Article

Symmetrical Simulation Scheme for Anomaly Detection in Autonomous Vehicles Based on LSTM Model

Abdulaziz A. Alsulami[1], Qasem Abu Al-Haija [2,*] , Ali Alqahtani[3] and Raed Alsini[4]

[1] Department of Information Systems, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah 21589, Saudi Arabia; aaalsulami10@kau.edu.sa (A.A.A)
[2] Department of Company Science/ Cybersecurity, Princess Sumaya University for Technology (PSUT), Amman, Jordan; q.abualhaija@psut.edu.jo (Q.A.A-H)
[3] Department of Networks and Communications Engineering, College of Computer Science and Information Systems, Najran University, Najran 61441, Saudi Arabia; asalqahtany@nu.edu.sa (A.A.Q.)
[4] Department of Information Systems, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah 21589, Saudi Arabia; ralsinie@kau.edu.sa (R.A.S.)

* Correspondence: q.abualhaija@psut.edu.jo

Abstract: Nowadays, technological advancement has transformed traditional vehicles into Autonomous Vehicles (A.V.s). In addition, in our daily lives, A.V.s play an important role since they are considered an essential component of smart cities. A.V. is an intelligent vehicle capable of maintaining safe driving by avoiding crashes caused by drivers. Unlike traditional vehicles, which are fully controlled and operated by humans, A.V.s collect information about the outside environment using sensors to ensure safe navigation. Furthermore, A.V.s reduce environmental impact because they usually use electricity to operate instead of fossil fuel, thus decreasing the greenhouse gasses. However, A.V.s could be threatened by cyberattacks, posing risks to human life. For example, researchers reported that Wi-Fi technology could be vulnerable to cyberattacks through Tesla and BMW AVs. Therefore, more research is needed to detect cyberattacks targeting the components of A.V.s to mitigate their negative consequences. This research will contribute to the security of A.V.s by detecting cyberattacks at the early stages. First, we inject False Data Injection (FDI) attacks into an A.V. simulation-based system developed by MathWorks. Inc. Second, we collect the dataset generated from the simulation model after integrating the cyberattack. Third, we implement an intelligent symmetrical anomaly detection method to identify FDI attacks targeting the control system of the A.V. through a compromised sensor. We use long short-term memory (LSTM) deep networks to detect FDI attacks in the early stage to ensure the stability of the operation of A.V.s. Our method classifies the collected dataset into two classifications: normal and anomaly data. The experimental result shows that our proposed model's accuracy is 99.95%. To this end, the proposed model outperforms other state-of-the-art models in the same study area.

Keywords: Autonomous vehicles (A.V.); Anomaly Detection (A.D.); Deep Learning (DL); Symmetry; Long Short-Term Memory (LSTM); False Data Injection (FDI) Attacks

1. Introduction

Recently, the number of operated autonomous vehicles (A.V.s) has increased significantly worldwide. It was reported that about 773,600 AVs were sold worldwide in 2016 [1]. In India, fossil fuel cars will be prohibited by 2030 because of their negative impact on the environment [2]. While the production and demand for the green energy source is an increase with the addition of renewable energy capacity in the subsequent years [3], A.V. is one such technology that uses renewable energy sources instead of fossil fuel sources which reduces greenhouse gas emissions produced by conventional fossil fuel vehicles [4].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 July 2022 doi:10.20944/preprints202207.0039.v1

Similar to any cyber-physical system (C.P.S.), A.V.s imply communication network infrastructure to transmit critical information in real-time. Therefore, integrating communication networks in A.V. introduces many benefits, such as exchanging information between embedded devices, sensors, actuators, and other technology. This is important to ensure the requirement for high connectivity between such synchronous cyber-physical systems [5]. In addition, integrating a communication network into physical components allows remote control and resource management [6]. This will enhance energy efficiency and consumption and make it much more convenient for A.V. owners to find available charging stations [6]. However, A.V.s can be exposed to cyber-attacks which cause negative impacts on the stability of the system. Cyberattacks can be launched stochastically anytime and anywhere [7] to target the connected devices of the A.V. whenever attackers find vulnerabilities in the system. Their impact is not limited to a single component of A.V., such as the control system but can involve the whole powertrain [8]. Besides, the effect of the attack can be observed in the short and long term.

Like any man-made object, in which symmetry is one of its main signatures [9], a typical A.V. includes symmetrical sensors to perceive the nearby environment, which must communicate effectively [10][11]. Therefore, sensors are the eyes of the A.V. onboard computer, which regularly provide the location, speed, and updates on nearby environments. In addition to that, A.V. is capable of exchanging data with other vehicles (V2V), pedestrians (V2P), and Infrastructures (V2I) [12]. The electronic control unit (E.C.U.) processes the measurements coming from sensors and transmits commands to actuators to control the devices nearby the vehicles [13]. E.C.U. Processes those measurements via software, which is vulnerable to an adversary. Imagine that E.V. contains many E.C.U.s, making it harder now to detect flaws in the software. Such adversaries, with their diversity, are collectively known as cyber-attacks or intrusions.

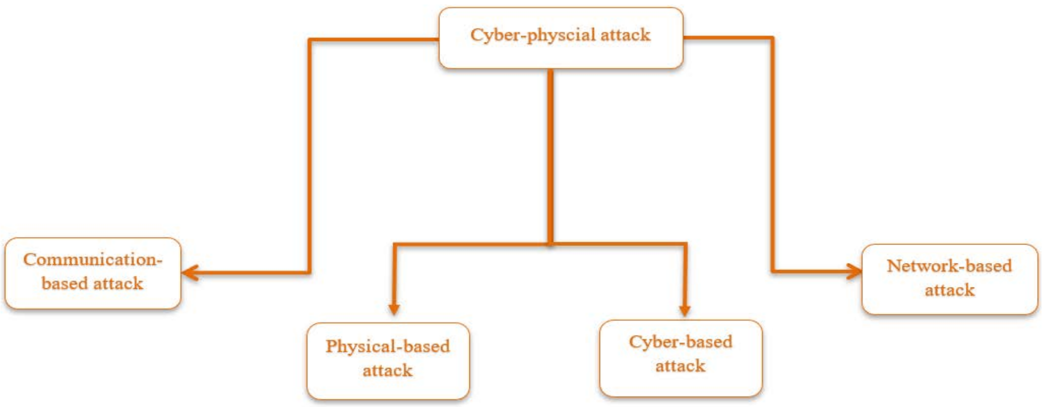

Figure 1. Classification of cyber-physical attacks

Network-based attacks, communication-based attacks, cyber-based attacks, and physical-based attacks are the taxonomy of Cyber-physical (C.P.) attacks, as displayed in Figure 1 [14]. The network-based attack involves passive and active attacks. A threat actor can compromise network security by gaining access to a node or nodes other than those under hijack. A.V.s networks also are susceptible to different types of attacks, such as access attacks, ransomware (RANSOM) attacks [15], denial of service (DoS) attacks, and reconnaissance attacks[14]. In communication-based attacks, the A.V.s rely heavily on sensors to exchange data with other sensors through a compromised communication channel. The attacker compromises the communication channel utilized for data exchange and transmits false data. Suppose the false data is shared over a network and reaches the relayed data node to the controller. In that case, a network-based attack takes place and affects the whole nodes associated with the relayed data node [13]. This kind of attack leads to cyber and physical catastrophic impacts. Therefore, it breaches the data integrity

 doi:10.20944/preprints202207.0039.v1

and shares the false data with the affected nodes [13] [16]. The cyber-based attack involves changing the system's code to a new code that can serve the adversary's plan. The most common attacks in A.V.s are malware injection attacks, FDI attacks, supply chain attacks, database manipulation attacks, and password cracking attacks [17]. In a physical-based attack, the adversary attempts to provide abnormal measurements to damage the physical device, such as the control system of an A.V. Many types of research have been conducted in anomaly detecting cyber-physical attacks and can be found in [18]. Understanding how a physical system works is crucial to building a predictive model to detect any malicious data that can damage the system. For instance, the programmable logic controller is widely used in A.V.s and is susceptible to an attack on Iran's nuclear power, known as the Stuxnet attack.

An intrusion detection system (I.D.S.) can be implemented into the onboard computer to detect any security flaws, monitor the system's events, and report incidents that violate the security policy [19]. However, like image recognition, in cybersecurity, more than 99% of new intrusions are symmetrical with very small mutations of previously existing ones [20]. This requires the development of very accurate I.D.S.s with high sensitivity in detecting cyber-attacks.

Advances have greatly influenced the development of self-driving vehicles in computing. For safety and speed reasons, all constraints must be considered when simulating the self-driving vehicle model. This simulation extensively uses deep learning models and strategies, making it possible to test an automated driving model. An approach such as long short-term memory (LSTM) has efficiently simulated the system. The LSTM is based on the symmetry of recurrent neural networks (R.N.N.).

As a result, academics in the autonomous system have quickly adopted it as a problem-solving method using deep learning. For example, LSTM can be used to predict a pedestrian's path and vehicle destination at an intersection. The effectiveness of LSTMs in time series prediction has been well established [11]. The ability of an LSTM network to predict the path of on-road vehicles is required for safe autonomous overtaking or lane changes. Furthermore, as shown in [12-14], several research articles investigated the advantages of using the LSTM in various systems

One of the most common C.P. attacks is the FDI attack, which involves fabricating the data and keeping the system's code the same. FDI attack is present in all classifications of C.P. attacks and can threaten A.V.s' applications, systems, and network layers [11]. Generally, it is challenging to detect the FDI attack because, for example, some of its effects cannot be noticed in the short term [20].

This research aims to develop a resilient cybersecurity method to mitigate the impact of FDI attacks on A.V. by detecting FDI attacks in the early stage. The contribution of this research includes the following:

- False Data Injection (FDI) attacks were injected into an autonomous vehicle (A.V.) simulation-based system developed by MathWorks Inc. for research purposes. We assumed an attacker compromised a smart sensor.
- The dataset was generated from the simulation model after integrating the cyberattack.
- An intelligent anomaly detection method based on LSTM was implemented to identify FDI attacks targeting the control system of the A.V. through a compromised sensor.
- The anomaly detection method classifies the generated dataset into two classifications: normal and anomaly data. It was developed using LSTM neural networks.
- We evaluated the accuracy of our model using precision, recall, and F1-Score measurements; the overall accuracy is equal to 99.95%.

This paper is organized as follows: the recent literature review is represented in section 2. Section 3 discusses the autonomous vehicle simulation model used in this research. Then, section 4 explains the system development and specifications. Finally, section 5 provides conclusions and future work related to this research.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 July 2022 doi:10.20944/preprints202207.0039.v1

2. Literature Review

Due to the rapid development in engineering and technology, cities have become increasingly smart. This can be achieved while relying on data and technology to improve several sectors such as mobility and transportation. As such, autonomous vehicles are an indispensable part of smart mobility that emerged to improve the life quality inside smart cities. Nevertheless, autonomous vehicles are vulnerable to a wide range of cyberattack vectors that might severely impact humans' life quality and safety. Therefore, several research studies have been conducted to analyze, identify, and mitigate autonomous vehicle cyberattacks and defense mechanisms.

For instance, in [21], the authors suggested a preemptive classification scheme for the cyber risk categories of connected and autonomous vehicles. Their predictive model uses Bayesian Networks to utilize the variables and fundamental relationships from the Common Vulnerability Scoring Scheme (CVSS) to parameterize the cyber risk of connected and autonomous vehicles. As a result of evaluating their model on an out-of-sample test, their B.N. predictive scheme exhibited high prediction accuracy for several risk scores and levels, scoring approximately 100%.

Also, in [22], the authors proposed a conceptual framework to classify the potential vulnerabilities of connected and autonomous vehicle systems. The suggested conceptual framework was developed using uniform modeling language (UML) using the KDD99 dataset to produce a new dataset modeling the cyberattacks targeting the communication processes of connected and autonomous vehicles, known as CAV-KDD-2020. CAV-KDD-2020 dataset is a communication-oriented dataset covering several types of attacks targeting different possible attack points of the connected and autonomous vehicle's systems, including the hardware parts of the autonomous vehicle such as LIDAR sensor, Camera, power system, software parts of the autonomous vehicle such as in-vehicle system, data processing system, and decision-making system, the data itself of the autonomous vehicle such as vehicle id, vehicle's speed, users personal information, and brake status, and the communication network protocols of the autonomous vehicle such as the vehicle to infrastructure communication, vehicle to cloud communication and vehicle to vehicle communication. To evaluate the new dataset, the authors employed two supervised machine learning classifiers, including a decision tree classifier (D.T.C.) and a Naive Bayes classifier (N.B.C.), to classify the cyberattacks on the autonomous vehicles into four attack groups, including probe Dos attacks, R2L attacks, and U2R attacks. Accordingly, the experimental results revealed that the D.T.C. model scored higher accuracy and precision proportions with a shorter runtime and thus is more applicable for the attack detection of autonomous vehicle communication.

Moreover, in [23], the authors proposed a real-time multi-stage deep-learning-based I.D.S. structure designed to recognize cyberattacks from the Intelligent Transportation Systems (ITS) to generate minimal false alarm rates (FAR). Their system employs the normal state-based and the Long Short-Term Memory (LSTM) deep learning model in a bidirectional mode to detect the potential attacks of connected and autonomous vehicle systems. To assess their implemented method's performance, the authors evaluated their model on two standard datasets: the UNSWNB-15 dataset and the CAR-HACK dataset. Consequently, their empirical investigations pointed out that their proposed multi-stage I.D.S. system surpassed other models scoring higher accuracy levels with 98.88% and 99.11% for UNSWNB-15 and CAR-HACK datasets, respectively. Such outcomes may enhance the cybersecurity for autonomous vehicles at both levels, the in-vehicle communications and out-vehicle communications (exterior).

Similarly, in [24], the authors developed a deep learning-based I.D.S. for autonomous vehicles in a real-time fashion. The proposed system is composed of two main stages. The first stage is responsible for features extraction leveraging auto-encoder-based long short-term memory. The second stage is responsible for anomaly detection and classification for every signal sequence using a convolutional neural network (CNN) in a real-time environment. Their experimental results reported 95.5% and 94.2% for model accuracy and precision, respectively.

Furthermore, in [25], the authors researched the cyber-security vulnerabilities of autonomous vehicles under sensor attacks. Specifically, they proposed a new rule-based I.D.S. system to identify the sensor attacks and sources for connected and autonomous vehicles. The proposed I.D.S. uses a combination of an extended Kalman filter (E.K.F.) to estimate the vehicle's location and a cumulative sum (CUSUM) discriminator to identify the possible variation of the sensor measurement. For higher resiliency against intrusion, multiple sensors were deployed to deliver real-time postures of the autonomous vehicle states. Besides, an auxiliary detector to examine the irregularity between multiple sensor measurements. Finally, a rule-based separation system is employed to analyze the detectors' results and provide information about the abnormal sensor. Extensive experimental results were reported, showing the developed model's usefulness in actual autonomous vehicle data.

Besides, in [26], the authors investigated and studied the threat classification concerning autonomous vehicles targeting three major security services: authentication, accountability, and availability. The authors elaborated on the various countermeasures for autonomous vehicle intrusions and their developmental aspects in this study. Specifically, the authors emphasized the vital role of blockchain to prevail over and mitigate such security and privacy concerns (for autonomous vehicles. Lastly, they end their investigational study by delving into the genuine concerns and questions of blockchain-based security systems for autonomous vehicles.

The authors in [27] proposed two deep learning algorithms to detect Denial of Service (DoS) attacks committed to the Electric Vehicle Charging Station (EVCS). The authors used python's long-short term memory (LSTM) and Deep Neural Network (D.N.N.) algorithms to classify the DoS attacks. It was assumed that attackers could use any weak network link to establish the DoS attack. The D.N.N. and LSTM algorithms were trained, tested, and validated. 50% of the data was used for training, 20% for validation, and 30% for testing. According to the authors, the accuracy of both deep learning algorithms has recorded high accuracy rates.

Unlike the studies mentioned above, where models are developed through the learning-based scheme (training and testing) using predefined systematic Attack-Aware datasets [28] that contain features of common cyber-attacks (intrusions), this research contributes to the cybersecurity of A.V.s by detecting the False Data Injection (FDI) cyber-attacks developed at this research. First, we inject False Data Injection (FDI) attacks into an A.V. simulation-based system developed by MathWorks. Inc. Second, we collect the dataset generated from the simulation model after integrating the cyberattack. Third, we implement an intelligent symmetrical anomaly detection method to identify FDI attacks targeting the control system of the A.V. through a compromised sensor. We use long short-term memory (LSTM) deep networks to detect FDI attacks in

 doi:10.20944/preprints202207.0039.v1

the early stage to ensure the stability of the operation of A.V.s. Our method classifies the collected dataset into two classifications: normal and anomaly data. The experimental result revealed a high-performant model outperforming other state-of-the-art models in the same study area.

3. Autonomous Vehicles Simulation Model

MathWorks developed the A.V. software system used in this research. Inc. It is a simulation based-model built using MATLAB, and Simulink focuses on using adaptive cruise control (A.C.C.) to regulate the A.V. velocity. It consists of two cars (1) the ego car and (2) the lead car. The ego car is a self-driving car that needs to maintain its speed and distance from the lead car in the same lane. Therefore, the ego car relies on an A.C.C. to regulate speed and distance. The A.C.C. system composes speed control and distance control to adjust the dynamic of the ego car to be symmetrically commensurate with the lead car [29].

The ego car uses sensors such as radar to collect information about the position of the ego car and the lead car. The sensor readings are fed to the A.C.C. system to regulate the speed of the ego car to maintain its position according to the lead car. The default safe distance between the ego and lead car is symmetrically set to 10 meters. Therefore, if the safe distance between the ego and the lead car is smaller than or equal to the relative distance (i.e., Asymmetric), the ego car needs to increase its speed ($D_{rel} \geq D_{Safe}$), as shown in Figure 2. However, if the safe distance is larger than the relative distance, the ego car needs to decrease its speed ($D_{rel} < D_{Safe}$), as shown in Figure 3.

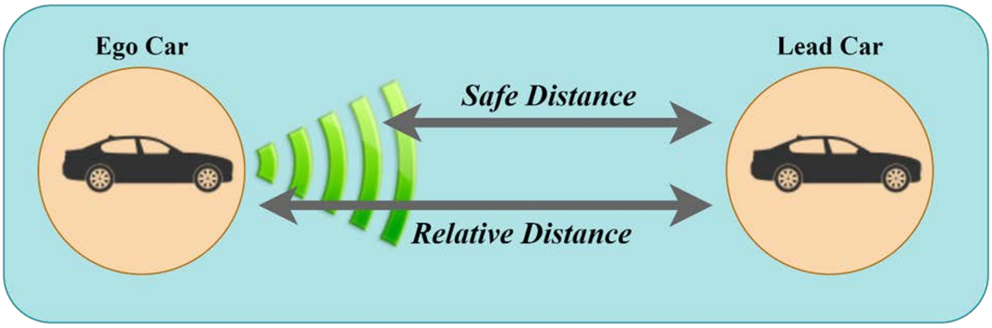

Figure 2. Safe Distance Vs. Relative Distance (1)

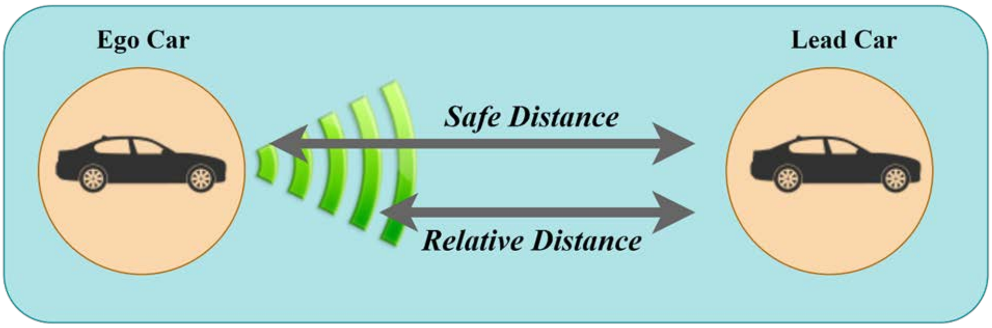

Figure 3. Safe Distance Vs. Relative Distance (2)

Figure 4 shows the velocity of the ego car and the lead car. The initial velocity of the ego car is 20 m/s, and the initial velocity of the lead car is 25 m/s. The desired velocity, in this case, is 30 m/s. It can be observed that the ego car symmetrically maintains its speed according to the lead car. Therefore, when the lead car reduces or increases its speed, the ego car follows accordingly.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 July 2022 doi:10.20944/preprints202207.0039.v1

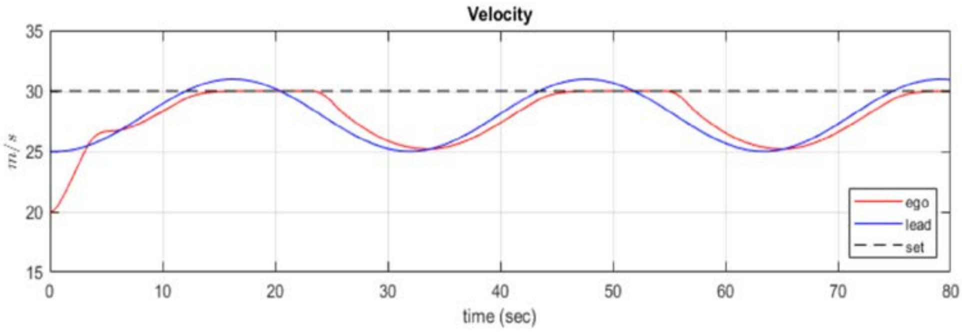

Figure 4. The velocity of the ego car and lead car

Figure 5 illustrates the distance between the ego car and the lead car. Overall, we can observe that the safe and relative distance is maintained symmetrically between the ego car and the lead car. When the relative distance decreases, the safe distance decreases too. Whereas when the relative distance increases, the safe distance increase accordingly.

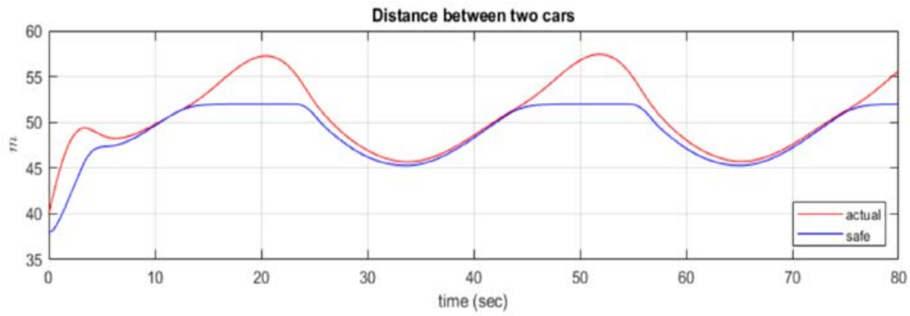

Figure 5. Distance between the two cars

4. System Development and Specifications

This section discusses generating the dataset used in this research and provides information about the data preprocessing procedure. In addition, it provides a detailed description of the implementation of the LSTM model used for the classification's procedure. Finally, it evaluates the performance of the LSTM model. Figure 6 illustrates the overall architecture of the system model used in this research.

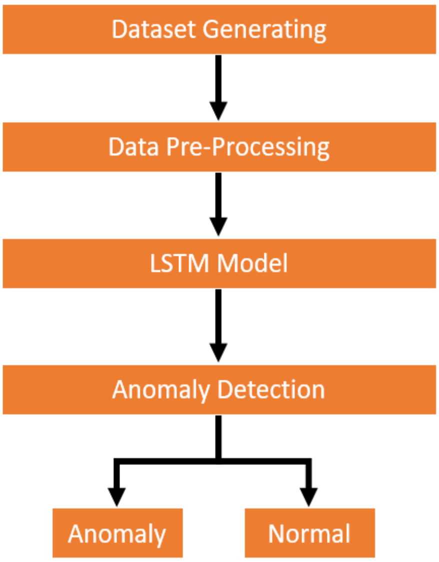

Figure 6. System Model

 doi:10.20944/preprints202207.0039.v1

4.1 A Scheme for Generating Dataset

This section will discuss the testbed used to generate a dataset for anomaly detection for the A.V. system. First, the development and integration of the cyberattack are explained. Second, it shows the collected dataset features. Third, it discusses the calculation of the anomaly detection feature.

4.1.1 Implementation of Cyberattack

The MATLAB/Simulink simulation model used in this research does not include cyberattacks. Therefore, FDI attacks were implemented and injected into the sensor responsible for measuring the position of the ego car. The equation of the FDI attack is shown in Equation 1. The $\boldsymbol{D_{rel}}$ Refers to the relative distance between the ego car and the lead car measured by the sensor. The attack percentage refers to the strength of the attack in percentage. The value of the attack percentage starts from 0.00001% to 100% to include a maximum number of possible attack strengths.

Figure 7 illustrates the velocity performance of both cars under FDI attack. The attack percentage value, in this case, was 60%. According to the figure, the velocity of the ego car was not stable due to the impact of the FDI attack compared with Figure 3.

Figure. 8 depicts the safe and relative distance between the two cars under the same attack percentage. By comparing the results of Figure 7 with Figure 4, the values of both parameters are no longer accurate because of the attack, which impairs the symmetry of Figure 4.

$$D_{rel+1} = D_{rel} + (attack\ percentage\ \times D_{rel}) \quad (1)$$

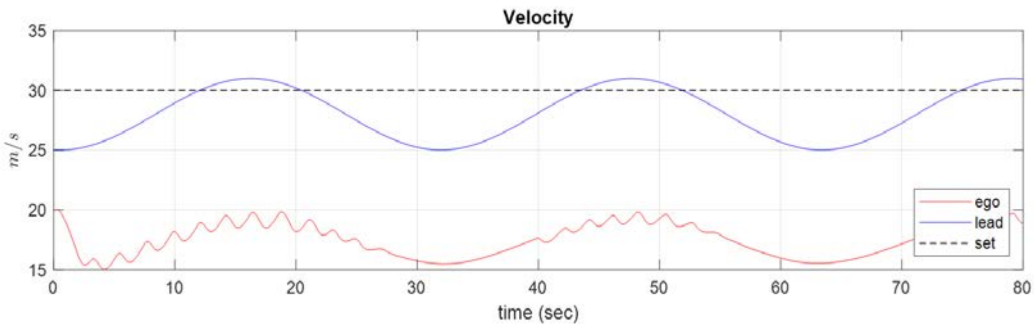

Figure 7. Velocity performance under FDI attack

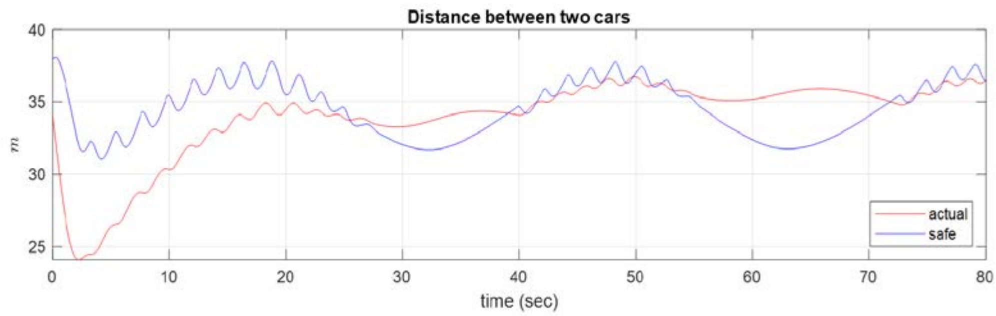

Figure 8. Distance performance under FDI attack

4.1.2 Dataset Features

The dataset used in this research was collected from a real-time simulation using the A.V. model with the integrated FDI attacks. We collected the response of the following parameters: the actual position of the ego car, the actual velocity of the ego car, the actual position of the lead car, and the actual velocity of the lead car. In addition, we extracted a new feature we

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 July 2022 doi:10.20944/preprints202207.0039.v1

called the anomaly detection label with two classifications label normal and anomaly, using the last four features. Therefore, our dataset consists of five features listed in Table. 1

Table 1. Dataset Features

Feature No.	Feature Name	Unit	Data Type
1	Actual position of the ego car	Meter	Double
2	Actual velocity of the ego car	Meter/Second	Double
3	Actual position of the lead car	Meter	Double
4	Actual velocity of the lead car	Meter/Second	Double
5	Anomaly detection label	Normal, Anomaly	Binary

4.1.3 Anomaly Detection Label

The anomaly detection label was calculated using features 1, 2, 3, and 4. We injected FDI attacks into the sensor responsible for measuring the actual position of the ego car. According to our observation, the ego car has a stable performance when the attacks' strength is between 0.00001% to 0.01%. For this reason, this anomaly detection label is marked as normal. The size of the normal dataset is 10,000 records, and each record has four features with 81 data lengths. However, when the strength of the FDI attack is larger than 0.01%, the ego car's performance is considered unstable. Therefore, we inject FDI attacks into the actual position of the ego car, but the strength of the FDI attack this time is between 0.011% to 100%. The size of the injected dataset (anomaly) is 10,000 records, and each record has four features with 81 data lengths. Therefore, the total size of the dataset is 20,000 records.

4.2 Implementation of LSTM

LSTM is a deep neural network first proposed by Hochreiter in 1997 [30] [31]. LSTM uses time-series data for classification by keeping track of cell states to preserve certain memory trends across time [32]. LSTM block consists of three gates: forget, update, and output, which works with the input for a time series, as shown in Figure 9 [33][34]. The model decides whether to forget, update, or output new data at each state stage. Therefore, LSTM is made to avoid the issue of long-term dependence. The forget gate determines whether a piece of data should be saved. In the LSTM, the input gate refreshes the cells, while the output gate always determines the hidden state. As a result, they determine which data should be shared with other cells and which should be ignored based on the outcome, which ranges from zero to one. Zero indicates rejection, but one indicates inclusion.

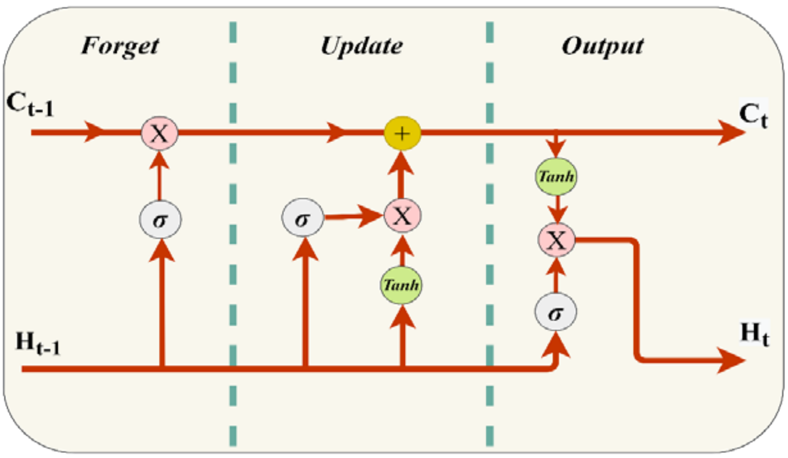

Figure 9. The LSTM Block

 doi:10.20944/preprints202207.0039.v1

The component of the LSTM architecture used for classification is illustrated in Figure 10. Initially, the data is fed to the sequence input layer, followed by the LSTM layer. Next, the prediction procedure is performed in the fully connected Softmax layers. Finally, the output is produced in the classification layer [35].

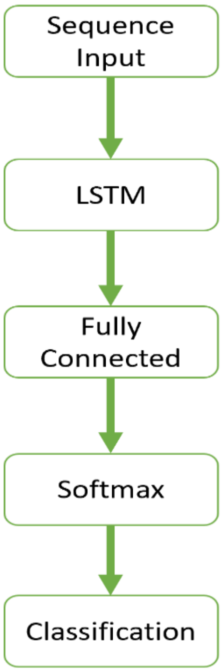

Figure 10. LSTM Architecture

4.3 Training Procedure

After preprocessing the dataset, we trained it using the LSTM model. We split our dataset into two groups; the first group is used for training with 70% of the dataset, and the second group is 30%, which will be used for testing. Therefore 17,000 records were used for training, and 3,000 were used for testing. We used the Adam optimization algorithm to train our LSTM networks, as discussed in the following section [36].

4.3.1 Adaptive Moment Estimation Optimization (ADAM)

Classification can be difficult when dealing with problems relating to the learning process. Several approaches have been proposed to help us arrive at an optimal learning level. The Adaptive moment estimation (ADAM) optimization algorithm is a recent deep learning extension of the stochastic gradient descent algorithm, which has recently been used in a variety of applications like on the Internet of Things, text detection, and so on [37] [38] [39] [40]. According to empirical outcomes, the method has performed well in practice and compares favorably to other stochastic optimization approaches [40]. Stochastic gradient descent is an efficient and effective optimization technique that has played an important role in many machines learning. According to the concept of the method, individual adaptive learning rates for distinct parameters are calculated using estimates of the gradient's first and second moments based on combining bothmsProp and AdaGrad [39]. RmsProp computes the average of recent changes in the magnitude of the internal signal gradient, while AdaGrad handles sparse gradients with uncentered variance [38]. These algorithms can be calculated as follows:

$$m_t = \beta m_t - 1 + (1 - beta_1) g_t \quad (2)$$

$$v_t = \beta_2 v_t - 1 + (1 - beta_2) {g^2}_t \quad (3)$$

4.4 Testing Procedure

We evaluated our model using the confusion matrix shown in Figure 11, which depends on the True Positive (T.P.) and False Negative (F.P.). The accuracy is calculated using equation 4 [41]. T.P. refers to the number of positive data classified correctly. F.N. refers to the number of positive misclassified data. Meanwhile, F.P. refers to the number of negative misclassified data, and T.N. refers to the number of negative data classified correctly. As was mentioned above, 30% of the data was used for testing. Figure 12 illustrates the confusion matrix of our proposed models. We can observe that only three records from the data were miss classified; however, the rest of the records were classified correctly.

$$Accuarcy = (TP + TN)/(TP + TN + FP + FN) \quad (4)$$

		Predicted Condition	
		Positive	Negative
True Condition	Positive	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)

Figure 11. Two Class Confusion Matrix for Calculation Accuracy

In addition, we evaluated our test method using precision, recall, and F1-score metrics, as represented in Table. 3 [42]. The accuracy of our model reached 99.95 %. Tabel.4 lists the accuracy of current existing deep learning methods developed by researchers compared with ours.

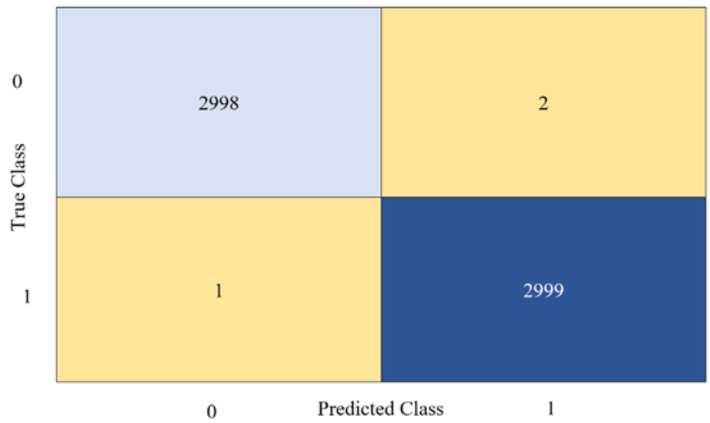

Figure 12. Confusion Matrix of Proposed Model

The precision is calculated using equation 7 :

$$\text{Precision} = TP/(TP + FP) \quad (7)$$

The recall is calculated using equation 8 :

$$\text{Recall} = TP/(TP + FN) \quad (8)$$

The F1-Score is calculated using equation 9 :

$$\text{F1} - \text{Score} = 2 * (\text{Precision} \times \text{Recall})/(\text{Precision} + \text{Recall}) \quad (9)$$

 doi:10.20944/preprints202207.0039.v1

Table 3. Precision, recall, and F1-score metrics.

Accuracy Parameter	Value
Precision	99.93%
Recall	99.97%
F1-Score	99.95%
Accuracy	**99.95%**

According to Table 4 which considers the performance comparison between our proposed model against several state-of-the-art models. The comparisons revealed the supremacy of our proposed model in terms of accuracy and the data generation process. In addition, it shows the number of features that were used in each study, and it can be observed that our proposed method used a lower number of features (only four) compared with other models. Generally, training a deep learning model with a few features can be challenging since most deep learning models require a sufficient number of features to reach higher accuracy [43]. While deep learning models' accuracy could also suffer from many number features [44].

Table 4. Comparing our proposed model's accuracy with existing deep learning models' accuracy.

Research	Task	No. of Features	ML Model	Accuracy
Hamza et al. [45]	Detection	NA	COSBO-BiLSTM	98.81%
Almasoud et al. [46]	Detection	24	RNN-GLSTM	7.96%
Roh et al. [47]	Detection	64	CNN-LSTM	92.03%
Sarwar et al. [48]	Detection	83	Random Forest	83%
Song et al. [49]	Classification	77	Deep-learning	97.4%
Alkahtani et al. [50]	Classification	80	CNN-LSTM	98.90
Al-Haija et al. [51]	Classification	43	CNN	98.2%
Ullah et al. [52]	Detection	83	SVM	80%
Proposed method	**Detection**	4	**LSTM**	**99.95%**

5. Conclusions

Rapid computing advances have significantly impacted the study and development of autonomous vehicles (Avs) in various fields. Since A.V.s are widely regarded as an essential component of smart cities, their roles in our daily lives are significant. However, autonomous vehicles may be vulnerable to cyberattacks that endanger human lives. Therefore, this study proposed an anomaly detection model based on symmetrical LSTM neural networks. First, FDI attacks were injected into an A.V. simulation-based model developed by MathWorks Inc. for research purposes. Second, the dataset was generated from the simulation model. Third, an intelligent anomaly detection method based on LSTM was implemented to detect FDI attacks targeting the control system of the A.V. through a compromised sensor. Finally, we evaluated the accuracy of our model using precision, recall, and F1-Score measurements; our overall LSTM model accuracy reached 99.95%. In addition, we compared our model performance with existing recent research that used deep network models to validate the accuracy of our proposed model. As was mentioned, we studied the effect of FDI attacks on the A.V. models; however, for future work, we will study the impact of denial of service (DDoS) and man in the middle (MITHM) attacks on the model.

Author Contributions: "Conceptualization, A.A.A.; methodology, A.A.A. and Q.A.A-H.; software, A.A.A..; validation, Q.A.A-H. and A.A.Q..; formal analysis, Q.A.A-H.; investigation, A.A.A. and R.A.S.; resources, A.A.A.; data curation, A.A.Q, and R.A.S.; writing—original draft preparation, A.A.A., Q.A.A-H, A.A.Q, and R.A.S..; writing—review and editing, A.A.A. and Q.A.A-H.; visualization, A.A.A., Q.A.A-H, A.A.Q, and R.A.S.; supervision, A.A.A. and Q.A.A-H; Project Administration, A.A.A; funding acquisition, A.A.A. All authors have read and agreed to the published version of the manuscript."

Funding: This research was funded by the ministry of education and the Deanship of Scientific Research at King Abdulaziz University, Jeddah 21589, Saudi Arabia, for financial and technical support under code number (G: 486-611-1443).

Acknowledgments: The authors are thankful to the Deanship of Scientific Research at King Abdulaziz University, Jeddah 21589, Saudi Arabia, for funding this work under the General Research Funding program grant code (G: 486-611-1443).

Conflicts of Interest: The authors declare no conflict of interest.

References

[1]. P. C. Zachary, B. Dustin, Y. Siyu, H. Andreas, L. Jun, F. Michael and C. Zhongwei, "Batteries and fuel cells for emerging electric Batteries and fuel cells for emerging electric," Nature Energy, 2018.
[2]. R. Syed, A. Haneen, M. Yemeserach and I. S. Arif, "A Study of E.V. B.M.S. Cyber Security Based on Neural Network S.O.C. Prediction," in EEE/PES Transmission and Distribution Conference and Exposition (T&D), Denver, CO, U.S.A., 2018.
[3]. Q. A. Al-Haija, M. I. Al Tarayrah and H. M. Enshasy, "Time-Series Model for Forecasting Short-term Future Additions of Renewable Energy to Worldwide Capacity," 2020 International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy (ICDABI), 2020, pp. 1-6, doi: 10.1109/ICDABI51230.2020.9325625.
[4]. J. Chan-Kuk, L. JaeHoon and Y. Okyeon, "Encryption scheme in portable electric vehicle charging infrastructure: Encryption scheme using symmetric key," in 4th International Conference on Computer Applications and Information Processing Technology (CAIPT), Kuta Bali, Indonesia, 2017.
[5]. Q. A. Al-Haija, C. D. McCurry and S. Zein-Sabatto, "A Real Time Node Connectivity Algorithm for Synchronous Cyber Physical and IoT Network Systems," 2020 SoutheastCon, 2020, pp. 1-8, doi: 10.1109/SoutheastCon44009.2020.9249730.
[6]. F. Yosra, B. A. Lamia, T. Wassim and A. S. Leila, "Cyber security issues of Internet of electric vehicles," in IEEE Wireless Communications and Networking Conference (WCNC), Barcelona, Spain, 2018.
[7]. Q. A. Al-Haija, "On the Security of Cyber-Physical Systems Against Stochastic Cyber-Attacks Models," 2021 IEEE International I.O.T., Electronics and Mechatronics Conference (IEMTRONICS), 2021, pp. 1-6, doi: 10.1109/IEMTRONICS52119.2021.9422623.
[8]. Y. Bowen, G. Lulu and Y. Jin, "Real-time Simulation of Electric Vehicle Powertrain: Hardware-in-the-Loop (H.I.L.) Testbed for Cyber-Physical Security," in IEEE Transportation Electrification Conference & Expo (ITEC), Chicago, IL, U.S.A., 2020.
[9]. Liu S, Huang Y, Zhang R. On-Road Vehicle Recognition Using the Symmetry Property and Snake Models. International Journal of Advanced Robotic Systems. December 2013. doi:10.5772/57382
[10]. J. Cui, L. S. Liew, G. Sabaliauskaite and ZhouFengjun, "A review on safety failures, security attacks, and available countermeasures for autonomous vehicles," Ad Hoc Networks, vol. 90, 2019.
[11]. S.A.E. International, "Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles," S.A.E. International in United States, 30 09 2016.
[12]. H. Hartenstein and L. P. Laberteaux, "A tutorial survey on vehicular ad hoc networks," IEEE Communications Magazine, vol. 46, no. 6, pp. 164-171, 2008.
[13]. G. Loukas, Cyber-Physical Attacks A Growing Invisible Threat, elsevier, 2015.
[14]. S. Musleh, G. Chen and Z. Y. Dong, "A Survey on the Detection Algorithms for False Data Injection Attacks in Smart Grids," IEEE Transactions on Smart Grid, vol. 11, no. 3, pp. 2218 - 2234, 2020.
[15]. Al-Haija, Q.A.; Alsulami, A.A. High Performance Classification Model to Identify Ransomware Payments for Heterogeneous Bitcoin Networks. Electronics 2021, 10, 2113. https://doi.org/10.3390/electronics10172113
[16]. E. Staddon, V. Loscri and N. Mitton, "Attack Categorisation for IoT Applications in Critical Infrastructures, a Survey," applied sciences, vol. 11, pp. 1-39, 2021.
[17]. M. Krotofil, A. A. Cárdenas, B. Manning and J. Larsen, "C.P.S.: driving cyber-physical systems to unsafe operating conditions by timing DoS attacks on sensor signals," in The 30th Annual Computer Security Applications Conference, 2014.

[18]. A. S. o. P.-B. A. D. i. C.-P. Systems, "Giraldo,Jairo ;Urbina,David ;Cardenas,Alvaro ;Valente,Junia ;Faisal,Mustafa ;Ruths,Justin ;Tippenhauer,Nils Ole;Sandberg,Henrik ;CandellRichard," ACM Computing Surveys, vol. 51, no. 4, pp. 1-36, 2019.
[19]. P. C. van Oorschot, "Information Security and Cryptography book series (I.S.C.)," in Intrusion Detection and Network-Based Attacks, Cham, Springer, 2021, pp. 309-338.
[20]. Abu Al-Haija Q, Top-Down Machine Learning-Based Architecture for Cyberattacks Identification and Classification in IoT Communication Networks. Front. Big Data 4:782902.2022. doi: 10.3389/fdata.2021.782902.
[21]. MathWorks, "Adaptive Cruise Control System Using Model Predictive Control," MAthWorks, 2017. [Online]. Available: https://www.mathworks.com/help/mpc/ug/adaptive-cruise-control-using-model-predictive-controller.html.
[22]. Sheehan, F. Murphy, M. Martin and C. Ryan, "Connected and autonomous vehicles: A cyber-risk classification framework," Transportation Research Part A: Policy and Practice, vol. 124, pp. 523-536, 2019.
[23]. Q. He, X. Meng, R. Qu and R. Xi, "Machine Learning-Based Detection for Cyber Security Attacks on Connected and Autonomous Vehicles," Mathematics, 2020.
[24]. Ahmed Khan, N. Moustafa, D. Pi, W. Haider, B. Li and A. Jolfaei, "An Enhanced Multi-Stage Deep Learning Framework for Detecting Malicious Activities From Autonomous Vehicles," IEEE Transactions on Intelligent Transportation Systems, pp. 1-10, 2021.
[25]. R. Oucheikh, M. Fri, F. Fedouaki and M. Hain, "Deep Real-Time Anomaly Detection for Connected Autonomous Vehicles," Procedia Computer Science, vol. 177, pp. 456-461, 2020.
[26]. Sheehan, F. Murphy, M. Mullins and C. Ryan, "Connected and autonomous vehicles: A cyber-risk classification framework," Transportation Research Part A: Policy and Practice, vol. 124, pp. 523-536, 2019.
[27]. R. Gupta, S. Tanwar, N. Kumar and S. Tyagi, "Blockchain-based security attack resilience schemes for autonomous vehicles in industry 4.0: A systematic review," Computers & Electrical Engineering, vol. 86, 2020.
[28]. Abu Al-Haija, Q.; Al-Badawi, A. Attack-Aware IoT Network Traffic Routing Leveraging Ensemble Learning. Sensors 2022, 22, 241. https://doi.org/10.3390/s22010241
[29]. B. Manoj and H. A. Mohd, "Deep Learning-based Intrusion Detection System for Electric Vehicle Charging Station," in 2nd International Conference on Smart Power & Internet Energy Systems (SPIES), Bangkok, Thailand,, 2020.
[30]. Balouji, I. Y. Gu, M. H. Bollen, A. Bagheri and M. Nazari, "A LSTM-based deep learning method with application to voltage dip classification," in 18th International Conference on Harmonics and Quality of Power (ICHQP), Ljubljana, Slovenia, 2018.
[31]. S. a. S. J. Hochreiter, "Long Short-term Memory,," Neural computation,, vol. 9 , pp. 1735-80, 1997.
[32]. S. Khafaga, A. A. Alhussan, E.-S. M. El-kenawy, A. Ibrahim, S. H. Abd Elkhalik, S. Y. El-Mashad and A. A. Abdelhamid, "Improved Prediction of Metamaterial Antenna Bandwidth Using Adaptive Optimization of LSTM," Computers, Materials & Continua, vol. 73, pp. 866-881, 2022.
[33]. S. Mahajan, H. R. and and K. Kotecha, "Prediction of network traffic in wireless mesh networks using hybrid deep learning model," IEEE Access, vol. 10, p. 7003–7015, 2022.
[34]. Q. Abu Al-Haija, "A machine learning based predictive model for time-series modelling and analysis", International Journal of Spatio-Temporal Data Science, Inderscience, 2021 1:3, 270-283.
[35]. Mathworks, "Long Short-Term Memory Networks," Mathworks, [Online]. Available: https://www.mathworks.com/help/deeplearning/ug/long-short-term-memory-networks.html.
[36]. D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," in 3rd International Conference for Learning Representations, San Diego, 2015.
[37]. S. B. Tahir, A. Jalal and and K. Kim, "Wearable inertial sensors for daily activity analysis based on Adam Optimization and the maximum entropy Markov Model," Entropy, , vol. 22, no. 5, p. 579, 2020.
[38]. H. Kohli, J. Agarwal and and M. Kumar, "An improved method for text detection using Adam Optimization algorithm," Global Transitions Proceedings, 2022.
[39]. D. Kingma and a. J. B.A., "Adam a Method for Stochastic Optimization. The 3rd International Conference for Learning Representations," in Adam a Method for Stochastic Optimization. The 3rd International Conference for Learning Representations, San Diego, 2022.
[40]. K. Jais, A. R. Ismail and and S. Q. Nisa, "Adam optimization algorithm for wide and deep neural network," Knowledge Engineering and Data Science," vol. 2, no. 1, p. 41, 2019.
[41]. Abu Al-Haija, Q.; Al-Saraireh, J. Asymmetric Identification Model for Human-Robot Contacts via Supervised Learning. Symmetry 2022, 14, 591. https://doi.org/10.3390/sym14030591.
[42]. D. Chicco and G. Jurman, "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation," B.M.C. Genomics, 2020.
[43]. B. Mikhail, D. H. M. Siyuan and M. Soumik, "Reconciling modern machine-learning practice and the classical bias–variance trade-off," Proceedings of the National Academy of Sciences, vol. 116, no. 32, pp. 15849-15854, 2019.

[44]. F. Jerry Alan and O. Dan R., "Interactive machine learning," in Proceedings of the 8th international conference on Intelligent user interfaces, Miami, Florida, USA, 2003 .
[45]. M. A. Hamza, H. S. B. Haj, L.-M.-S. Souad, M. K. Nour, F. N. Al-Wesabi, A. Motwakel and A. M. a. A. D. Hilal, "Optimal Bidirectional LSTM for Modulation Signal Classification in Communication Systems," Computers, Materials & Continua, vol. 72, pp. 3056-3071, 2022.
[46]. S. Almasoud, T. A. E. Eisa, F. N. Al-Wesabi, A. Elsafi, M. Al Duhayyim, I. Yaseen and M. A. a. M. Hamza, "Parkinson's Detection Using RNN-Graph-LSTM with Optimization Based on Speech Signals," Computers, Materials & Continua, vol. 72, pp. 872-886, 2022.
[47]. Roh, S. Oh, H. Song, J. Han and S. Lim, "Deep Learning-based Wireless Signal Classification in the IoT Environment," Computers, Materials & Continua, vol. 71, pp. 5718-5732, 2022.
[48]. S. H. a. W. U. K. A. Sarwar, "Design of an Advance Intrusion Detection System for IoT Networks," in ICAI, Islamabad, Pakistan, 2022.
[49]. S. H. a. Y.-G. C. Y. Song, "Analysis of Autoencoders for Network Intrusion Detection," Sensors, , 2021.
[50]. H. A. a. T. H. H. Aldhyani, "Intrusion Detection System to Advance Internet," Complexity, , vol. 2021, 2021.
[51]. S. Z.-S. Q. Abu Al-Haija., "An Efficient Deep-Learning-Based Detection and Classification System for Cyber-Attacks in IoT Communication Networks," Electronics , vol. 9, p. 2152, 2020.
[52]. Q. H. M. I. Ullah, "A Scheme for Generating a Dataset for Anomalous Activity Detection in IoT Networks," Advances in Artificial Intelligence, p. 508–520, 2020.