CYBERTHREAT DETECTION BASED ON ARTIFICIAL NEURAL NETWROKS USING EVENT PROFILES

[1]D.Sravani, [2]Ch.Anusha, [3]G.Sathvika, [4]K.Pranathi, [5]S.Varsha Reddy,
[1,2,3,4,]U.G.Scholor, Department of IT, Sri Indu College Of Engineering & Technology, Ibrahimpatnam, Hyderabad.
[5]Assistant Professor, Department of IT, Sri Indu College Of Engineering & Technology, Ibrahimpatnam, Hyderabad.

ABSTRACT

One of the major challenges in cyber security is the provision of an automated and effective cyber-threats detection technique. In this paper, we present an AI technique for cyber-threats detection, based on artificial neural networks. The proposed technique converts multitude of collected security events to individual event profiles and use a deep learning-based detection methodforenhancedcyber-threatdetection.For this work, we developed an AI-SIEM system based on a combination of event profiling for data preprocessing and different artificial neural network methods, including FCNN, CNN, and LSTM. The system focuses on discriminating between true positive and false positive alerts, thus helpingsecurityanalysts to rapidlyrespond to cyber threats.
All experiments in this study are performed by authors using two benchmarkdatasets (NSLKDDandCICIDS2017)andtwodatasets collected in the real world. To evaluate the performance comparison with existing methods, we conducted experiments using the five conventional machine-learning methods (SVM, k-NN, RF, NB, and DT). Consequently, the experimentalresultsofthisstudyensurethatour proposed methods are capable of being employed as learning-based models for network intrusion- detection, and show that although it is employed in the real world, the performance outperforms the conventional machine-learning methods.

1. INTRODUCTION

Contrasted with the past, improvements in PC and correspondence innovations have given broad and propelled changes. The use of new innovations gives incredible advantages to people, organizations, and governments, be that as it may, messes some up against them. For instance, the protection of significant data, security of put away information stages, accessibility of information and so forth. Contingent upon these issues, digital fear-based oppression is one of the most significant issues in this day and age. Digital fear, which made a great deal of issues people and establishments, has arrived at a level that could undermine open and nation security by different gatherings, for example, criminal association, proficient people and digital activists. Alongthese lines, Intrusion Detection Systems (IDS) has been created to maintain a strategic distance from digital assaults. Right now, learning the bolster support vector machine (SVM) calculations wereutilized to recognize port sweep endeavors dependent on the new CICIDS2017 dataset with 97.80%, 69.79% precision rates were accomplished individually. Rather than SVMwe can introduce some other algorithms like random forest, CNN, ANN where these algorithms can acquire accuracies like SVM – 93.29,CNN–63.52,RandomForest–99.93, ANN–99.11.

MOTIVATION

The use of new innovations gives incredible advantages to people, organizations, and governments, be that as it may, messes some up against them. For instance, the protection of significant data, security of put away information stages, accessibility of information and so forth. Contingent upon these issues, digital fear-based oppression is one of the most significant issues in this day and age. Digital fear, which made a great deal of issues people and establishments, has arrived at a level that could undermine open and nation security by different gatherings, for example, criminal association, proficient people and digital activists. Along these lines, Intrusion Detection Systems (IDS) has been created to maintain a strategic distance from digital assaults.

OBJECTIVES

Objectiveofthisprojectistodetectcyber-attacks by using machine learning algorithms like

CNN:

A Convolutional Neural Network (CNN) is a type of deep learning neural network that is well-suited for image and video analysis.Random forest: It is the tech industry's definitive destination for sharing compelling, first-person accounts of problem- solving on the road to innovation.

ANN:

ANN algorithm accepts only numeric and structured data. Convolutional Neural Networks (CNN) and Recursive Neural Networks (RNN) are used to accept unstructured and non-numeric data forms such as Image, Text, and Speech

OUTCOMES:

These predictions can be done by four algorithms like SVM, ANN, RF, CNN thispaper helps to identify which algorithm predicts the best accuracy rates which helps to predict best results to identify the cyber-attacks happened or not.

APPLICATIONS:

This strategy used in Detection of Cyber Attack inNetworkusingMachine LearningTechniques

2. LITERATURESURVEY

Literature Survey Intrusion detection is very important for network situation awareness. While a few methods have been proposed to detect network intrusion, they cannot directly and effectively utilize semi-quantitative information consisting of expert knowledge and quantitative data.

Hence, this paper proposes a new detection model based on a directed acyclic graph (DAG) and a belief rule base (BRB). In the proposed model,calledDAG-BRB,theDAGisemployed toconstruct amulti-layeredBRBmodelthatcan avoid explosion of combinations of rule number because of a large number of types of intrusion. To obtain the optimal parameters of the DAG-BRB model, an improved constraint covariance matrix adaption evolution strategy (CMA-ES) is developed that can effectively solve the constraint problem in the BRB. A case study was used to test the efficiency of the proposed DAG-BRB.

The results showed that compared with other detection models, the DAG-BRB model has a higher detection rate and can be used in real network.

Nowadays, IT organizations generate colossal amounts of data. Handling these chunks of data itself is critical in the IT world. Hence centralizing the log management system improves security thereby enhances data protection in an organization. Such enterprises require a high profiling tool that helps in managing the information and events data to improve the level of security.

SecurityInformationandEventManagement (SIEM) :

It is a procedure for security analysis that prominence an overview of security in an organization.SIEMtoolscollect,analyze,

normalize and correlates all files and analyze data coming from the various device and give a centralized view of logs. This paper articulates an abstraction of SIEM tools and event correlation engines, furnishing a description of their technical comparative study, focusing on most popular SIEM tools and open source rule- based correlation engines and profiles them.

Distributed computing has become an effective approach to enhance capabilities of aninstitution or organization and minimize requirements for additional resource. In this regard, the distributed computing helps in broadening institutes IT capabilities. One needs to note that distributed computing is now integral part of most expanding IT business sector. It is considered novel and efficientmeans for expanding business. As more organizations and individuals start to use the cloud

Literature Survey Intrusion detection is very important for network situation awareness. While a few methods have been proposed to detect network intrusion, they cannot directly and effectively utilize semi-quantitative information consisting of expert knowledge and quantitative data. Hence, this paper proposes a new detection model based on a directed acyclic graph (DAG) and a belief rule

base (BRB). In the proposed model, called DAG-BRB, theDAG is employed to construct a multi-layered BRB model that can avoid explosion of combinations of rule number because of a large number of types of intrusion.
To obtain the optimal parameters of the DAG-BRB model, an improved constraint covariance matrix adaption evolution strategy (CMA-ES) is developed that can effectively solve the constraint problem in the BRB. A case study was used to test the efficiency of the proposed DAG-BRB. The results showed that compared with other detection models, the DAG-BRB model has a higher detection rateand can be used in real network.
Nowadays, IT organizations generate colossal amountsofdata.Handlingthesechunksofdata itself is critical in the IT world. Hence centralizing the log management system improves security thereby enhances data protection in an organization. Such enterprises require a high profiling tool that helps in managing the information and events data to improve the level of security. Security InformationandEventManagement(SIEM)isa procedure for security analysis that prominence an overview of security in an organization. SIEM tools collect, analyze, normalize and correlatesallfilesandanalyzedatacomingfrom the various device and give a centralized viewof logs.
This paper articulates an abstraction of SIEM tools and event correlation engines, furnishing a description of their technical comparative study, focusing on most popular SIEM tools and open source rule-based correlation engines and profiles them. Distributed computing has become an effective approach to enhance capabilities of an institution or organization and minimize requirements for additional resource. Inthisregard,thedistributedcomputinghelpsin broadening institutes IT capabilities. One needs to note that distributed computing is now integral part of most expanding IT business sector. It is considered novel and efficientmeans for expanding business. As more organizations and individuals start to use the cloud to store their data and applications, significant concerns have developed to protect sensitive data from external and internal attacks over internet.
Due to security concern many clients hesitate in relocating their sensitive data on the clouds, despite significant interest in cloud-based computing. Security is a significant issue, since data much of an organizations data provides a tempting target for hackers and those concerns will continue to diminish the development of distributed computing if not addressed. Therefore, this study presents a new test and insight into a honeypot.
It is a device that can be classified into two types:

Handling
1. researchhoneypots.

Handling honeypots are used to mitigate real life dangers:
A research honeypot is utilized as anexploration instrument to study and distinguish the dangers on the internet. Therefore, the primary aim of this research project is to do an intensive network security analysis through a virtualized honeypot for cloud servers to tempt an attacker and provide a new means of monitoring.

2. EXISTINGSYSTEM

EXISTINGAPPROACH:
Blameless Bayes and Principal Component Analysis (PCA) were been used with theKDD99 dataset by Almansob and Lomte. Similarly, PCA, SVM, and KDD99 were used Chithik and Rabbani for IDS. In Aljawarneh et Al's.Consequently, KDD99 is old and doesn't give any data about cutting edge new assault types, example, multi day misuses and so forth.

Drawbacks
- StrictRegulations
- Difficult to work with for non-technical users
- Restrictiveto resources
- Constantlyneeds Patching
- Constantlybeing attacked

3. PROPOSEDSYSTEM

ProposedSystem
Important steps of the algorithm are given in below.
- Normalizationofeverydataset.
- Convert that dataset into the testing and training.
- Form IDS models with the help of using RF, ANN, CNN and SVM algorithms.
- Evaluateeverymodel'sperformance

ALGORIT
HMS CNN:

AConvolutionalNeuralNetwork(CNN)isatypeof deeplearningneuralnetworkthatiswell-suitedfor imageandvideoanalysis.

Randomforest:

It is the tech industry's definitive destination for sharingcompelling,first-personaccountsofproblem-solvingontheroadtoinnovation.

ANN:

ANNalgorithmacceptsonlynumericandstructured data.ConvolutionalNeuralNetworks(CNN)and RecursiveNeuralNetworks(RNN)areusedtoaccept unstructuredandnon-numericdataformssuchas Image,Text,andSpeech

USECASEDIAGRAM

A use case diagram in the Unified Modeling Language (UML) is a type of behavioral diagram defined by and created from a Use-case analysis. Its purpose is to present a graphical overview of the functionality provided by a system in terms of actors, their goals (represented as use cases), and any dependencies between those use cases. The mainpurposeofausecasediagramistoshow what system functions are performed for which actor. Roles of the actors in the system can be depicted.

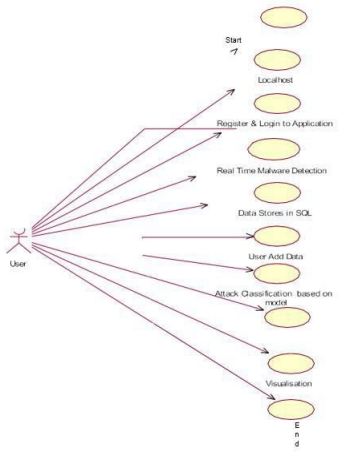

Fig4.1:UseCaseDiagram

CLASSDIAGRAM

In software engineering, a class diagram in the Unified Modeling Language (UML) is a type of static structure diagram that describes the structure of a system by showing the system's classes,theirattributes,operations(ormethods), and the relationships among the classes. It explains which class contains information.

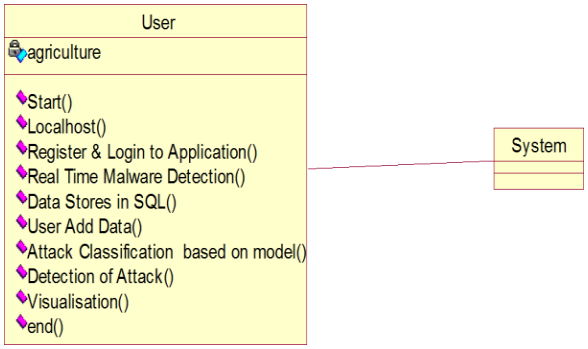

Fig4.2:ClassDiagram

SEQUENCEDIAGRAM

A sequence diagram in Unified Modeling Language (UML) is a kind of interaction diagram that shows how processes operate with one another and in what order. It is a constructof a Message Sequence Chart. Sequence diagrams are sometimes called event diagrams, event scenarios, and timing diagrams.

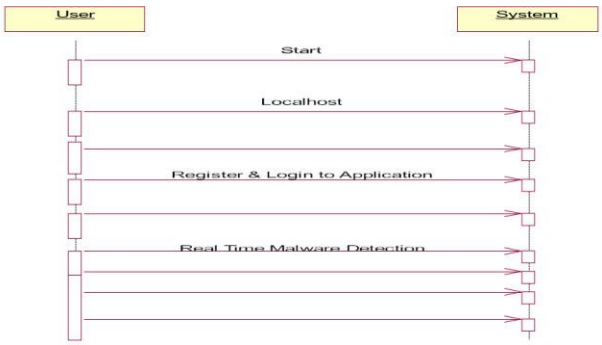

Fig4.3:Sequence Diagram

4 RESULTS

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
```

```
import itertools
import seaborn as sns
import pandas_profiling
import statsmodels.formula.api as sm
from statsmodels.stats.outliers_influence import variance_inflation_factor
from patsy import dmatrices
```

```
/usr/local/lib/python3.6/dist-packages/statsmodels/tools/_testing.py:19: FutureWarning: pandas.util.testing is dep
recated. Use the functions in the public API at pandas.testing instead.
  import pandas.util.testing as tm
```

```
from sklearn import datasets
from sklearn.feature_selection import RFE
import sklearn.metrics as metrics
from sklearn.svm import SVC
from sklearn.linear_model import LogisticRegression
from sklearn.feature_selection import SelectKBest
from sklearn.feature_selection import chi2, f_classif, mutual_info_classif
```

```
train=pd.read_csv('/content/drive/My Drive/kdd/NSL_Dataset/Train.txt',sep=',')
test=pd.read_csv('/content/drive/My Drive/kdd/NSL_Dataset/Test.txt',sep=',')
```

Fig7.1:InputData

```
columns=["duration","protocol_type","service","flag","src_bytes","dst_bytes","land",
"wrong_fragment","urgent","hot","num_failed_logins","logged_in",
"num_compromised","root_shell","su_attempted","num_root","num_file_creations",
"num_shells","num_access_files","num_outbound_cmds","is_host_login",
"is_guest_login","count","srv_count","serror_rate","srv_serror_rate",
"rerror_rate","srv_rerror_rate","same_srv_rate","diff_srv_rate","srv_diff_host_rate","dst_host_count","dst_host_sr
"dst_host_diff_srv_rate","dst_host_same_src_port_rate",
"dst_host_srv_diff_host_rate","dst_host_serror_rate","dst_host_srv_serror_rate",
"dst_host_rerror_rate","dst_host_srv_rerror_rate","attack", "last_flag"]
```

```
train.columns=columns
test.columns=columns
```

```
train.head()
```

	duration	protocol_type	service	flag	src_bytes	dst_bytes	land	wrong_fragment	urgent	hot	num_failed_logins	logged_in	num_compromised	root_s
0	0	udp	other	SF	146	0	0	0	0	0	0	0	0	
1	0	tcp	private	S0	0	0	0	0	0	0	0	0	0	
2	0	tcp	http	SF	232	8153	0	0	0	0	0	1	0	
3	0	tcp	http	SF	199	420	0	0	0	0	0	1	0	
4	0	tcp	private	REJ	0	0	0	0	0	0	0	0	0	

```
test.head()
```

Fig7.2:Data Processing

```
# Protocol type distribution
plt.figure(figsize=(9,8))
sns.countplot(x="protocol_type", data=train)
plt.show()
```

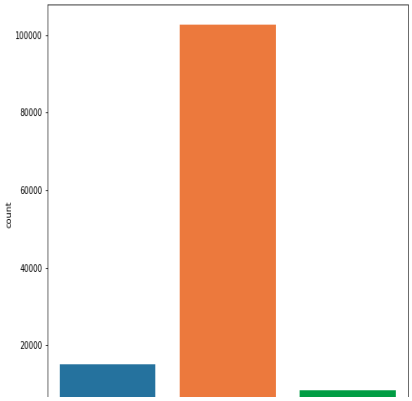

Fig7.3:DATAEDA

Model Building

```
train_X=train_new[cols]
train_y=train_new['attack_class']
test_X=test_new[cols]
test_y=test_new['attack_class']
```

Fig7.4:ModelBuilding

Logistic Regression

```
# Building Models
from sklearn.linear_model import LogisticRegression
logreg = LogisticRegression(random_state=0,solver='lbfgs',multi_class='multinomial')
logreg.fit( train_X, train_y)
logreg.predict(train_X)   #by default, it use cut-off as 0.5
```

```
list( zip( cols, logreg.coef_[0] ) )
```

```
logreg.intercept_
```

```
logreg.score(train_X,train_y)
```

Fig7.5: LogisticRegression

Decision Trees

```
train_X.shape
```

```
param_grid = {'max_depth': np.arange(2, 12),
              'max_features': np.arange(10,15)}
```

```
train_y.shape
```

```
from sklearn.model_selection import GridSearchCV
from sklearn.tree import DecisionTreeClassifier, export_graphviz, export
tree = GridSearchCV(DecisionTreeClassifier(), param_grid, cv = 10,verbose=1,n_jobs=-1)
tree.fit( train_X, train_y )
```

```
tree.best_score_
```

```
tree.best_estimator_
tree.best_params_
```

```
train_pred = tree.predict(train_X)
```

```
print(metrics.classification_report(train_y, train_pred))
```

```
test_pred = tree.predict(test_X)
```

Fig7.6:DecisionTrees

Random Forest

```
from sklearn.ensemble import RandomForestClassifier
pargrid_rf = {'n_estimators': [50,60,70,80,90,100],
              'max_features': [2,3,4,5,6,7]}
```

```
from sklearn.model_selection import GridSearchCV
gscv_rf = GridSearchCV(estimator=RandomForestClassifier(),
                       param_grid=pargrid_rf,
                       cv=10,
                       verbose=True, n_jobs=-1)

gscv_results = gscv_rf.fit(train_X, train_y)
```

```
gscv_results.best_params_
```

```
gscv_rf.best_score_
```

```
radm_clf = RandomForestClassifier(oob_score=True,n_estimators=80, max_features=5, n_jobs=-1)
radm_clf.fit( train_X, train_y )
```

```
radm_test_pred = pd.DataFrame( { 'actual':  test_y,
                                 'predicted': radm_clf.predict( test_X ) } )
```

Fig7.7:Random forest

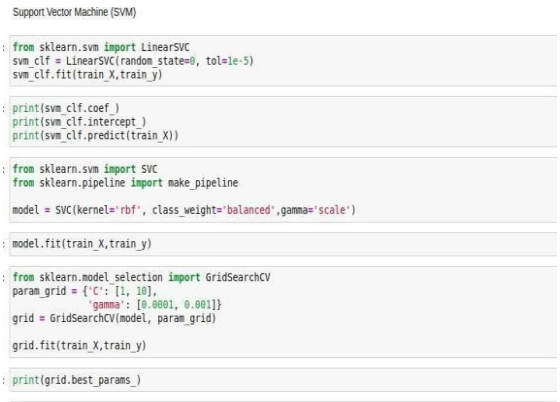

Fig7.8:SupportVectormachine (SVM)

Fig7.9:LocalHostincmdpython app.py

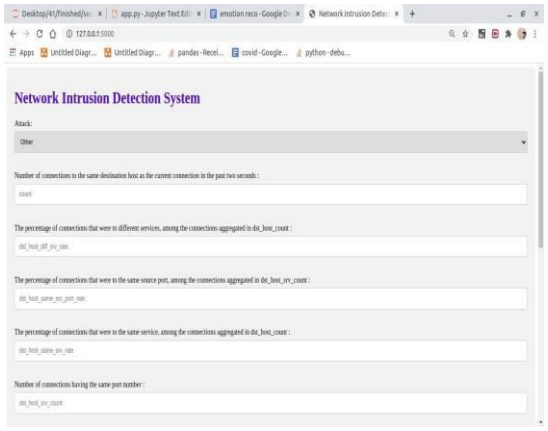

Fig7.10:NetworkIntrusionDetection System

Fig7.11: Enteringthe input data

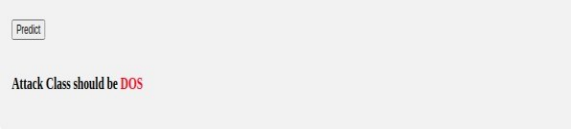

Fig7.12:Predict attack

4. CONCLUSIONANDFUTURESCOPE

CONCLUSION

Right now, estimations of help vector machine, ANN, CNN, Random Forest and profound learning calculations dependent on modern CICIDS2017 dataset were introduced relatively. Results show that the profound learning calculation performed fundamentally preferable outcomes over SVM, ANN, RF and CNN. We aregoingtoutilizeportsweependeavorsaswell as other assault types with AI and profound learning calculations, Apache Hadoop and sparkle innovations together dependent on this dataset later on.

All these calculation helps us to detect thecyber-attack in network. It happens in the way that when we consider long back years there maybesomanyattackshappenedsowhenthese attacksarerecognizedthenthefeaturesatwhich valuestheseattacksarehappeningwillbestored insomedatasets.Sobyusingthesedatasetswe

are going to predict whether cyber-attack isdone or not. These predictions can be done by four algorithms like SVM, ANN, RF, CNN this paper helps to identify which algorithm predicts the best accuracy rates which helps to predict best results to identify the cyber-attacks happened or not.

5.2FUTURESCOPE

The future scope of cyber threat detection using ANNs based on event profiles is promising,with potential benefits in terms of detection accuracy, real-time monitoring, adaptive learning, and scalability. However,organizations need to consider the challenges and invest in the necessary resources to implementandmaintaintheseadvancedsystems effectively.

REFERENCES

[1] K. Graves, Ceh: Official certified ethical hacker review guide: Exam 312-50. John Wiley & Sons, 2007.

[2] R. Christopher, "Port scanning techniques and the defense against them,"SANS Institute, 2001.

[3] M. Baykara, R. Das¸, and I. Karadogan, "Bilgig¨uvenli˘gisistemlerindekullanilanarac¸ larinincelenmesi," in 1st International Symposium on Digital Forensics and Security (ISDFS13), 2013,pp. 231–239.

[4] S. Staniford, J. A. Hoagland, and J. M. McAlerney, "Practical automated detection of stealthy portscans," Journal of Computer Security, vol. 10, no. 1-2, pp. 105–136, 2002.

[5] S. Robertson, E. V. Siegel, M. Miller, and S. J. Stolfo, "Surveillance detection in high bandwidth environments," in DARPA Information Survivability Conference and Exposition, 2003.Proceedings, vol. 1. IEEE, 2003, pp. 130–138.

[6] K. Ibrahimi and M. Ouaddane, "Management of intrusion detection systems based-kdd99: Analysis with lda and pca," in Wireless Networks and MobileCommunications (WINCOM),2017 InternationalConferenceon. IEEE,2017,pp.1– 6.

[7] N. Moustafa and J. Slay, "Thesignificant features of the unsw-nb15 and the kdd99 data sets for network intrusion detection systems," in Building Analysis Datasets and Gathering Experience Returns for Security (BADGERS), 2015 4th International Workshop on. IEEE, 2015, pp. 25–31.

[8] L.Sun,T.Anthony,H.Z.Xia,J.Chen, X. Huang, and Y. Zhang, "Detection and classification of malicious patterns in network traffic using Binford's law," in Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), 2017. IEEE, 2017, pp. 864– 872.

[9] S. M. Almansob and S. S. Lomte, "Addressing challenges for intrusion detection system usingnaive bayes and pca algorithm," in Convergence in Technology (I2CT), 2017 2nd International Conference for. IEEE, 2017, pp. 565–568.

[10] M. C. Raja and M. M. A. Rabbani, "Combined analysis of support vector machine and principle component analysis for ids," in IEEE International Conference on CommunicationandElectronicsSystems,2016, pp. 1–5.

[11] S. Aljawarneh, M. Aldwairi, and M. B. Yassein, "Anomaly-based intrusion detection system through feature selection analysis and building hybrid efficient model," Journal of ComputationalScience,vol.25,pp.152–160, 2018.

[12] I. Sharafaldin, A. H. Lashkari, and A. A. Ghorbani, "Toward generating a new intrusion detection dataset and intrusion traffic characterization."inICISSP,2018,pp.108– 116.

[13] D.Aksu,S.Ustebay,M.A.Aydin,and T. Atmaca, "Intrusion detection with comparative analysis of supervised learning techniques and fisher score feature selection algorithm," in International Symposium on Computer and Information Sciences. Springer, 2018, pp. 141– 149.

[14] N.Marir,H.Wang,G.Feng,B.Li,and M. Jia, "Distributed abnormal behaviordetection approach based on deep beliefnetwork and ensemble svm using spark," IEEE Access, 2018.

[15] P. A. A. Resende and A. C. Drummond, "Adaptive anomaly-based intrusion detection system using genetic algorithm and profiling," Securityand Privacy, vol. 1, no. 4, p. e36, 2018.

[16] C. Cortes and V. Vapnik, "Support-vectornetworks,"Machinelearning,vol.20,no. 3,pp. 273– 297, 1995.

[17] R. Shouval, O. Bondi, H. Mishan, A. Shimoni,R. Unger,and A. Nagler, "Application of machine learning algorithms for clinical predictive modeling: a data-mining approach in sct," Bone marrow transplantation, vol. 49, no. 3, p. 332, 2014.