Cukurova University
Journal of
Natural & Applied Sciences

Original Paper

Anomaly Detection in Network Traffic Using Machine Learning

Roaa MOHAMMED[1], M. Fatih AKAY[1]

[1] Cukurova University, Department of Computer Engineering, Adana, Turkey

ORCID IDs of the authors: R.M. 0009-0001-2973-2470; M. F. A. 0000-0003-0780-0679.

Cite this article as: Mohammed, R., Akay, M.F. (2023). Anomaly Detection in Network Traffic Using Machine Learning, Cukurova University Journal of Natural & Applied Sciences 2(3): 5-12

Abstract

The primary theme of this paper revolves around the detecting anomalies and measuring the device health for Central Processing Unit (CPU), memory utilization, and allocation; for Key Performance Indicator (KPI) dataset which assembled throw twenty-one-day, by improving models using machine learning (ML) methods; namely, Convolutional Neural Network (CNN), and Long Short-Term Memory (LSTM), with Auto Encoders (AE), One-Class Support Vector Machine (Oc-SVM), also k-Nearest Neighbors (k-NN). The accuracy of all methods was measured by using a confusion matrix. According to the observed results, the deep learning methods yield great performance results compared to classification methods for all models. In general, CNN/AE and LSTM/AE models show higher accuracy than the other methods. The ranking of models from best to worst based on accuracy in the confusion matrix are; CNN/AE, LSTM/AE, as for the deep learning models, while for classification models the favorable order for the methods are; k-NN, and Oc-SVM.

Keywords: Anomaly detection, Machine learning (ML)

1. Introduction

Anomaly detection involves identifying points or errors that distort data, which can manifest as frauds, defects, or errors. In the context of network traffic, anomalies can occur suddenly and unpredictably, requiring a high level of preparedness to detect and protect the network early.

Over the years, various methods have been developed to detect anomalies, including Recurrent Neural Networks (RNN), Convolutional Neural Networks (CNN), Support Vector Machines (SVM), OPTICS, DBSCAN, Ward Hierarchical Clustering, and Mean-Shift. These techniques can be classified into supervised, unsupervised, or semi-supervised learning methods, collectively known as anomaly detection techniques. Supervised techniques require labeled and classified datasets as normal and abnormal, and iterative training is used to detect anomalies.

In contrast, unsupervised algorithms use unlabeled test datasets, where abnormal data are instances that do not fit with the remainder of the dataset. Semi-supervised methods create a model with a normal attitude from a normal training dataset and test the likelihood of a test pattern generated by the utilized model. The detection of outlier data in network traffic is a significant challenge today, and this thesis aims to detect such data using machine learning techniques. The models utilized for this purpose are CNN/AE, LSTM/AE, Oc-SVM, and k-NN. CNN/AE and LSTM/AE are deep learning models that use artificial neural networks to effectively label and signal for unlabeled data, while Oc-SVM and k-NN are unsupervised and supervised learning methods,

Respectively. In this study, the dataset was split into 70% training data and 30% testing data using hold-out and k-fold techniques. Grid-search was used to tune different hyperparameters for the dataset to develop classification and deep learning models. The performance of all models was evaluated using the confusion matrix, and the results indicate that deep learning models outperformed classification models across all methods. Specifically, CNN/AE and LSTM/AE models exhibited better performance than other methods.

Address for Correspondence:
Roaa Mohammed, e-mail: roaarafih@gmail.com.

Received: Jul 2, 2023
Accepted: Nov 20, 2023

The favorable order of methods for the prediction model based on accuracy is CNN/AE and LSTM/AE, while for the classification model, the favorable order is k-NN and Oc-SVM. This study provides valuable insights into the use of machine learning techniques for detecting anomalies in network traffic, with potential applications in various fields.

The study utilizes several key elements that enable the application of methods containing multiple features simultaneously. These elements are listed as follows:

1. Dataset was gathered; Key Performance Indicator (KPI) for daily CPU usage, as well as DISK, and MEMORY usages in network traffic.
2. Two different approaches to detection were suggested; two models of deep learning-based methods are; CNN/AE, and LSTM/AE. On the other hand; two models of classification-based methods are; Oc-SVM, and k-NN.
3. Defined the part of the dataset which will be used for train/test the models by two techniques, the 1st technique is hold-out; (70 ratio) for train, and (30 ratio) for test, and the 2nd technique is k-fold; (10 n_sample) for train, and test.
4. Applied different values of parameters for detection methods to obtain the best hyperparameters depend on accuracy by use grid search technique.
5. High Accuracy (ACC) rates results were compared; based on Confusion Matrix (CM) for each model.
6. The model that has a high accuracy was compared; with previous works to find the performance and better accuracy (ACC) to detect anomalous and recognizing the health of devices.

2. Material and Method

2.1. Dataset

Through this, an intrusion prevention system can determine the unusual activities and define it as a cyber attack. In addition, it can also send an alert to selected administrators or technicians when it identifies suspicious activities. This allows them to quickly identify the root causes of the problem and prevent further attacks. All kinds of Networks are often exposed to anomalous behavior. For example, attacks or large data transfers in IP networks, the presence of intruders in surveillance systems, and sudden congestion in the network, so one of the biggest challenges that both network administrators and researchers face is detecting anomalies in network traffic, where the rapid accurate detection of anomalies is important to prevent many serious problems.
To ensure the decives in order the effectively detect anomalies in network traffic, it is imperative to monitor various Key Performance Indicators (KPIs) related to device utilization and allocation, to ensure consistent and reliable services. KPIs are time-series data that are generated through continuous measurements over time, such as network throughput, serving latency, page views, and the number of online users. However, due to their inherent complexity and lack of diversity, including periodic, stable, and fluctuating patterns, detecting anomalies in KPIs has been a significant challenge. Therefore, developing effective anomaly detection techniques for various KPIs has become a critical research area, with the potential to significantly improve decision-making and operational efficiency across diverse domains. In this work, a dataset was prepared using KPIs related to CPU and memory utilization, as well as memory allocation, to measure device health. The datasets were aggregated every ten minutes over a period of 24 hours for twenty-one days. The range index comprised 6300 entries, ranging from 0 to 6299, and the data columns totaled four, as illustrated in Fig.1 This study presents valuable insights into the use of KPIs for detecting anomalies in network traffic, with potential applications in various fields.

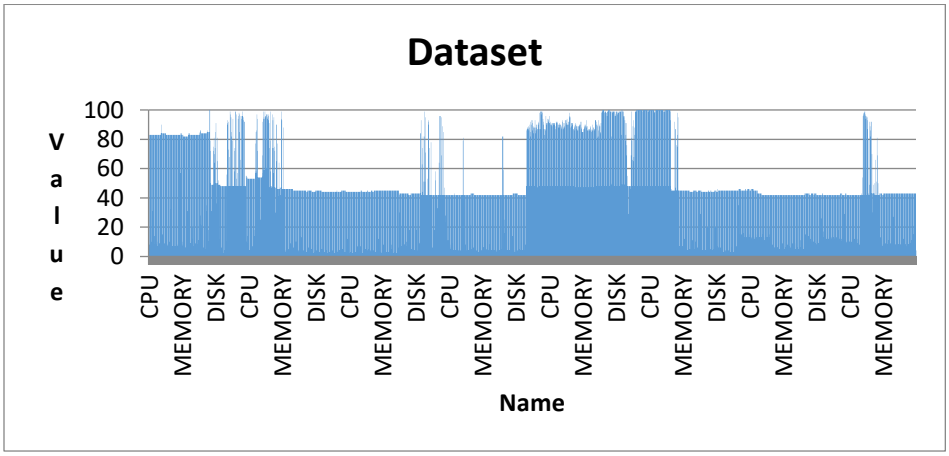

Figure 1. KPI datasets.

2.2. Methods

2.2.1 One Class SVM (Oc-SVM)

One Class SVM (Oc-SVM) was usually defined as a popular unsupervised method to disclose anomalies [1], The performance of the Oc-SVM model is dependent on the hyperparameter that controls the sensitivity of the support vectors used to separate data points from the origin in feature space F while maximizing the distance from the hyperplane to the origin. The model is trained on 'normal' data to learn the boundaries of these points. It can then be used to classify outliers that lie outside the boundary. Table 1 specifies the parameter value intervals for the Oc-SVM method.

Table 1. The intervals for the parameters and values of the Oc-SVM methods.

Parameters	Values
Hyper parameter (nu)	0.5
Gamma	Auto
Kernel	Rbf

2.2.2 k-Nearest Neighbors (k-NN)

k-NN, which is a well known and used classification algorithm and consider one of the simplest supervised ML methods, proposed by Evelyn Fix and Joseph Hodges in 1951 [2], it used widely for classification and also occasionally using for regression problems, So; It is well used to detect the anomolious and building recommender systems etc. The main work of this algorithm is that depends on observations that are close and similar to each other are considered normal and which are distant from the cluster of similar observations and which be single are considered abnormal observations. The classification accuracy of the k-Nearest Neighbors (k-NN) method is heavily influenced by the choice of distance metric used. This method classifies an object by identifying its k nearest neighbors and assigning it to the category that is most common among them. The value of k is typically a positive integer. Table 2 displays the parameter value intervals for the k-NN method.

Table 2. The intervals for the parameters and values of the k-NN method.

Parameters	Values
n-neighbors	5
Weights	Uniform

2.2.3 Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM), is an optimization method used to find artificial neural networks weights to avoid long-term dependency problems [3]. this method was explicitly proposed as a solution to avoid the vanishing gradient problem in recurrent neural networks (RNN) [4]. The performance of an LSTM model is dependent on the number of neurons in the hidden layer and the number of epochs used during training. Increasing the number of neurons can improve the model's ability to learn the structure of the problem, but may lead to overfitting and longer training times. The number of epochs should be chosen such that the gap between the test error and the training error is minimized. Table 3 provides the parameter value intervals for LSTM models.

Table 3. The intervals for the parameters and values of the LSTM/AE models.

Parameters	Values
Mean	None
Sd	None
Units	64-32
Rate	0.5-0.2
Return_sequences	True

2.2.4 Convolutional Neural Network (CNN)

Convolutional Neural Network (CNN), is one of the most famous neural networks after the best results were shared on the ImageNet Large Scale Visual Recognition Competition (ILSVRC), [5, 6]. The CNN method was inspired by the visual cortex for the eye, that is responsible for detecting light in small parts of the visual partions [7]; the visual cortex consists of, two kinds of layers are convolution and sampling layers used to extract, and map features sequentially as in CNN. The input assumed is a two-dimensional image with a set of pixels. For this reason, CNN has been basically used to classify images, and network security [8]. Optimizing the performance of CNN models can be achieved through careful selection of the number of neurons and activation function, where a small number of neurons may lead to underfitting and an excessive number of neurons could lead to overfitting. The Rectified Linear Units (ReLU) activation function has become the industry standard due to its superior performance and ease

of training. The use of a flatten layer between convolutional and dense layers can reduce feature maps to a one-dimensional vector. Table 4 displays the parameter value intervals for CNN/AE models.

Table 4. The intervals for the parameters and values of the CNN/AE models.

Parameters	Values
Mean	None
Sd	None
Kernel_size	1
Padding	'Valid'
Strides	1
Activation	'relu'
Input_dim	3
Activation	'sigmoid'

2.3. Performance Metrics

As a standard evaluation metric for Intrusion Detection Systems (IDSs), accuracy is considered as a metric that estimates the overall percentages of detection and false alarms an IDS model produces, which reflects the overall success rate of any IDS. The accuracy is evaluated by the confusion matrix, which includes terms such as; the T..rue-P..ositive (TP), and T..rue-N..egative (TN) that denote correctly predicted conditions, whilst the False-Positive (FP), and False-Negative (FN) misclassified ones.In other words, TPs and TNs refer to correctly classified attack and normal records, respectively, and conversely, FPs and FNs refer to misclassified normal and attack records, respectively as shown in Fig.2.

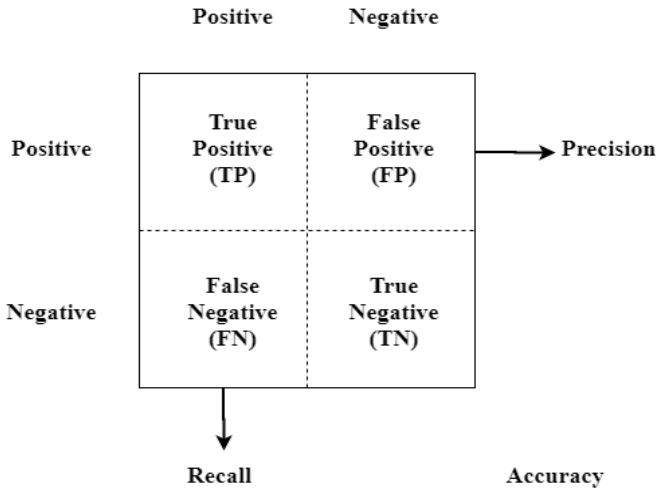

Figure 2. Confusion Matrix.

These four terms are used to calculate A.ccuracy, P.recision, R.ecall, and A.rea U.nder C.urve (AUC) as elucidated in formulas from (1) to (7). The strength of a system to handle network traffic with high accuracy and low level loss while running in real-time is evaluated by performance.

$$\text{Accuracy} = \frac{\Sigma TN+TP}{\Sigma TP+FP+TN+FN} \quad (1)$$

$$\text{Precision} = \frac{\Sigma TP}{\Sigma TP+FP} \quad (2)$$

$$\text{Recall} = \frac{\Sigma TP}{\Sigma TP+FN} \quad (3)$$

$$\text{True Positive Rate (TPR)} = \frac{\Sigma TP}{\Sigma TP+FN} \quad (4)$$

$$\text{False Positive Rate (FPR)} = \frac{\Sigma FP}{\Sigma FP+TN} \quad (5)$$

$$\text{True Negative Rate (TNR)} = \frac{\Sigma TN}{\Sigma TN+FP} \quad (6)$$

$$\text{False Negative Rate (FNR)} = \frac{\Sigma FN}{\Sigma FN+TP} \quad (7)$$

3. Results and Discussion

In this paper, four different methods including Oc-SVM, k-NN, CNN/AE, and LSTM/AE were used to classify any kind of anomalies in network traffic, and find out which method has higher accuracy to ensure the health of the device.Two techniques were applied to splits the dataset for each method in order to adopt classification models, and deep learning models. the 1st technique; one combination of the hold-out split was used. The split ambit is 70% for training and 30% for testing, respectively. While the 2nd technique; 10 of k-fold splits have been used to build machine learning models. Taking into consideration that these techniques were applied with grid search to tunning hyperparameters to get the best parameter for each method.

Tables 5, to 10 show the detection models' results of three experiments for CPU, Memory, and Disk are; key performance indicator datasets. We have developed machine learning models using KPI dataset that is labeled as normal and abnormal. The observations that have a reconstruction error greater than the threshold will be classified as freeze, whereas the ones with a reconstruction error less than the threshold as non-freeze, traffic data. We have illustrated the efficacy of the proposed approach by reporting different hold-out and k-fold values and representing their impact on Accuracy, Area Under Curve (AUC), Precision, and Recall. Tables 5, to 10 summarize the evaluation metrics performance for different values in terms.

Table 5. CPU Hold-out, split range (70% TRAIN - 30% TEST)

Model Name	Accuracy	Auc	Precision	Recall	Best Model Parameter
Oc-SVM	0.620	0.390	0.600	0.610	Kernel = 'rbf', Gamma = 'auto', nu = 0.1
k-NN	0.900	0.900	0.990	0.990	N_neighbors = 3
CNN/AE	0.996	0.998	0.990	0.998	Epoch = 50, batch size = 20
LSTM/AE	0.989	0.996	0.971	0.996	Epoch = 100, batch size = 60

Table 6. CPU KFold, split range (10)

Model Name	Accuracy	Auc	Precision	Recall	Best Model Parameter
Oc-SVM	0.430	0.560	0.440	0.440	Kernel = 'rbf', Gamma = 'auto', nu = 0.3
k-NN	0.980	0.990	0.990	0.980	N_neighbors = 3
CNN/AE	0.997	0.997	0.979	0.999	Epoch =50, batch size = 20
LSTM/AE	0.992	0.999	0.992	0.998	Epoch = 90, batch size = 30

Table 7. MEMORY Hold-out, split range (70% TRAIN - 30% TEST)

Model Name	Accuracy	Auc	Precision	Recall	Best Model Parameter
Oc-SVM	0.320	0.660	0.350	0.340	Kernel = 'rbf', Gamma = 'auto', nu = 0.6
k-NN	0.900	0.990	0.990	0.990	N_neighbors = 2
CNN/AE	0.996	0.999	0.989	1.000	Epoch = 30, batch size = 20
LSTM/AE	0.994	0.999	0.983	0.996	Epoch = 60, batch size = 40

Table 8. MEMORY KFold, split range (10)

Model Name	Accuracy	Auc	Precision	Recall	Best Model Parameter
Oc-SVM	0.590	0.430	0.560	0.570	Kernel = 'rbf', Gamma = 'auto', nu = 0.1
k-NN	0.970	0.999	1.000	1.000	N_neighbors = 2
CNN/AE	0.995	0.999	0.986	1.000	Epoch =70, batch size = 40
LSTM/AE	0.995	0.999	0.986	0.998	Epoch = 90, batch size = 30

Table 9. DISK Hold-out, split range (70%-30%)

Model Name	Accuracy	Auc	Precision	Recall	Best Model Parameter
Oc-SVM	0.570	0.430	0.560	0.570	Kernel = 'rbf', Gamma = 'auto', nu = 0.1
k-NN	0.900	0.990	1.000	1.000	N_neighbors = 3
CNN/AE	0.989	0.997	0.968	1.000	Epoch = 100, batch size = 30
LSTM/AE	0.976	0.996	0.933	0.994	Epoch = 100, batch size = 60

Table 10. DISK KFold, split range (10)

Model Name	Accuracy	Auc	Precision	Recall	Best Model Parameter
Oc-SVM	0.470	0.510	0.500	0.490	Kernel = 'rbf', Gamma = 'auto', nu = 0.1
k-NN	0.960	0.999	1.000	1.000	N_neighbors = 2
CNN/AE	0.998	0.999	0.987	1.000	Epoch = 60, batch size = 16
LSTM/AE	0.995	0.999	0.996	1.000	Epoch = 100, batch size = 16

Fig.3 to Fig.5 illustrate average Accuracy of the Oc-SVM, k-NN, LSTM/AE, and CNN/AE methods of anomaly detection models separately. The accuracy rates between hold-out and k-fold in the whole dataset on the average, and the other paradigms; while the Figures show the average accuracy of all methods for anomaly detecting models.

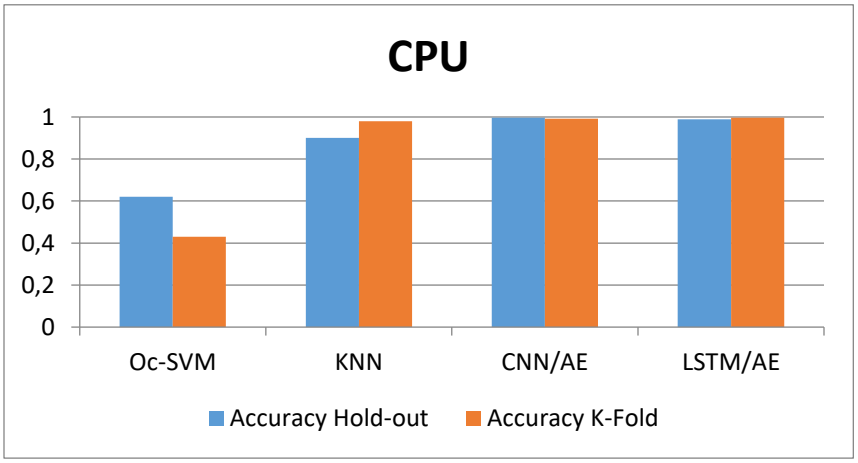

Figure 3. Accuracy of CPU, split range (70%), k-fold (10).

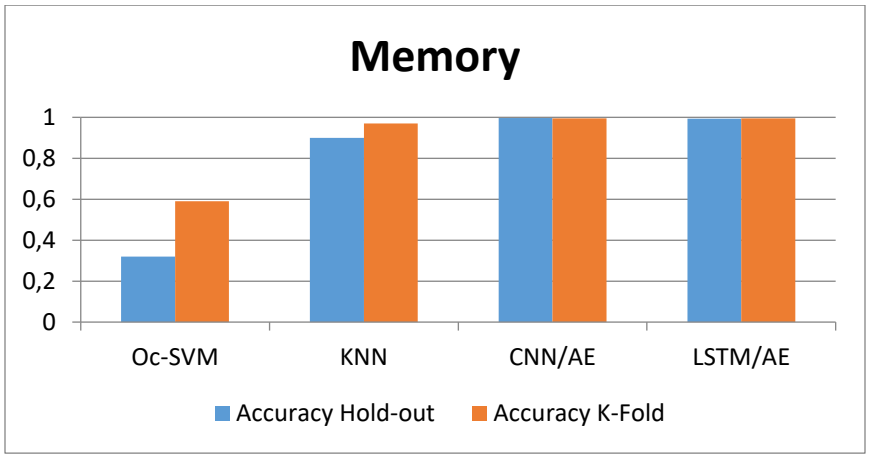

Figure 4. Accuracy of Memory, split range (70%), k-fold (10).

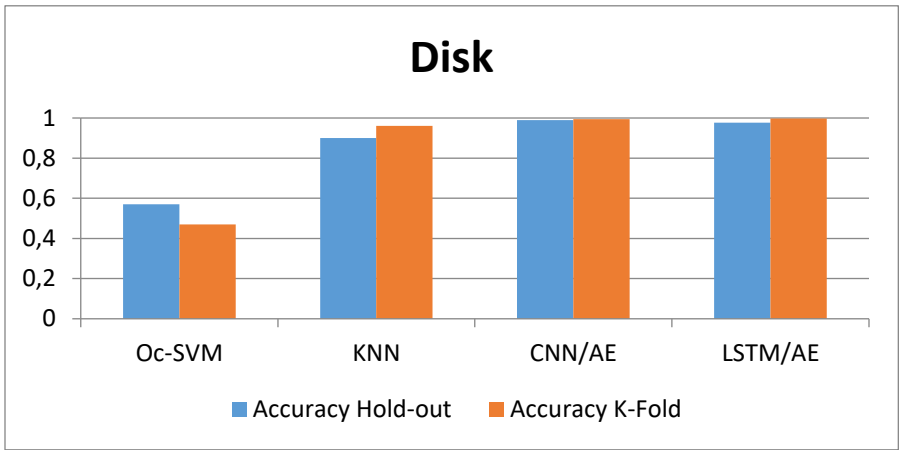

Figure 5. Accuracy of Disk, split range (70%), k-fold (10).

3.1.Discussion of the Comparison with Previous Works

In this section, obtained results of developed models were discussed and compared with previous works to make sure that classification models, and deep learning models have parallel performance with previous works. The comparison was summarized in Table 11.

Table 11. Comparing the results of accuracy for various techniques.

Ser.	Methods	Performance Metric	Values
#1	LSTM, [9]	ACC	93.82 %
#2	LSSVM-IDS, [10]	ACC	97.33 %
#3	ODC, [11]	ACC	98.30 %
#4	AE/ k-Means, [11]	ACC	82.93 %
#5	HAST-NAD, [12]	ACC	99.91 %
#6	*CNN/AE [Our Studies]*	*ACC*	*99.98 %*
#7	*LSTM/AE [Our Studies]*	*ACC*	*99.97 %*
#8	*k-NN [Our Studies]*	*ACC*	*95.00 %*

As illustrated in the table, the accuracy ratio for Convolutional Neural Network with Auto Encoder (CNN/AE) and Long Short-Term Memory with Auto Encoder (LSTM/AE) models is higher than that of Long Short-Term Memory (LSTM). [9]; Least Square Support Vector Machine-Intrusion Detection System **(LSSVM-IDS)**, [10]; Optimized Deep Clustering **(ODC)**, and Auto Encoder **(AE)** with Kmeans, [11]; Hierarchical Spatiotemporal Feature Learning **(HAST-NAD)**, [12]; wherein the values of Accuracy **(ACC)** are; 99.98 %, 99.97 %, 93.82 %, 97.33 %, 99.45 %, 98.30 %, 82.93 %, and 99.91%, respectively; This is in line with the findings of [12], indicating that the models learn more features and improve the accuracy rate.

The results confirm that the *CNN/AE* and *LSTM/AE* classifiers provide superior performance in detecting anomalies and recognizing the health of devices, when compared to the results previously published for strong static classifiers in terms of Accuracy.

4. Conclusion

In conclusion, our study demonstrates the effectiveness of machine learning methods such as LSTM/AE, CNN/AE, Oc-SVM, and k-NN in detecting anomalies and recognizing the health of devices. Our results suggest that deep learning models, particularly CNN/AE models, outperform classification models in detecting anomalies. These findings have the potential to contribute to decision-making and operational efficiency across diverse domains, with applications in network traffic and beyond.

The main objective of this study is to employ accurate models for detecting anomaly and recognize the health of devices, by CPU, MEMORY, and DISK rates using Machine learning methods including LSTM/AE, CNN/AE, Oc-SVM, and k-NN to test KPI daily datasets. Two different detection approaches are created by the development of environment, which are the classification models, and deep learning models.

For the detection model (train-test%), and (k)fold that used CPU, Memory, and Disk datasets have been partitioned into two different ranges. In the first division, 70% of dataset content has been utilized for training while the rest of the dataset has been

utilized for testing. Additionally, 10 of k-fold also used to train and test divisions have been utilized as second partitioning ranges for datasets.

The performance of all models were evaluated by calculating the value of the confusion matrix. A total of 24 different models were developed, including 12 classification models and 12 deep learning models.

According to the obtained results, the deep learning model produced better results compared to the classification models of all methods. In general, CNN/AE models show better performance than models developed by other methods. The order of the methods for the prediction model in terms of accuracy-based performance, from best to worst, is CNN/AE, LSTM/AE, k-NN, and Oc-SVM, while for the classification model, the order of favor of these methods is shown as k-NN , and Oc-SVM.

In general, when recognizing the health of the device, we deducting that the device is good during all the execution time, and also the memory has some of bozening. the order in terms of speed of the methods producing detection models is based on execution time, from fast to slow, CNN/AE, LSTM/AE, k-NN, and Oc-SVM.

As a suggestion for future work, a similar study can be conducted in several different areas by employing a hyperdeep learning neural networks methods or models to detect online attacks early with high accuracy.

References

[1] Raghavendra Chalapathy, Aditya Krishna Menon, Sanjay Chawla, (2019). " ANOMALY DETECTION USING ONE-CLASS NEURAL NETWORKS.

[2] Evelyn Fix, Joseph Hodges L., (2020). " Discriminatory Analysis. Nonparametric Discrimination: Consistency Properties ".USAF School of Aviation Medicine, Randolph Field, Texas. Archived (PDF) from the original on 26, September.

[3] Sara A. Althubiti , Eric Marcell Jones Jr, Kaushik Roy, (2018)." LSTM for Anomaly-Based Network Intrusion Detection ".978-1-5386-7177-1/18, IEEE

[4] Felix A. Gers, Jurgen Schmidhuber, Fred Cummins, (2000)." Learning to Forget: Continual Prediction with LSTM". Neural Computation 12, 2451–2471 °c 2000 Massachusetts Institute of Technology.

[5] Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton, (2017)." ImageNet Classification with Deep Convolutional Neural Networks", communıcatıons of the acm, june, vol. 60, no. 6.

[6] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, et. al., (2015)."ImageNet Large Scale Visual Recognition Challenge ", Int J Comput Vis 115:211–252 DOI 10.1007/s11263-015-0816-y,

[7] Weibo Liu, Zidong Wang, Xiaohui Liu, Nianyin Zeng, Yurong Liu, Fuad E. Alsaadi, (2017). "A survey of deep neural netwo Felix A. Gers rk architectures and their applications", Neurocomputing 234, 11–26.

[8] Taejoon Kim, Sang C. Suh, Hyunjoo Kim, Jonghyun Kim, Jinoh Kim, (2018). " An Encoding Technique for CNN-based Network Anomaly Detection", 978-1-5386-5035-6/18 ©IEEE.

[9] Ralf C. Staudemeyer, (2015). "Applying long short-term memory recurrent neural networks to intrusion detection", SACJ No. 56.

[10] M. A. Ambusaidi, X. He, P. Nanda, Z. Tan, (2016). "Building an intrusion detection system using a filter-based feature selection algorithm," IEEE Transactions on Computers, vol. 65, no. 10, pp. 2986–2998.

[11] Annie Gilda Roselin et.al., (2021). "Intelligent Anomaly Detection for Large Network Traffic With Optimized Deep Clustering (ODC) Algorithm", 10.1109/ACCESS.2021.3068172.

[12] Guanglu Wei et.al., (2021). "Adoption and realization of deep learning in network traffic anomaly detection device design". Soft computing 25:1147–1158.