in the input feature that permit us to efficiently yield exact output label. Note that “accurate” output is selected completely from the learning features, so although we do have a crushed fact that our system will shoulder is exact, it is not to say that information labels are continuously exact in actual states. Noisy, or inappropriate, data labels will obviously decrease the efficiency of your method.

When showing supervised learning, the chief thoughts are system complexity, and the bias-variance compromise. Letter that together of these are unified.

system complexity denotes to the complexity of the purpose you are trying to learn—comparable to the point of a polynomial. The good stage of system complexity is usually selected by the wildlife of your training features. If you have a lesser amount of feature, or if your feature is not regularly feast through dissimilar likely situations, you must opt for a low-complexity system. This is because a high-complexity system will over fit if applied on a minor amount of feature facts. Overfitting denotes to learning a purpose that fits your learning feature actual glowing, but does not simplify to other feature facts—furthermore, you are severely learning to yield your learning feature without learning the real tendency or organization in the feature that mains to this yield. Visualize annoying to appropriate a line between 2 facts. In model, you can apply a function of any grade, but in practice, you would stingily add complexity, and go with a linear function.

The bias-variance use also tells to system simplification. In any system, there is a equilibrium between prejudice, which is the continuous error term, and modification, which is the quantity by which the error might differ between dissimilar training groups. Consequently, great bias and small modification would be a system that is reliably incorrect 20% of the period, while a small bias and great modification system would be a system that can be mistaken wherever from 5%-50% of the period, contingent on the feature applıed to train it. Note that bias and modification naturally change in conflicting instructions of each additional; growing bias will typically chief to minor modification, and wickedness versa. When creating your system, your exact problem and the countryside of your feature should authorization you to create an learned decision on where to reduction on the bias-variance field. Commonly, growing bias outcomes in systems with comparatively guaranteed standard stages of accuracy, which may be serious in sure errands. Moreover, in instruction to yield system that simplify well, the modification of your system would scale through the extent and complexity of your training feature—minor, humble features-sets must commonly be learned with minor-modification systems, and huge, complex feature-sets will frequently need major- modification systems to completely learn the organization of the feature.

The block diagram that presented in Figure 2.2. shows the Supervised learning technique [15],[16],[17].

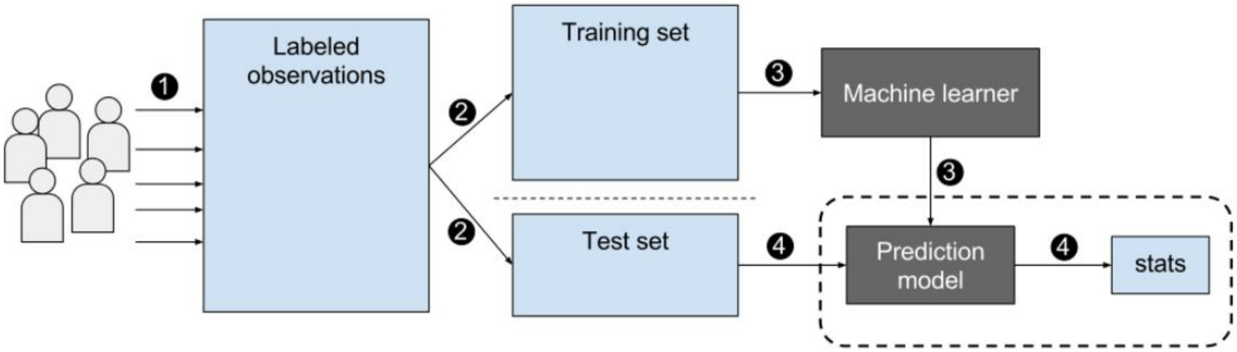


Figure 2.2: Simple Supervised Learning Structure [17].

2.4.2 Semi-Supervised Learning

Semi-supervised machine learning is a mixture of supervised and unsupervised learning techniques.

With extra mutual supervised machine learning techniques, you train a machine learning techniques on a labeled data in which individually greatest contains the output facts. This permits the techniques to infer patterns and recognize relations between your goal mutable and the break of the data depend on data it already has. In difference, unsupervised machine learning techniques train from a data without the output facts. In semi-supervised training, a techniques trained from a data that contains together labeled and unlabeled data, typically frequently unlabeled.

When you don’t have sufficient labeled data to yield suitable accuracy model and you don’t have the capability or incomes to grow extra, you can use semi-supervised methods to increase the number of your training dataset [18],[19]. The basic structure of semi-supervised learning presented in Figure 2.3.

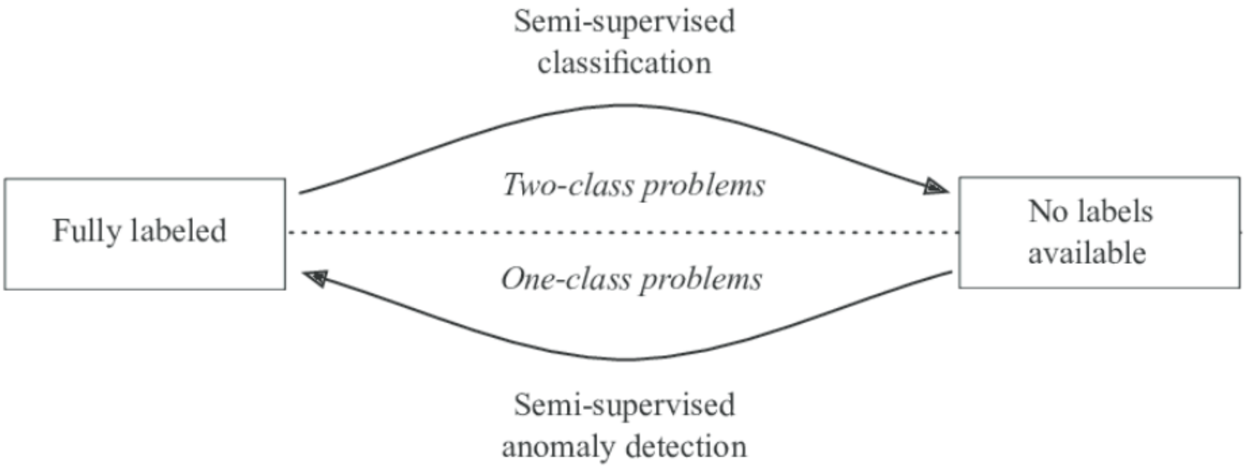


Figure 2.3: Semi-Supervised Learning Structure [22].

2.4.3 Unsupervised Learning

Unsupervised machine learning algorithms infer patterns after a dataset deprived of orientation to recognized, or labeled, outcomes. Different supervised machine learning, unsupervised machine learning approaches cannot be straight used to a reversion or a cataloguing problematic because you take no impression what the standards for the production data strength be, making it intolerable for you to train the technique the way you usually would. Unsupervised learning can in its place be applied for learning the fundamental construction of the facts.

Unsupervised machine learning purports to expose before unidentified designs in information, but greatest of the period these designs are deprived estimates of what supervised machine learning can realize. Moreover, meanwhile you don’t know what the consequences must be, here is no method to control how precise they are, creation supervised machine learning additional appropriate to practical difficulties.

The greatest period to apply unsupervised ML is when you don't have information on wanted consequences, similar decisive a board marketplace for a completely novel creation that your commercial has not ever vended beforehand. Though, if you are annoying to become a healthier sympathetic of your current customer dishonorable, supervised learning is the best method [20],[21]. The structure of unsupervised learning presented in Figure 2.10.

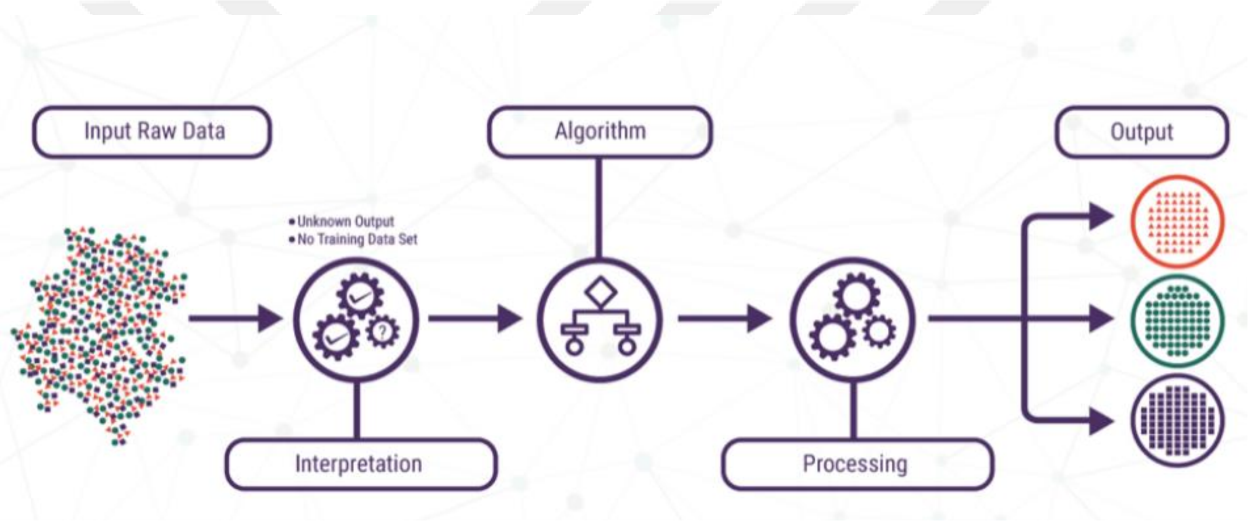


Figure 2.4: Unsupervised Learning Structure [22].

The techniques of these kinds of data mining techniques are presented in Figure 2.5 and will explained in detail form in chapter 3.

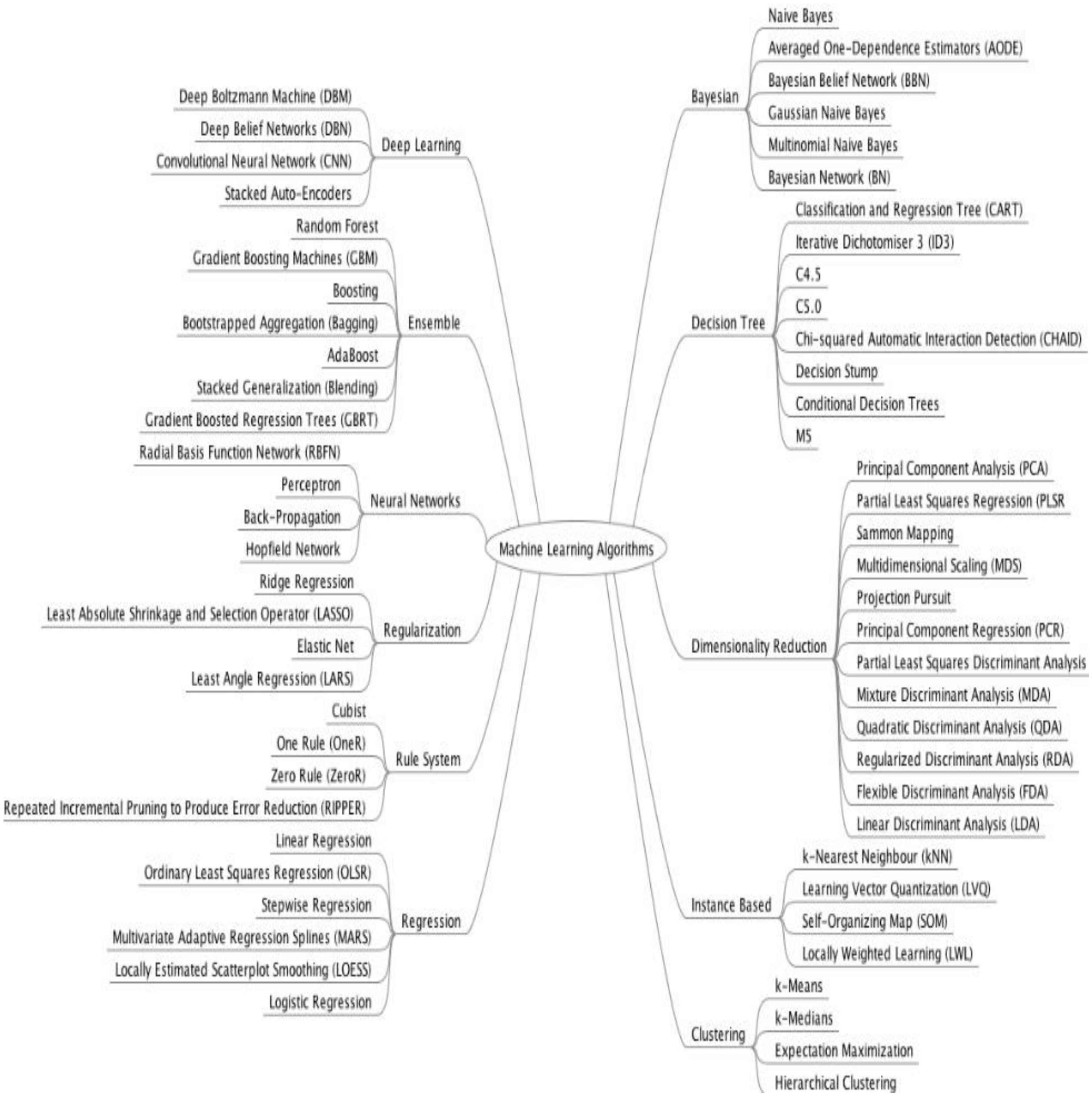

Machine Learning Algorithms
Deep Learning
Deep Boltzmann Machine (DBM)
Deep Belief Networks (DBN)
Convolutional Neural Network (CNN)
Stacked Auto-Encoders
Ensemble
Random Forest
Gradient Boosting Machines (GBM)
Boosting
Bootstrapped Aggregation (Bagging)
AdaBoost
Stacked Generalization (Blending)
Gradient Boosted Regression Trees (GBRT)
Neural Networks
Radial Basis Function Network (RBFN)
Perceptron
Back-Propagation
Hopfield Network
Regularization
Ridge Regression
Least Absolute Shrinkage and Selection Operator (LASSO)
Elastic Net
Least Angle Regression (LARS)
Rule System
Cubist
One Rule (OneR)
Zero Rule (ZeroR)
Repeated Incremental Pruning to Produce Error Reduction (RIPPER)
Regression
Linear Regression
Ordinary Least Squares Regression (OLSR)
Stepwise Regression
Multivariate Adaptive Regression Splines (MARS)
Locally Estimated Scatterplot Smoothing (LOESS)
Logistic Regression
Bayesian
Naive Bayes
Averaged One-Dependence Estimators (AODE)
Bayesian Belief Network (BBN)
Gaussian Naive Bayes
Multinomial Naive Bayes
Bayesian Network (BN)
Decision Tree
Classification and Regression Tree (CART)
Iterative Dichotomiser 3 (ID3)
C4.5
C5.0
Chi-squared Automatic Interaction Detection (CHAID)
Decision Stump
Conditional Decision Trees
M5
Dimensionality Reduction
Principal Component Analysis (PCA)
Partial Least Squares Regression (PLSR
Sammon Mapping
Multidimensional Scaling (MDS)
Projection Pursuit
Principal Component Regression (PCR)
Partial Least Squares Discriminant Analysis
Mixture Discriminant Analysis (MDA)
Quadratic Discriminant Analysis (QDA)
Regularized Discriminant Analysis (RDA)
Flexible Discriminant Analysis (FDA)
Linear Discriminant Analysis (LDA)
Instance Based
k-Nearest Neighbour (kNN)
Learning Vector Quantization (LVQ)
Self-Organizing Map (SOM)
Locally Weighted Learning (LWL)
Clustering
k-Means
k-Medians
Expectation Maximization
Hierarchical Clustering

Figure 2.5: Machine Learning Structure [23].

2.5 RELATED WORK

Malware attacks have been very hot topic in last 10 years for researchers. Individually technique has its own compensations and difficulties. Moreover, each research group compose their own dataset and they apply their method on their dataset. In that case, there is no possibility to work on those datasets for other working groups to compare their results.

Sun et al. proposed a method to count number of SYN-FIN packets difference. Under normal conditions, each TCP connection which completes the three-way handshake starts with a SYN packet, and finishes with a FIN or RST packet. In this case, there should be equal number of SYN and FIN packets in a long run. If there is inequality of these packets, like SYN packets are more than FIN packets, it means that there is attack going on [24].

Nashat et al stated that the difference between SYN packets and SYN/ACK packets can be good way to detect an attack. During normal 3-way handshake process, SYN packets arrive and SYN/ACK packets transmit. In this case, there should be similar number of SYN packets and SYN/ACK packets in a long run. However, during an attack, there will be so many SYN packets and victim will not be able to send SYN/ACK packets back [25].

Malware attacks are usually done by packets with spoofed IP addresses. In this case, entropy of source IP addresses seem to be logical feature to detect Malware attacks [26]. Also, the target of attacks is usually only one victim and it is expected that entropy of destination IP addresses would decrease, since attack packets' target will be the only victim [27].

A group of packets those have the same source IP, destination IP, source port, destination port are called "flow". During SYN flood attack, there will be so many packets with different IP addresses and ports and in this case, there will be less packets in a flow, comparing with normal time [28].

Correlative flow means a group of packets those have opposite source IP – destination IP and source port - destination port couples, such as packet 1: source IP is A destination IP B source port is K destination port is L, packet 2: source IP is B destination IP A source port is L destination port is

K. It is expected that, there will be so many incoming traffic but no outgoing traffic during a Malware attack. So, percentage of correlative ows will reduce in case of an attack [29].

The author used a "connection size distribution" feature to detect attacks. Connection size distribution depends on the service which customer use, such as FTP, HTTP, DNS. Normally, it is assumed that this distribution is generally stable in a short term. Also, new IP addresses are suspicious in the system. If there are new IP addresses, their connection size distribution is checked and decided whether it is attack or not [30]. In the other study, source IP address distribution and packet number in a certain time interval are features are used to detect attacks. Lee et al. used hash functions to map source IP of incoming traffic. And they also checked the number of total packets. So, it is expected that, source IP of incoming traffic and total number of packets should increase during Malware attack [31].Sharma et al. investigated the difference of some entropy-based features. During an attack, uncertainty will increase because of nature of Malware attack. So, they used source IP, destination IP, source port, destination port, TCP ag set, protocol and length distributions. Then, they calculated these features' entropies to detect attack presence [32].

In this study the author meant for emergent a discovery scheme depend on numerous adapted perceptron procedures. For dissimilar procedures, he attained the correctness of 69.90%-96.18%. It must be specified that the procedures that caused in finest accurateness also shaped the uppermost amount of false-positives: the greatest precise one caused in 48 false positives. The greatest” balanced”s procedure with suitable correctness and the little false-positive degree had the correctness of 93.01%. [33].

In this study author uses machine learning to Contribution in Central Safety in Initiative Networks” deliberates the discovery technique depend on adapted RF technique in mixture with Info Improvement for healthier feature depiction. It must be observed, that the information usual contains chastely of moveable executable records, for which feature removal is usually calmer. The consequence attained is the correctness of 97% and 0.03 false-positive rate [34].

The author planned mining approaches depend on PE headers, DLLs and API purposes and approaches depend on NB, J48 DT, and SV M. Uppermost general correctness was attained with the J48 procedure [35].

In this study Supervised Learning Processes used to detect malware of API call Names", the API purposes remained applied for data illustration over. The finest result was attained through SVM technique with regularized polykernel. The accuracy of 97.6% was attained, with a false-positive degree of 0.025 [36].

As it can be realized, entirely researches finished awake with dissimilar fallouts. After here, we can accomplish that not at all united practice was shaped hitherto neither for discovery nor feature depiction. The correctness of respectively distinct circumstance be contingent on the particulars of malware relations applied and on the real application.

3. MATERIAL AND METHODS

In this section, several supervised and unsupervised techniques are presented in detail. Furthermore, the new malware detection method are presented and explained detail phase.

3.1 UNSUPERVISED TECHNIQUES

3.1.1 Factor analysis

D denotes the data set which contains of D-dimension actual standards D= {y1, y2, y3,}. The input dataset is formed by applying the Equation 3.1.

$$y= \Lambda x+ \varepsilon \tag{3.1}$$

where x is a k-dimensional zero-mean Gaussian vector with data competition to concealed features, $\wedge$ is atmosphere of factors in the method of D X K. The particulars of the mathematical model were illustrated in [36].

3.1.2 K-Means Clustering

Its aims to divider n items into k groups in which individually item fits to the cluster with the adjacent nasty. This technique yields precisely k dissimilar clusters of highest conceivable difference. The greatest amount of clusters k foremost to the utmost parting (space) is not identified as a priori and necessity be planned from the feature. The purpose of K-Means clustering is to decrease whole intra-cluster alteration. The procedures of the technique presented below [37].

1. Clusters the features into k sets where k is determined by the user.
2. determine k facts at accidental as cluster middles.
3. Assign objects to their closest cluster center according to the *Euclidean distance* function.
4. Compute the centroid or mean of all items in every cluster.

5. Recurrence stages 2, 3 and 4 till the similar facts are allocated to every cluster in successive circles.

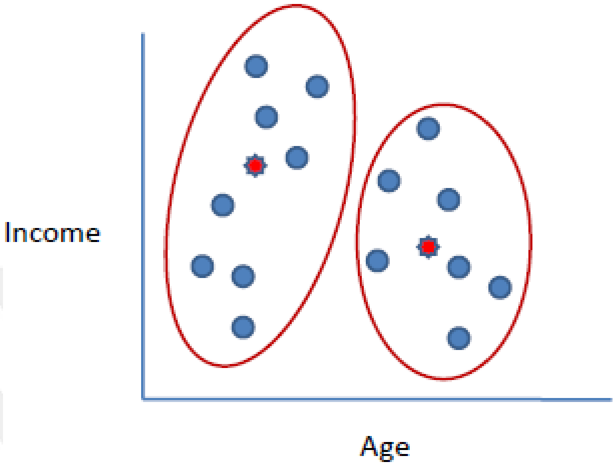


Figure 3.1: K-Means Clustering [37].

K-Means is comparatively a efficient technique. Though, we essential to stipulate the numeral of clusters, in spread and the last outcomes are subtle to initialization and often dismisses at a local optimum. Inappropriately, there is no general theoretical technique to discover the best amount of clusters. An applied method is to associate the consequences of manifold runs with dissimilar k and select the optimal one depend on a predefined standard. In overall, a great k perhaps reductions the error but upsurges the danger of overfitting.

3.1.3 Principal Component Analysis (PCA)

PCA is a common method applied in statistical learning methods. PCA can be applied to realize dimensionality decrease in regression settings letting us to explain a high-dimensional data set with a minor numeral of illustrative features which, in mixture, designate greatest of the erraticism originate in the original high-dimensional data.

PCA can besides be applied in unsupervised learning fields to learn, imagine a discover patterns in high-dimensional data sets when there is not exact reply feature. Moreover, PCA can assistance in clustering trainings and subdivision systems [38]. The method of PCA presented in Figure 3.2.

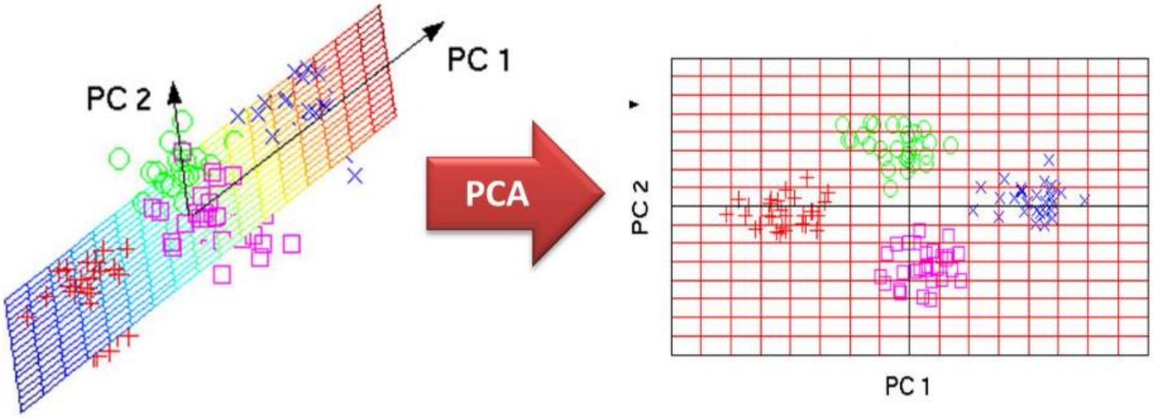


Figure 3.2: PCA [38].

3.1.4 Power Spectral Density (PSD)

In this chunk fleetingly we will clarify how the forte of a signal is separated in the frequency domain, relative to the fortes of any signals in the atmosphere, is principal to the project of all Linear time-invariant filter deliberate to suppress or assortment the signal [39]. This decent once signal is deterministic, and it change out to be only as precise in the vigorous of accidental signals. For instance, if a musical waveform is audio signals with collective disorder signals, it's necessarily make a little permit Linear time-invariant filter for mining the audio and curb the complaint signals [40],[41]. PSD purpose current the power of the energy in the signal as a function of frequency. The unit of PSD function is energy (difference) for each frequency(width) and can gain energy in a certain frequency domain by merge Power spectral density in that frequency domain. Figure 3.3 is example about power spectral for a signal.

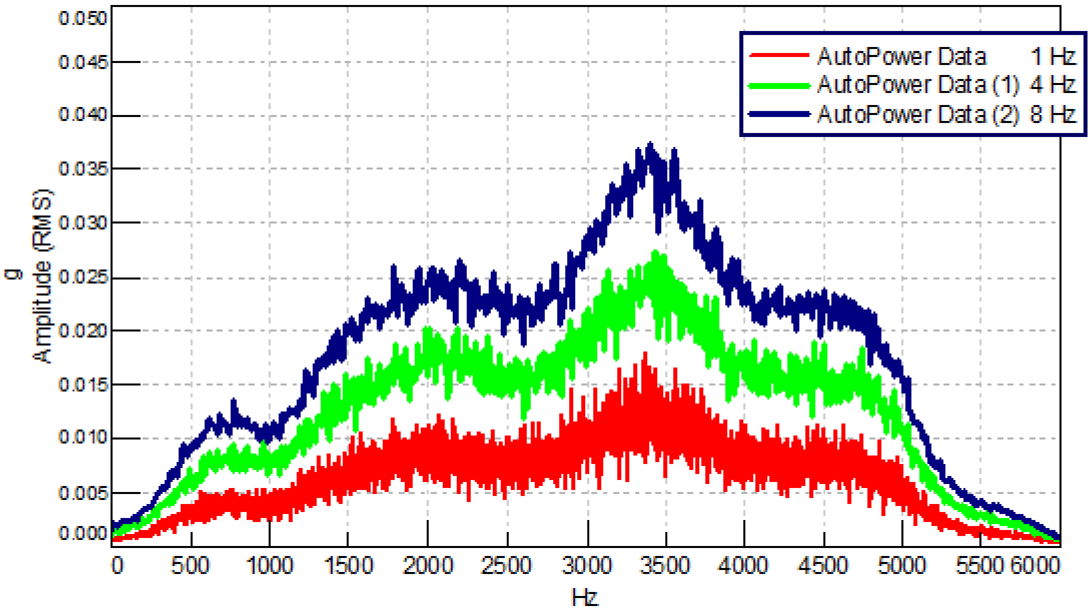


Figure 3.3: Power Spectral Density

3.2 SUPERVISED TECHNIQUES

3.2.1 Support Vector Machine

SVM is a discriminative classifier officially definite by a splitting hyperplane. In other study, assumed labeled training features, the technique productions a best hyperplane which classifies new instances. In two dimensional interplanetary this hyperplane is a line separating a plane in 2 sections where in each lesson lay in either side. Then, classifying any data have more than one Possible hyperplanes. In Figure 3.4 and 3.5 the possible hyperplanes presented for one equation.

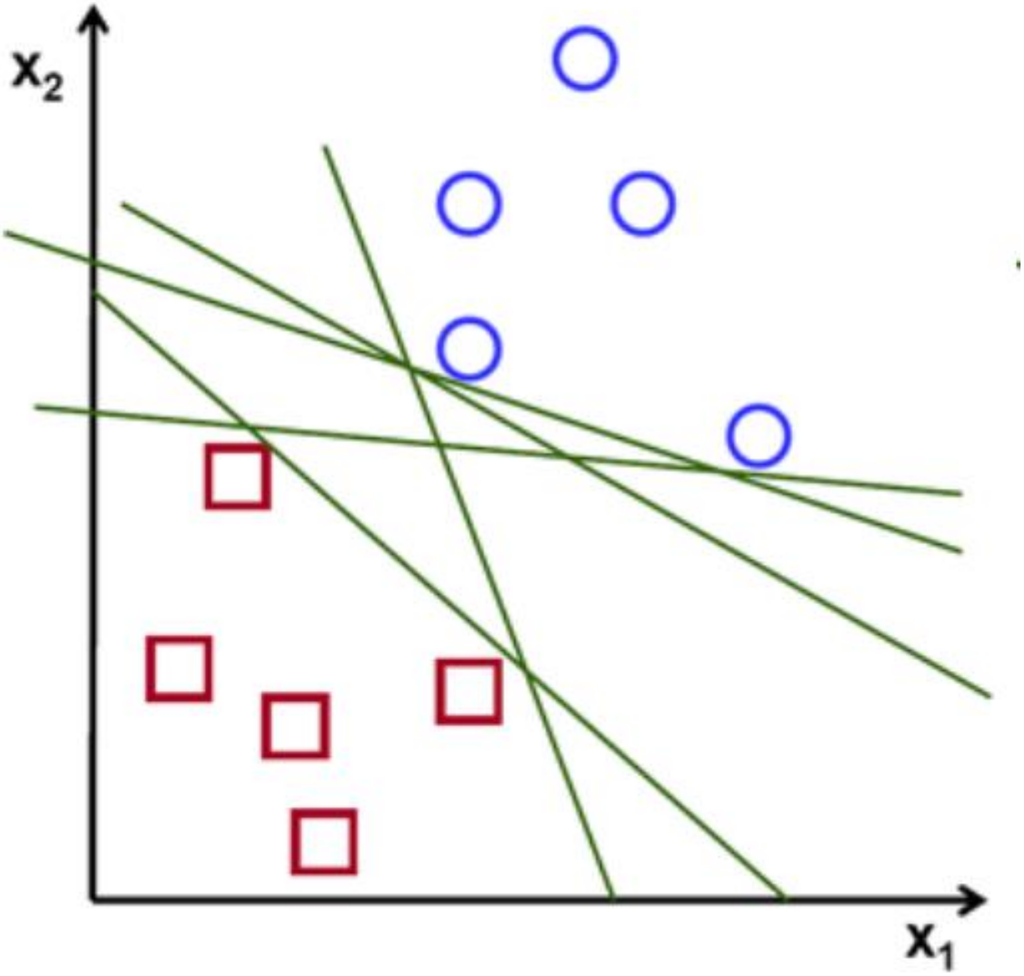


Figure 3.4: Possible hyperplane 1[42].

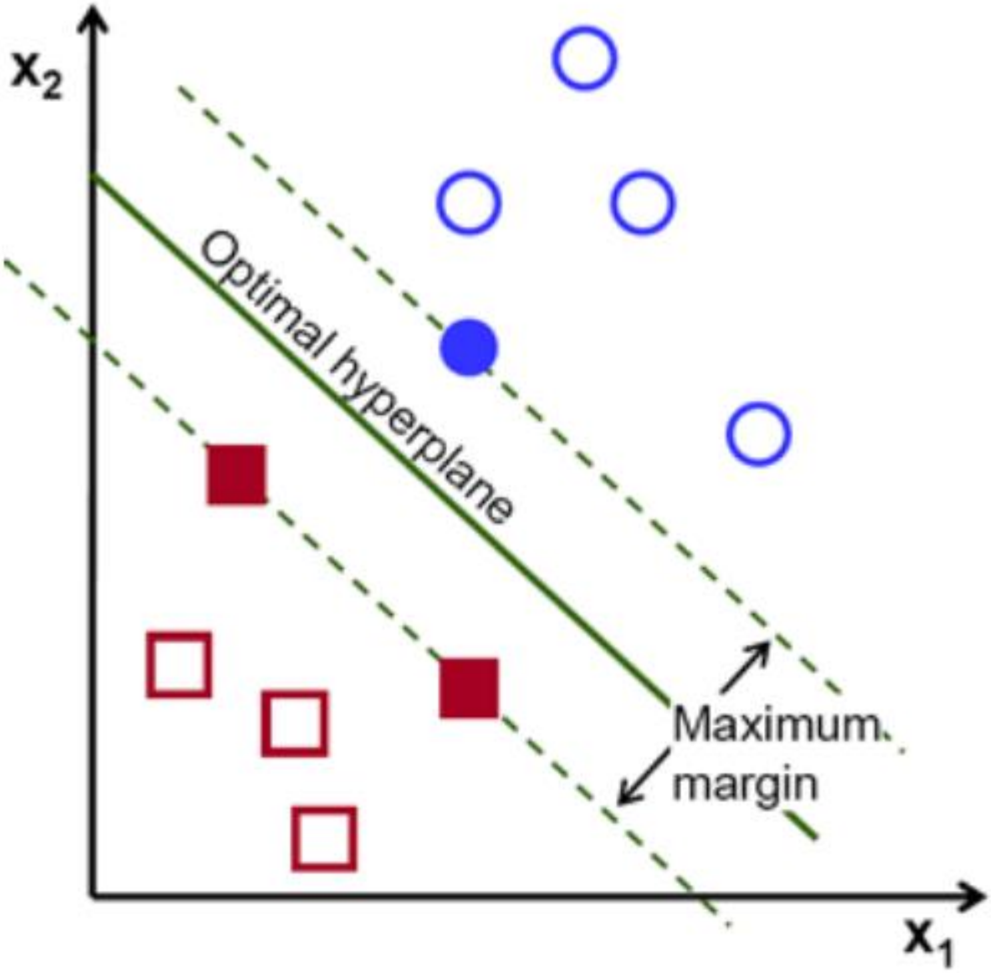


Figure 3.5: Possible hyperplane 2 [43].

Using these SVM, we increase the margin of the classifier. Deleting the support vectors will modification the location of the hyperplane. The, margin is very important issue when building SVM classifier.

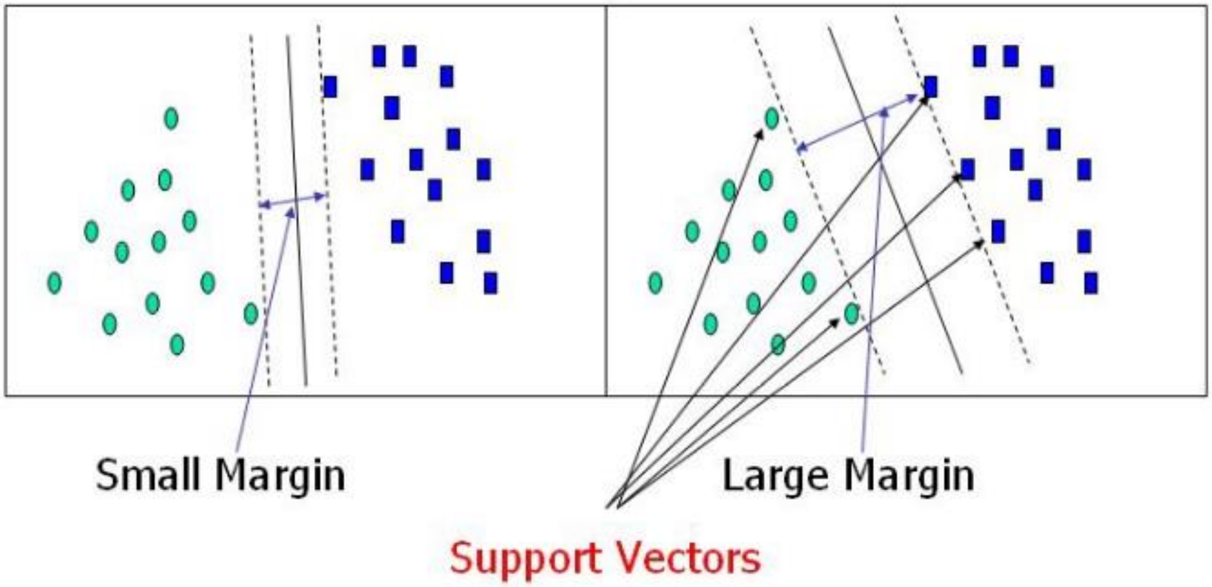


Figure 3.6: SVM margin Size [44].

.

3.2.2 k-Nearest Neighbors (KNN)

KNN is a family of classification technique and regression technique is frequently mentioned to as memory-depend training or example-depend training. Occasionally, it is also entitled lazy learning.

These relations agree to the chief idea of KNN. The idea is to substitute system formation by learning the training dataset and formerly apply this information to brand estimates.

Let's proceeds a humble situation to comprehend this technique. Subsequent is a feast of red circles and green squares [45]:

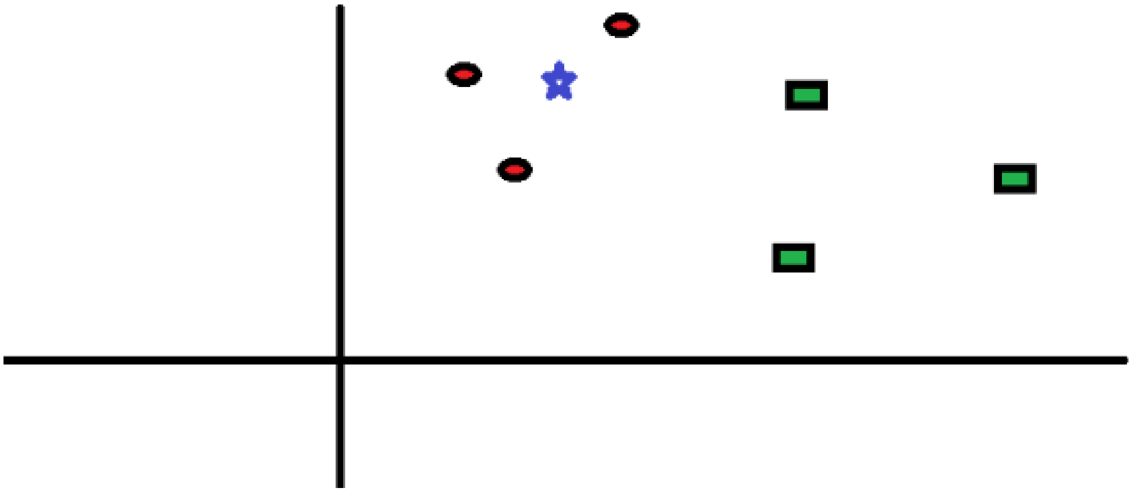


Figure 3.7: Data from 3 different points [46].

This mean to bargain out the lesson of the blue star. blue star can also be of red circles or green squares and nonentity different. The "K" is KNN technique is the nearest neighbors we request to income ballot from. Let's say K = 3. Therefore, this will nowadays brand a circle with blue star as midpoint fair as large as to encircle only three facts points on the flat. Mention to next diagram for more particulars

The 3 neighboring points to blue star is all red circles. Therefore, with decent sureness stage we can say that the blue star should fit to the lesson red circles. Now, the excellent became actual clear as all 3 ballots from the neighboring neighbor left to red circles. The excellent of the factor K is actual critical in this technique. Following we will comprehend what are the parameters to be careful to accomplish the finest K. The procedures of KNN are presented in Figures 3.8.

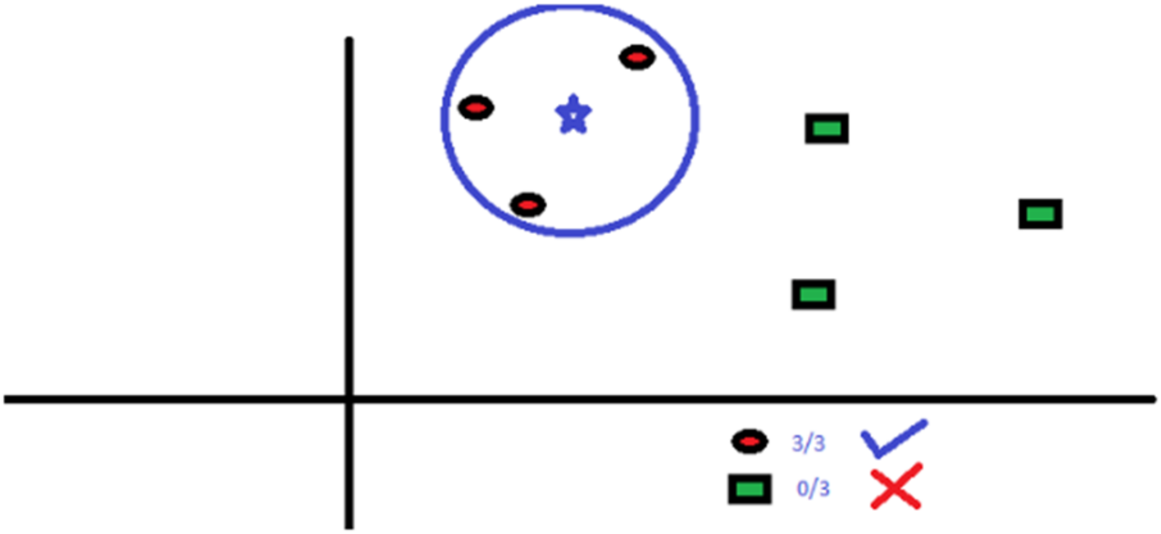


Figure 3.8: Classification using KNN [46].

3.2.3 Radial Basis Function (RBF)

A RBFN is a specifickind of NN. In this thesis, it will be explainin gistapplies as a non-linear classifier.

Commonly, when person'sconversation aroundNN or "ANN" they are denoting to the MLP. Individually neuron in an MLP incomes the weighted amount of its input standards. That is, apiece input cost is increased by a constant, and the outcomes are wholly summed collected. A lone MLP neuron is a humble linear classifier, but multifaceted nonlinear classifiers can be constructed by mixing these neurons into a net.

To me, the RBFN method is additional instinctive than the MLP. An RBFN achieves classification by calculating the input's resemblance to instances from the training group. Separately RBFN neuron supplies a "prototype", which is fair one of the instances from the training group. When we poverty for classifying a new input, individually neuron calculates the Euclidean distance among the input and its example. Unevenly language, if the input extra carefullylooks like the lessonA examples than the lesson B examples, it is identified as lesson A[47]. the construction of RBF shown in Figure 3.9.

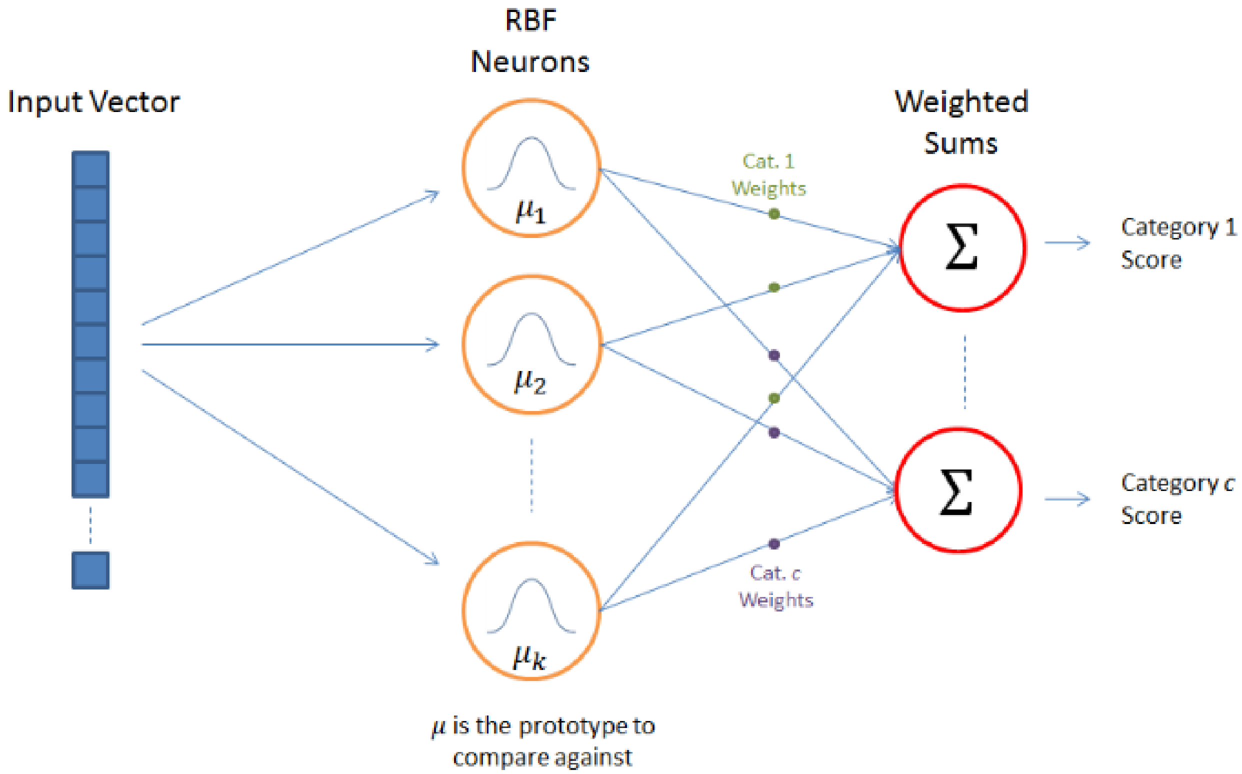


Figure 3.9: Radial Basis Function [47].

The beyond diagram displays the characteristic building of an RBF Network. It contains of an input array, a layer of RBF neurons, and an output layer with one nodule per group or lesson of data.

The input features are the n-dimensional features that you are annoying to categorize. The whole input feature is exposed to each of the RBF neurons see equation (3.7).

$$\phi(r) = \exp(-\frac{r^2}{2\sigma^2}) \qquad (3.2)$$

where r is a receptor and σ value of shaping parameter.

h(x) is the Gaussian system with the variable r (radius) and c (center) distinct distinctly at apiece RBF component. The training procedure is depend on regulating the limits of the net to repeat a traditional of input-output designs. There are 3 kinds of variables; the weight w among the concealed bulges and the output bulges, the center c of individually neuron of the hidden layer and the component width r. The manufacture of the net covers of a collection of nodes, unique each collection that we are maddening for categorizing. Unconnectedly produce node computes a kind of entire for the connected collection. Characteristically, agroup excellent is comprehensive by broadcast the input to the collection with the highest notch..

3.3 DATASET

The malware dataset [48] is creating to represent the security in android operating system. The data were gained by a procedure that contained to generate a binary array of consents applied for separately application investigated {1=used, 0=no used}. Furthermore, the examples of malware/kind were separated by "Kind"; 1 abnormal and 0 normal.

3.4 PROPOSED METHOD

The proposed method involves from 2 portions: feature mining and classifier. Formerly, the feature mining built by applying the PSD which attempt to get the greatest and subtle topographies from input features by measuring power spectral density. see equation 3.3.

$$\boldsymbol{P} = \int_{-\infty}^{\infty} x(f)^2 df \tag{3.3}$$

Where ∞ and - ∞ are the areas of purpose, f is the frequency standards. The PSD applied to measure PSD for respectively dated strongminded by operator rendering to the measurement and conduct of the input feature, Also, the production of the wired to the classifier which applied SVM, RBF and MLP. Formerly, numerous tests will consume complete to control planned technique

(PSD based MLP and RBF and SVM), which consequences are related with greatest presented approaches in this arena that’s proposed in section 2. The flowchart shown in Figure 3.10.

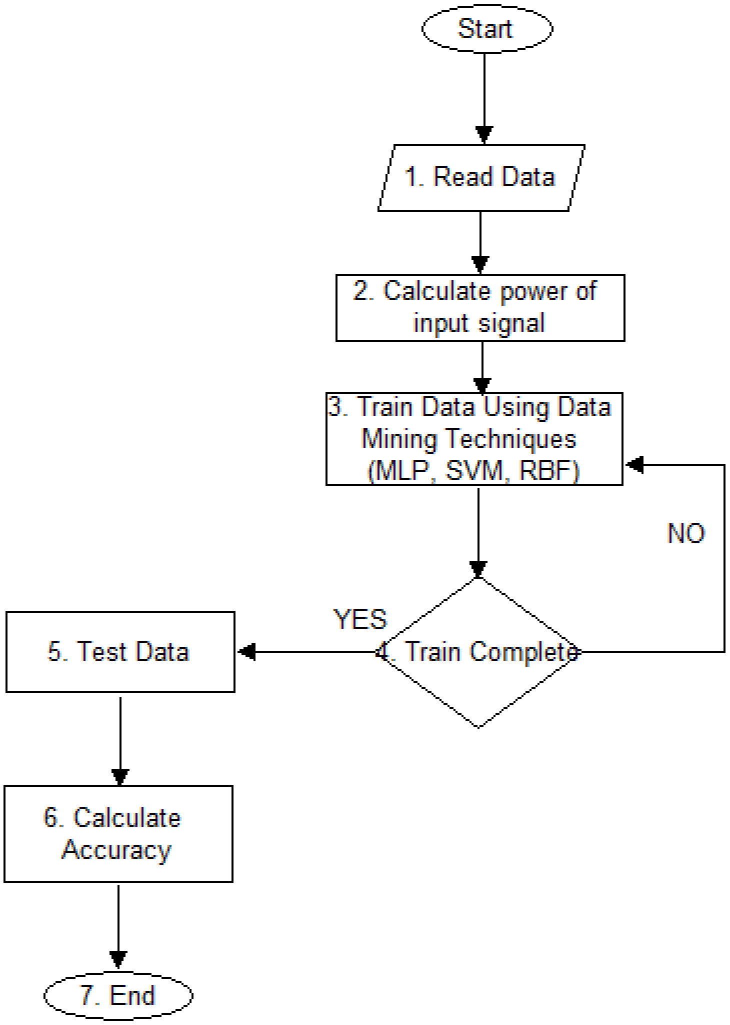

Start
1. Read Data
2. Calculate power of input signal
3. Train Data Using Data Mining Techniques (MLP, SVM, RBF)
NO
YES
4. Train Complete
5. Test Data
6. Calculate Accuracy
7. End

Figure 3.10: Proposed Method.

The details step of Figure 1 presented in details in the flowing:

1. Read the data [48] by using MATLAB program.
2. Calculate the input data power by using Eq (1).

3. Train classifier to classify the features that extracted by Eq(1).
4. If train complete continue with testing section or return to the training.
5. Then, test the trained model by using [48] dataset.
6. Calculate the accuracy of the results.
7. Complete

4. EXPERIMENTS AND DISSCUSION

In this unit, the experiment will be offered beside with the outcome and the calculation, separated into chapter as the follows:

4.1 MATLAB TOOL

MATLAB is a user interface enterprise instrument projected exactly for engineers and academics? The sensitivity of MATLAB is the MATLAB language, a matrix based programing language authorizing the uppermost normal onset of computational mathematics.

- Data study
- Algorithms grow
- Generate requests and replicas

The language, apps, and built-in math purposes permit the scientist to quickly learn various approaches to spread at an answer. MATLAB leases the specialists produce your opinions from study to production by putting to inventiveness applications and rooted plans, as well as fraternization with Simulink® and Model Founded Design.

4.2 MEAN SQUARE ERROR

In statistics, the mean squared error (MSE) or mean squared deviation (MSD) of an estimator (of a process for approximating an unnoticed amount) events the regular of the squares of the errors—that is, the regular squared change between the assessed standards and what is predictable. The MSE can calculated in equation (4.1).

$$\text{MSE} = \frac{1}{m}\sum_{k=1}^{n}(Y_k - F_k)^2 \tag{4.1}$$

Where m is amount of instance, k is the numeral of features F characterized the input features. Y characterized output data.

4.3 IMPLEMENTATION

Numerous experiments are performed in several settings, the test and the outcome were measured applying numerous capacities, the presentation of numerous tests were related and the outcomes were highlighted.

The experimental applied by using MATLAB2018 as tool. The dataset features that clarified in the chapter 3 used as input to the planned model that written by using MATLAB2018.

Moreover, the construction of the system consists from 513 input and 2 output classes. Then, the input features shown in Figure 4.1.

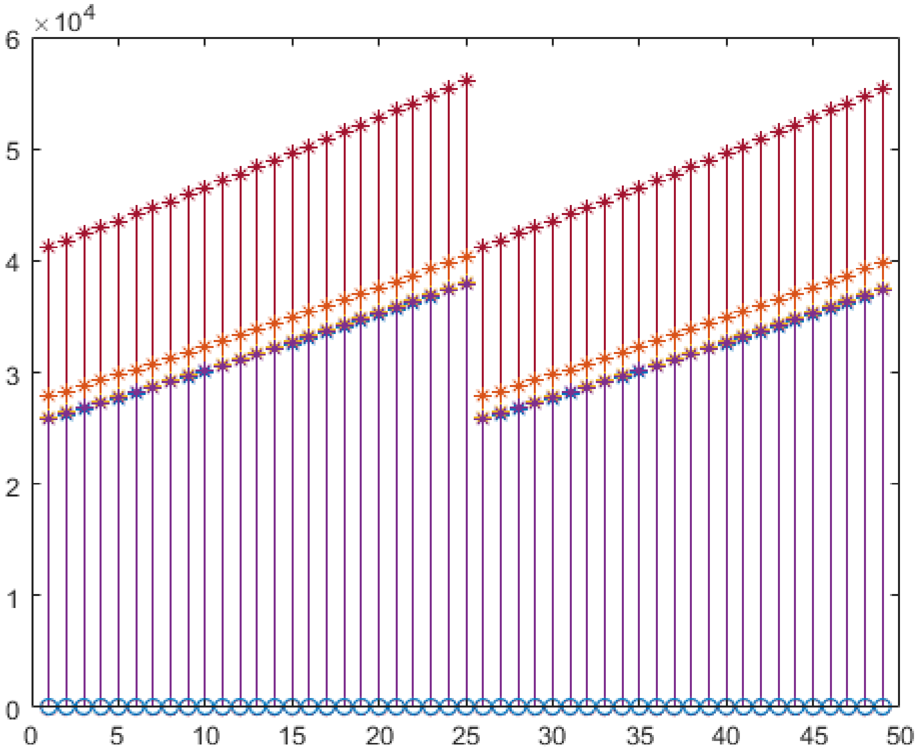


Figure 4.1: Input Features.

Also, one case which represented from 512 features are represented in Figure 4.2. Moreover, only 256 features are extracted which represented best and sensitive features and these features are shown in Figure 4.2.

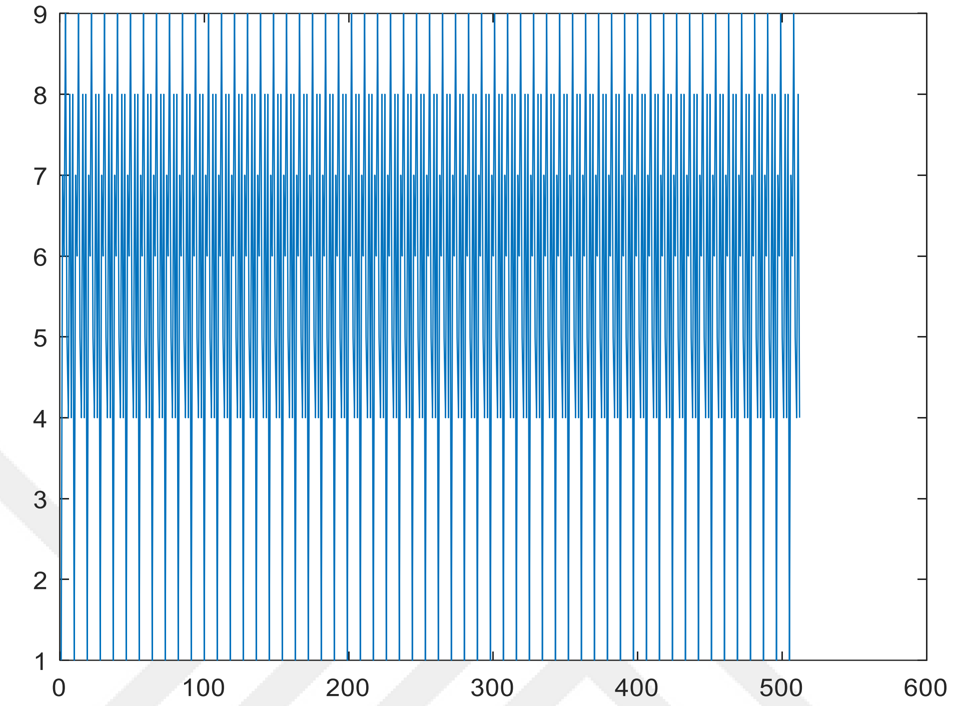


Figure 4.2: Features of one case.

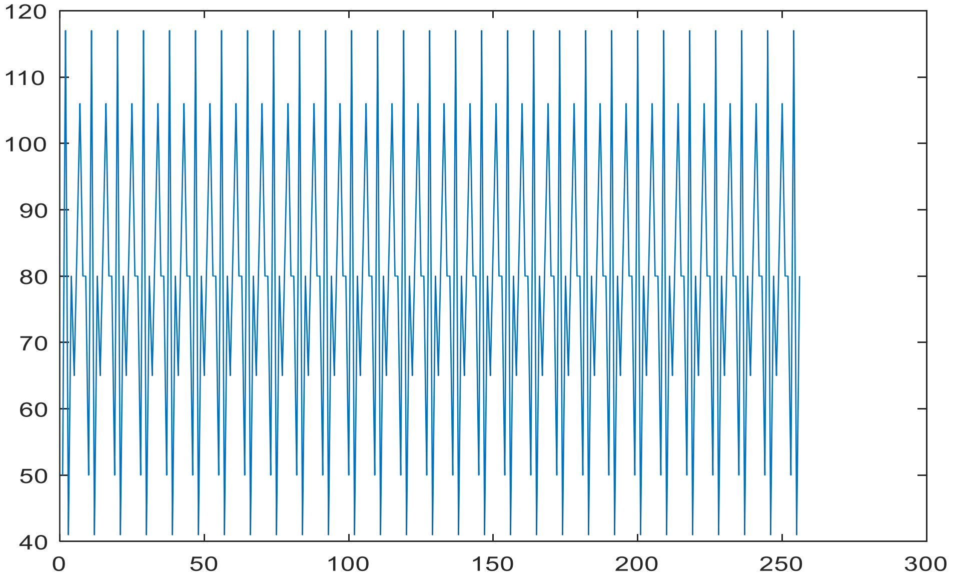


Figure 4.3: Input data after feature extraction.

In the first experiment, the output of the PSD Wired to RBF which used to classify features in to Attack and normal. The training process of RBF presented in Figure 4.4.

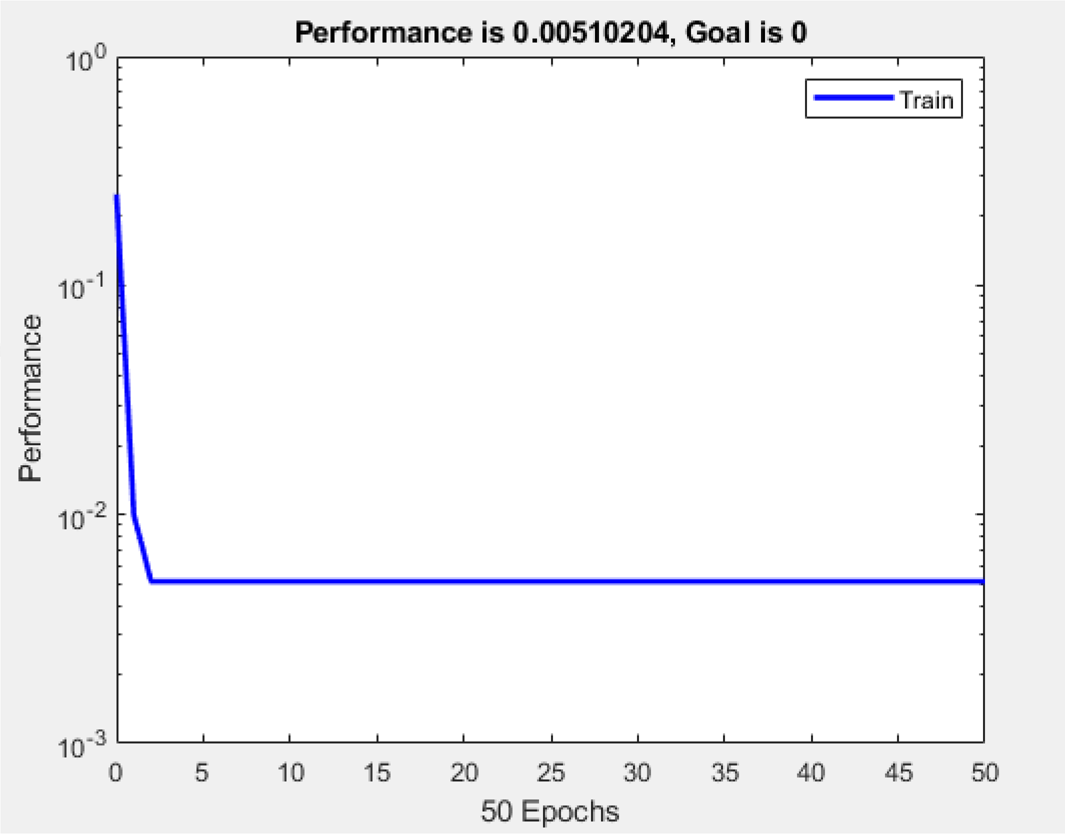


Figure 4.4: RBF Training Process.

Then, the classification results shown in Figure 4.5 which 99.00 accuracy are presented in Figure 4.5. Additionally, extra statistical parameters are presented in Table 4.1.

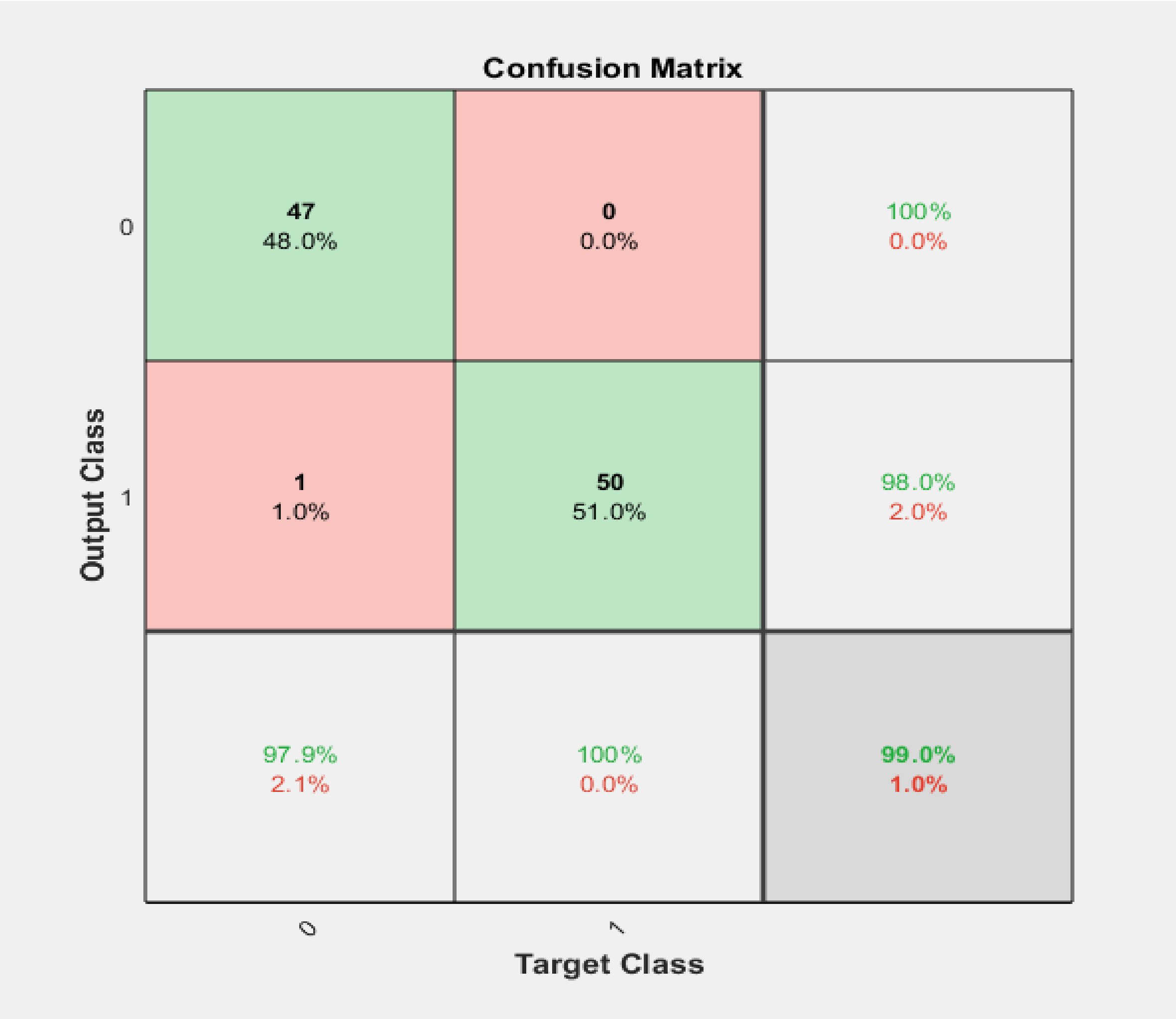
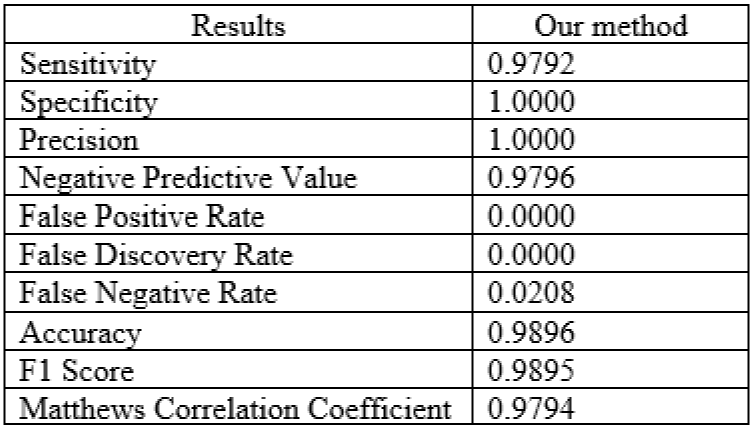


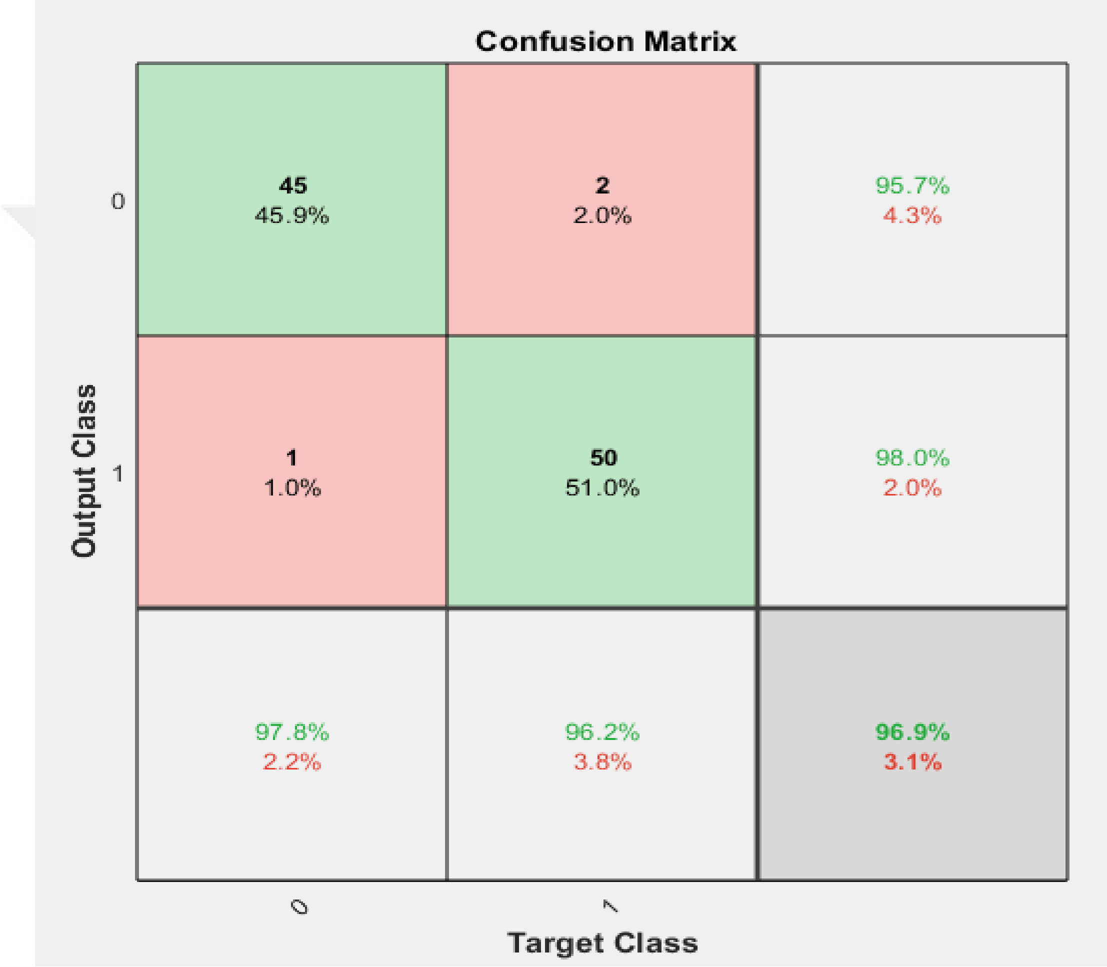
Figure 4.5: Confusion Matrix of RBF.

Table 4.1: Statistical Parameters For RBF.



Results	Our method
Sensitivity	0.9792
Specificity	1.0000
Precision	1.0000
Negative Predictive Value	0.9796
False Positive Rate	0.0000
False Discovery Rate	0.0000
False Negative Rate	0.0208
Accuracy	0.9896
F1 Score	0.9895
Matthews Correlation Coefficient	0.9794

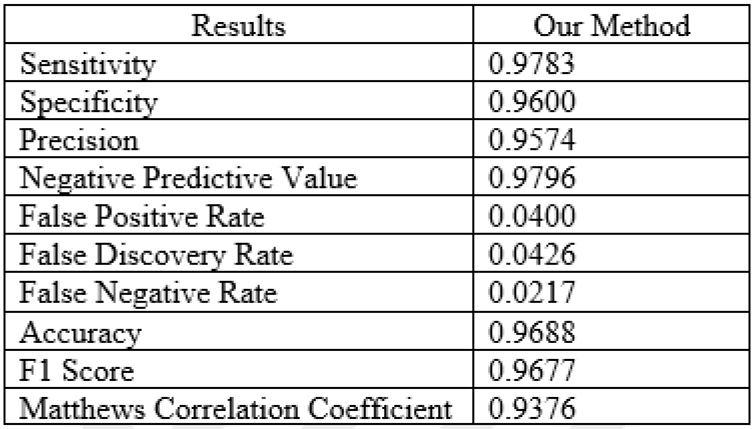
In the second experiment, the output of the PSD Wired to SVM which used to classify features in to Attack and normal. Confusion Matrix of SVM presented in Figure 4.6. Furthermore, statistical parameters presented in Table 4.2.



	0	1	
0	45 45.9%	2 2.0%	95.7% 4.3%
1	1 1.0%	50 51.0%	98.0% 2.0%
	97.8% 2.2%	96.2% 3.8%	96.9% 3.1%

Figure 4.6: Confusion Matrix Of SVM.

Table 4.2: statistical parameters for SVM.



Results	Our Method
Sensitivity	0.9783
Specificity	0.9600
Precision	0.9574
Negative Predictive Value	0.9796
False Positive Rate	0.0400
False Discovery Rate	0.0426
False Negative Rate	0.0217
Accuracy	0.9688
F1 Score	0.9677
Matthews Correlation Coefficient	0.9376

In the last experiment, the output of the PSD Wired to neural network which used to classify features in to Attack and normal. Confusion Matrix of neural network presented in Figure 4.7. Furthermore, statistical parameters presented in Table 4.3.

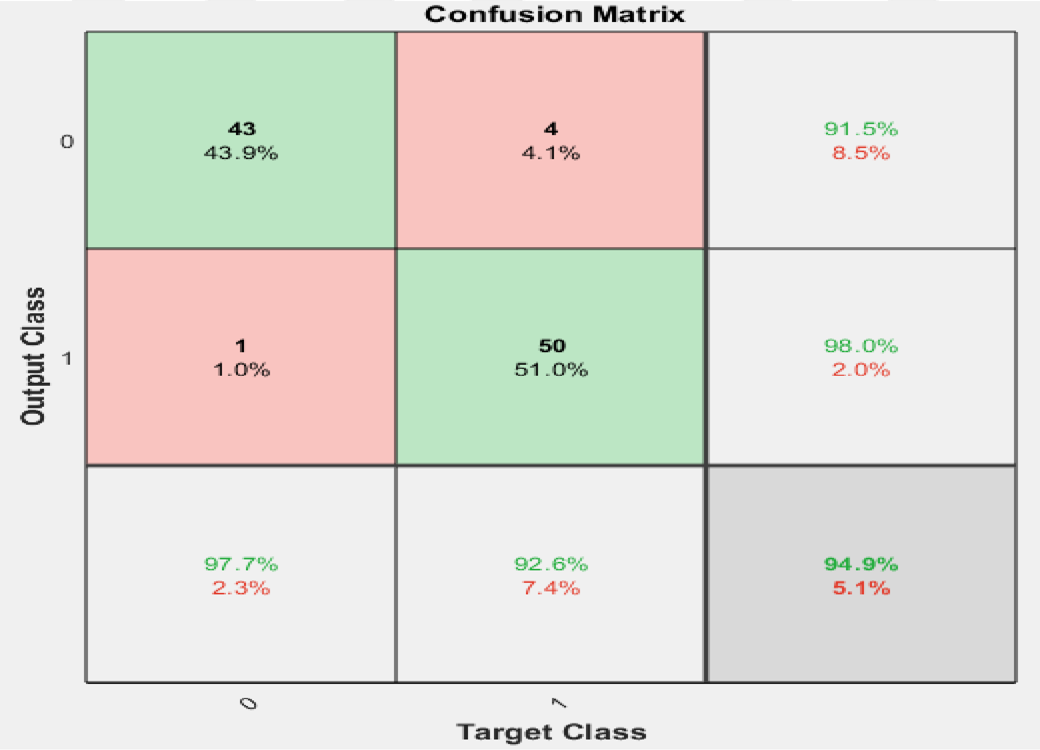
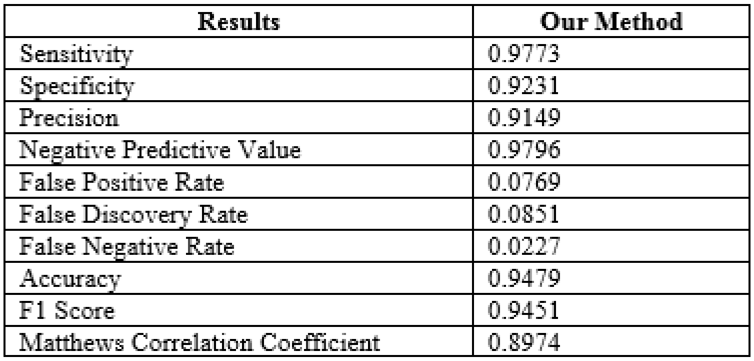


Figure 4.7: Confusion Matrix Of Neural Network.

Table 4.3: statistical parameters for neural network.



Results	Our Method
Sensitivity	0.9773
Specificity	0.9231
Precision	0.9149
Negative Predictive Value	0.9796
False Positive Rate	0.0769
False Discovery Rate	0.0851
False Negative Rate	0.0227
Accuracy	0.9479
F1 Score	0.9451
Matthews Correlation Coefficient	0.8974

The results of our methods that are presented in Table 4.1, Table 4.2 and Table 4.3 are listed in Table 4.4 and compared with well known studies in this field.

Table 4.4: Results Comparison.

Methods		ACC
[48]	SVM	96.30
[49]	Naïve-Bayes	93
[49]	J48	98
Our Method	RBF SVM NN	99 96.9 94.9

5. CONCLUSION

This thesis presented a new data mining technique based on PSD function for malware detection. The data are analyzed by using several auto-encoders to extract high level features from malware dataset and then the extracted features are classified using data mining techniques (SVM, RBF, MLP). The new method presented 99.00 accuracy which is remarkable results when compared with previous studies.

The presented method is new idea which combine the power of signal processing with data mining techniques. The aim of this study extracts sensitive features by using PSD to obtain high accuracy results. The extracted features wired to data mining classifier which trained in supervised learning to classify the extracted features.

Then, PSD can have combined with various classifiers to solve dissimilar kinds of problems because PSD presented remarkable results when compared with convolution feature extraction techniques.

The proposed method can easily apply to various classification problems such as EEG signal classification, EGC signal classification, face recognition, fingerprint recognition and disease detection by modifying only number of parameters like input features, hidden nodes and output classes.

REFERENCES

[1] J. Mirkovic and P. Reiher, "A taxonomy of Malware attack and Malware defense mechanisms," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 34, no. 2, p. 39, 2004.

[2] M. Boy, "Malware Detection types."Southeast Asian J. Trop. Med. Public Health, vol. 44, no. 4, pp. 568–573, 2016.

[3] N. Patani and R. Patel, "A Mechanism for Prevention of Flooding based Malware Attack," vol. 13, no. 1, pp. 101–111, 2017.

[4] I. Users, "No Title," 2015. [Online]. Available: internetlivestats.com/internet-users.

[5] A. Meek, "Malware attacks are getting much more powerful and the Pentagon is scrambling for solutions," 2015.

[6] Kaspersky, "Kaspersky Malware Intelligence Report Q3 2015," *Kaspersky Lab*, 2015.

[7] "Malware ATTACKS," 2011. [Online]. Available: https://www.incapsula.com/ddos/ddos-attacks.html.

[8] B. Hang, R. Hu, and W. Shi, "An enhanced SYN cookie defence method for Malware attack," *J. Networks*, vol. 6, no. 8, pp. 1206–1213, 2011.

[9] S. Taghavi Zargar, J. Joshi, D. Tipper, and S. Member, "A Survey of Defense Mechanisms Against Distributed Denial of Service (Malware) Flooding Attacks," pp. 1–24, 2013.

[10] M. Bogdanoski, T. Shuminoski, and A. Risteski, "Analysis of the SYN Flood DoS Attack," *Int. J. Comput. Netw. Inf. Secur.*, vol. 5, no. 8, pp. 15–11, 2013.

[11] G. Zhang and M. Parashar, "Cooperative Defense against Malware Attacks," 2002.

[12] M. Sachdeva, G. Singh, K. Kumar, and K. Singh, "Malware incidents and their impact: A review," *Int. Arab J. Inf. Technol.*, vol. 7, no. 1, pp. 14–20, 2010.

[13] S. Sinha and M. Sharma, "Simulation and Analysis of DDoS Attacks by Specialized

Simulator using Virtualization," vol. 3, no. 2, pp. 2–4, 2014.

[14] Cisco, "Internet Protocols."Comput. Stat. Data Anal., vol. 125, pp. 44–56, 2014.

[15] X. Rui, W. L. Ma, and W. L. Zheng, "Defending against UDP Flooding by negative selection algorithm based on eigenvalue sets," *5th Int. Conf. Inf. Assur. Secur. IAS 2009*, vol. 2, pp. 342–345, 2009.

[16] M. Rouse, "UDP (User Datagram Protocol) definition." [Online]. Available: https://searchnetworking.techtarget.com/definition/UDP-User-Datagram-Protocol.

[17] cloudflare, "What is a DDoS Attack?" [Online]. Available: https://www.cloudflare.com/learning/ddos/what-is-a-ddos-attack/.

[18] E. Alpaydin, "Introduction to machine learning," *Methods Mol. Biol.*, vol. 1107, pp. 105–28, 2004.

[19] X. Zhu, "Semi-Supervised Learning Tutorial (ICML 2007)," 2007.

[20] K. Sun, J. Zhang, C. Zhang, and J. Hu, "Generalized extreme learning machine autoencoder and a new deep neural network," *Neurocomputing*, vol. 230, no. November 2016, pp. 374–381, 2017.

[21] Z. Ghahramani, "LNAI 3176 - Unsupervised Learning," *Mach. Learn.*, pp. 72–112, 2004.

[22] T. Hastie and R. Tibshirani, "Week 9: Unsupervised Learning ," *Stat. Learn. Course*, no. 9, pp. 1–60, 2013.

[23] C. Sun, C. Hu, Y. Tang, and B. Liu, "More accurate and fast SYN flood detection," *Proc. - Int. Conf. Comput. Commun. Networks, ICCCN*, 2009.

[24] H. Liu, Y. Sun, and M. S. Kim, "Fine-grained DDoS detection scheme based on bidirectional count sketch," *Proc. - Int. Conf. Comput. Commun. Networks, ICCCN*, 2011.

[25] J. Tajer, M. Adda, and B. Aziz, “Comparison Between Divergence Measures for Anomaly Detection of Mobile Agents in IP Networks,” *Int. J. Wirel. Mob. Networks*, vol. 9, no. 3, pp. 01-13, 2017.

[26] C. Sun, J. Fan, and B. Liu, “A robust scheme to detect SYN flooding attacks,” *Proc. Second Int. Conf. Commun. Netw. China, ChinaCom 2007*, no. September 2007, pp. 397–401, 2008.

[27] D. Nashat, X. Jiang, and S. Horiguchi, “Router based detection for low-rate agents of DDoS attack,” *2008 Int. Conf. High Perform. Switch. Routing, HPSR 2008*, no. March, pp. 177–182, 2008.

[28] J. Singh, M. Sachdeva, and K. Kumar, “Detection of DDoS Attacks Using Source IP Based Entropy,” vol. 3, no. 1, pp. 201–210, 2013.

[29] A. M. Brues, “Genetic effects of the atom bomb,” *J. Hered.*, vol. 38, no. 5, pp. 137–137, 1947.

[30] D. Wang, Z. Yufu, and J. Jie, “A multi-core based DDoS detection method,” *Proc. - 2010 3rd IEEE Int. Conf. Comput. Sci. Inf. Technol. ICCSIT 2010*, vol. 4, pp. 115–118, 2010.

[31] H. Rahmani, N. Sahli, and F. Kamoun, “DDoS flooding attack detection scheme based on F-divergence,” *Comput. Commun.*, vol. 35, no. 11, pp. 1380–1391, 2012.

[32] P. Xu, H. Peng, and T. Huang, “Unsupervised learning of mixture regression models for longitudinal data,” *Comput. Stat. Data Anal.*, vol. 125, pp. 44–56, 2018.

[33] L. Castro, E. A. Wasserman, and M. Lauffer, “Unsupervised learning of complex associations in an animal model,” *Cognition*, vol. 173, no. October 2017, pp. 28–33, 2018.

[34] S. Shabani and Y. Norouzi, “Speech recognition using Principal Components Analysis and Neural Networks,” *2016 IEEE 8th Int. Conf. Intell. Syst.*, pp. 90–95, 2016.

[35] U. Rajendra Acharya, S. Vinitha Sree, A. P. C. Alvin, and J. S. Suri, “Use of principal component analysis for automatic classification of epileptic EEG activities in wavelet framework,” *Expert Syst. Appl.*, vol. 39, no. 10, pp. 9072–9078, 2012.

[36] A. M. Karim, M. S. Güzel, M. R. Tolun, H. Kaya, and F. V Çelebi, “A new framework

using deep auto-encoder and energy spectral density for medical waveform data classification and processing,” *Biocybern. Biomed. Eng.*, vol. 39, no. 1, pp. 148–159, 2019.

[1] J. Mirkovic and P. Reiher, “A taxonomy of DDoS attack and DDoS defense mechanisms,” *ACM SIGCOMM Comput. Commun. Rev.*, vol. 34, no. 2, p. 39, 2004.

[2] M. Boy, “DDOS Detection types.”Southeast Asian J. Trop. Med. Public Health, vol. 44, no. 4, pp. 568–573, 2016.

[3] N. Patani and R. Patel, “A Mechanism for Prevention of Flooding based DDoS Attack,” vol. 13, no. 1, pp. 101–111, 2017.

[4] I. Users, “No Title,” 2015. [Online]. Available: internetlivestats.com/internet-users.

[5] A. Meek, “DDoS attacks are getting much more powerful and the Pentagon is scrambling for solutions,” 2015.

[6] Kaspersky, “Kaspersky DDoS Intelligence Report Q3 2015,” *Kaspersky Lab*, 2015.

[7] “DDOS ATTACKS,” 2011. [Online]. Available: https://www.incapsula.com/ddos/ddos-attacks.html.

[8] B. Hang, R. Hu, and W. Shi, “An enhanced SYN cookie defence method for TCP DDoS attack,” *J. Networks*, vol. 6, no. 8, pp. 1206–1213, 2011.

[9] S. Taghavi Zargar, J. Joshi, D. Tipper, and S. Member, “A Survey of Defense Mechanisms Against Distributed Denial of Service (DDoS) Flooding Attacks,” pp. 1–24, 2013.

[10] M. Bogdanoski, T. Shuminoski, and A. Risteski, “Analysis of the SYN Flood DoS Attack,” *Int. J. Comput. Netw. Inf. Secur.*, vol. 5, no. 8, pp. 15–11, 2013.

[11] G. Zhang and M. Parashar, “Cooperative Defense against DDoS Attacks,” 2002.

[12] M. Sachdeva, G. Singh, K. Kumar, and K. Singh, “DDoS incidents and their impact: A review,” *Int. Arab J. Inf. Technol.*, vol. 7, no. 1, pp. 14–20, 2010.

[13] S. Sinha and M. Sharma, “Simulation and Analysis of DDoS Attacks by Specialized Simulator using Virtualization,” vol. 3, no. 2, pp. 2–4, 2014.

[14] Cisco, “Internet Protocols.”Comput. Stat. Data Anal., vol. 125, pp. 44–56, 2014.

[15] X. Rui, W. L. Ma, and W. L. Zheng, “Defending against UDP Flooding by negative selection algorithm based on eigenvalue sets,” *5th Int. Conf. Inf. Assur. Secur. IAS 2009*, vol. 2, pp. 342–345, 2009.

[16] M. Rouse, “UDP (User Datagram Protocol) definition.” [Online]. Available: https://searchnetworking.techtarget.com/definition/UDP-User-Datagram-Protocol.

[17] cloudflare, “What is a DDoS Attack?” [Online]. Available: https://www.cloudflare.com/learning/ddos/what-is-a-ddos-attack/.

[18] E. Alpaydin, “Introduction to machine learning,” *Methods Mol. Biol.*, vol. 1107, pp. 105–28, 2004.

[19] X. Zhu, “Semi-Supervised Learning Tutorial (ICML 2007),” 2007.

[20] K. Sun, J. Zhang, C. Zhang, and J. Hu, “Generalized extreme learning machine autoencoder and a new deep neural network,” *Neurocomputing*, vol. 230, no. November 2016, pp. 374–381, 2017.

[21] Z. Ghahramani, “LNAI 3176 - Unsupervised Learning,” *Mach. Learn.*, pp. 72–112, 2004.

[22] T. Hastie and R. Tibshirani, “Week 9: Unsupervised Learning ,” *Stat. Learn. Course*, no. 9, pp. 1–60, 2013.

[23] C. Sun, C. Hu, Y. Tang, and B. Liu, “More accurate and fast SYN flood detection,” *Proc. - Int. Conf. Comput. Commun. Networks, ICCCN*, 2009.

[24] H. Liu, Y. Sun, and M. S. Kim, “Fine-grained DDoS detection scheme based on bidirectional count sketch,” *Proc. - Int. Conf. Comput. Commun. Networks, ICCCN*, 2011.

[25] J. Tajer, M. Adda, and B. Aziz, “Comparison Between Divergence Measures for Anomaly Detection of Mobile Agents in IP Networks,” *Int. J. Wirel. Mob. Networks*, vol. 9, no. 3, pp. 01-13, 2017.

[26] C. Sun, J. Fan, and B. Liu, “A robust scheme to detect SYN flooding attacks,” *Proc. Second Int. Conf. Commun. Netw. China, ChinaCom 2007*, no. September 2007, pp. 397–401, 2008.

[27] D. Nashat, X. Jiang, and S. Horiguchi, “Router based detection for low-rate agents of DDoS attack,” *2008 Int. Conf. High Perform. Switch. Routing, HPSR 2008*, no. March, pp. 177–182, 2008.

[28] J. Singh, M. Sachdeva, and K. Kumar, “Detection of DDoS Attacks Using Source IP Based Entropy,” vol. 3, no. 1, pp. 201–210, 2013.

[29] A. M. Brues, “Genetic effects of the atom bomb,” *J. Hered.*, vol. 38, no. 5, pp. 137–137, 1947.

[30] D. Wang, Z. Yufu, and J. Jie, “A multi-core based DDoS detection method,” *Proc. - 2010 3rd IEEE Int. Conf. Comput. Sci. Inf. Technol. ICCSIT 2010*, vol. 4, pp. 115–118, 2010.

[31] H. Rahmani, N. Sahli, and F. Kamoun, “DDoS flooding attack detection scheme based on F-divergence,” *Comput. Commun.*, vol. 35, no. 11, pp. 1380–1391, 2012.

[32] P. Xu, H. Peng, and T. Huang, “Unsupervised learning of mixture regression models for longitudinal data,” *Comput. Stat. Data Anal.*, vol. 125, pp. 44–56, 2018.

[33] L. Castro, E. A. Wasserman, and M. Lauffer, “Unsupervised learning of complex associations in an animal model,” *Cognition*, vol. 173, no. October 2017, pp. 28–33, 2018.

[34] S. Shabani and Y. Norouzi, “Speech recognition using Principal Components Analysis and Neural Networks,” *2016 IEEE 8th Int. Conf. Intell. Syst.*, pp. 90–95, 2016.

[35] U. Rajendra Acharya, S. Vinitha Sree, A. P. C. Alvin, and J. S. Suri, “Use of principal component analysis for automatic classification of epileptic EEG activities in wavelet framework,” *Expert Syst. Appl.*, vol. 39, no. 10, pp. 9072–9078, 2012.

[36] A. M. Karim, M. S. Güzel, M. R. Tolun, H. Kaya, and F. V Çelebi, “A new framework

using deep auto-encoder and energy spectral density for medical waveform data classification and processing," *Biocybern. Biomed. Eng.*, vol. 39, no. 1, pp. 148–159, 2019.

[37] A. M. Karim, Ö. Karal, and F. V Çelebi, "A New Automatic Epilepsy Serious Detection Method by Using Deep Learning Based on Discrete Wavelet Transform," no. 4, pp. 15–18, 2018.

[38] A. M. Karim, M. S. Güzel, M. R. Tolun, H. Kaya, and F. V Çelebi, "A New Generalized Deep Learning Framework Combining Sparse Auto-encoder and Taguchi Method for Novel Data Classification and Processing," pp. 1–22, 2018.

[39] T. Han, K. Hao, Y. Ding, and X. Tang, "Neurocomputing A sparse autoencoder compressed sensing method for acquiring the pressure array information of clothing," *Neurocomputing*, vol. 275, pp. 1500–1510, 2018.

[40] M. L. Huang, Y. H. Hung, W. M. Lee, R. K. Li, and B. R. Jiang, "SVM-RFE based feature selection and taguchi parameters optimization for multiclass SVM Classifier," *Sci. World J.*, vol. 2014, 2014.

[41] W. M. Zuo, W. G. Lu, K. Q. Wang, and H. Zhang, "Diagnosis of cardiac arrhythmia using kernel difference weighted KNN classifier," *Comput. Cardiol.*, vol. 35, pp. 253–256, 2008.

[42] Z. Yu *et al.*, "Prostatic Schistosoma japonicum with atypical immunophenotyping of individual glandular tubes: a case report and review of the literature.," *Southeast Asian J. Trop. Med. Public Health*, vol. 44, no. 4, pp. 568–573, 2013.

[42] Q. C. Hsu and A. T. Do, "Minimum porosity formation in pressure die casting by taguchi method," *Math. Probl. Eng.*, vol. 2013, 2013.

[43] V. Sze, Y.-H. Chen, T.-J. Yang, and J. Emer, "Efficient Processing of Deep Neural Networks: A Tutorial and Survey," vol. 105, no. 12, pp. 2295–2329, 2017.

[44] Dept. of Information and comm., Engineering Chosun University, D. Jha, and G.-R. Kwon, "Alzheimer's Disease Detection Using Sparse Autoencoder, Scale Conjugate Gradient and Softmax Output Layer with Fine Tuning," *Int. J. Mach. Learn. Comput.*, vol. 7, no. 1, pp. 13–17, 2017.

[45] C. Yin, Y. Zhu, J. Fei, and X. He, “A Deep Learning Approach for Intrusion Detection Using Recurrent Neural Networks,” vol. 5, 2017.

[46] S. Singh, S. K. Pandey, U. Pawar, and R. R. Janghel, “Classification of ECG Arrhythmia using Recurrent Neural Networks,” *Procedia Comput. Sci.*, vol. 132, no. Iccids, pp. 1290–1297, 2018.

[47] H. Deng, L. Zhang, and X. Shu, “Feature memory-based deep recurrent neural network for language modeling,” *Appl. Soft Comput. J.*, vol. 68, pp. 432–446, 2018.

[48] Mohammad M. Masud, Latifur Khan, and Bhavani Thuraisingham., A Hybrid Model to Detect Malicious Executables, IEEE Communications Society subject matter experts for publication in the ICC 2007 proceedings..

[49] Usukhbayar Baldangombo, Nyamjav Jambaljav, and Shi-Jinn Horng “A STATIC MALWARE DETECTION SYSTEM USINGDATA MINING METHODS”, International Journal of Artificial Intelligence & Applications (IJAIA), Vol. 4, No. 4, July 2013, 10.5121/ijaia.2013.4411.