IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 7, July 2022

DOI: 10.17148/IJARCCE.2022.11743

A DEEP LEARNING APPROACH FOR INTRUSION DETECTION USING RECURRENT NEURAL NETWORKS

Pavananag T N[1], Divakar H R [2]

Research Scholar, Dept. of MCA, P.E.S College of Engineering, Mandya, India[1].

Associate Professor, Dept. of MCA, P.E.S College of Engineering, Mandya, India[2].

Abstract: Intrusion detection plays an important role in ensuring information security, and the key technology is to accurately identify various attacks in the network. In this project we explore how to model an intrusion detection system based on deep learning and we propose a deep learning approach for intrusion detection using recurrent neural networks (RNN-IDS). Moreover, we study the performance of the proposed model in binary classification and multiclass classification and the number of neurons and different learning rate impacts on the performance of the proposed model. We compare it with those of artificial neural network, random forest, support vector machine and other machine learning methods. The experimental results show that RNN-IDS is very suitable for modelling a classification model with high accuracy and that is performance is superior to that of traditional machine learning classification methods in both binary and multiclass classification. The RNN-IDS model improves the accuracy of the intrusion detection and provides a new research method for intrusion.

Keywords: Intrusion detection, Recurrent neural networks, Deep Learning.

I. INTRODUCTION

Machine learning is becoming a more important tool in both the business and scientific realms nowadays. Classification techniques are helped by machine learning's deep learning technology, which is one of its many components. The referral procedure is facilitated. Since there are more security concerns to social networking sites now than ever before, network attack detection has gained more and more attention. Although the way people live, learn, and work is changing as a result of the inexorable profound Internet-society integration, many security risks we face are becoming more and more real. A significant piece of data security research, the intrusion detection system (IDS), can tell apart an assault from a recent interruption from an ongoing infiltration. In this study, we are identifying if system Transportation behaviour is normal or abnormal, or a five-tiered categorization arrangement issue, determining if it is typical or any of the remaining four forms of assault User to Root (U2R), Probe (Probing), and Root to Local are examples of denial-of-service attacks (R2L). So put it simply, the main motivation for interference recognition to increase the precision of classifiers is correctly identifying interfering behaviors.

II. RELATED WORK

B.V Elsevier [1] proposes a system that provides deep learning architectures to develop an adaptive and resilient network intrusion detection system (IDS) to detect and classify network attacks. The focus is on how deep learning or deep neural networks (DNNs) might enable flexible IDS with learning power to detect known and new or zero-day network behavioral features, hence expelling the systems invader and lowering the risk of compromise. We used the UNSW-NB15 dataset, which reflects actual contemporary network communication behavior with artificially manufactured attack activities, to show the model's efficacy. The proposed deep learning classification architecture coupled with the semi-dynamic hyper parameter tuning approach demonstrated significant improvements to multiclass models compared to the results of similar deep learning-based network IDS. The models showed that our proposed approach obtained an overall accuracy of 95.4% and 95.6% for the pre-partitioned and user-defined multiclass classification.

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 7, July 2022

DOI: 10.17148/IJARCCE.2022.11743

D. Jing and H.B Chen [2] proposes a system that Support Vector Machine (SVM) with a new scaling method for binary-classification and multi-classification experiments. Network Intrusion Detection System (NIDS) becomes an integral part of the IoT environment. The method's efficiency is evaluated using its accuracy, detection rate, and false positive rate. Compared with other methods. The accuracy of the proposed method reaches 85.99% for binary-classification, compared to 78.47% by Expectation-Maximization (EM) clustering. For multi-classification, the proposed SVM method can achieve the testing accuracy of 75.77%, which is 6.17% higher than that of Naive Bayes (NB).

D.S Yeung and E.C.C Tsang [3] proposes a system that Denial of service (DoS) problem is one of serious attacks in the Internet. In an effort to stop authorized users from utilizing the system, the attackers try to deplete the service provider's resource pool. Most DoS detection software use threshold and rule-based detection techniques. This paper aims to apply machine learning techniques, such as radial-basis function neural network (RBFNN) and support vector machines (SVM), to solve the DoS problem and compare which technique, is better to detect DoS. The experimental result of this detection method is that it has the ability to detect or predict new attacks when some patterns are similar to the attack patterns learnt in the past.

F.Ullah [4] proposes a system that Denial of service (DoS) problem is one of serious attacks in the Internet. In an effort to stop authorized users from utilizing the system, the attackers try to deplete the service provider's resource pool. Most DoS detection software use threshold and rule-based detection techniques. This paper aims to apply machine learning techniques, such as radial-basis function neural network (RBFNN) and support vector machines (SVM), to solve the DoS problem and compare which technique, is better to detect DoS. The experimental result of this detection method is that it has the ability to detect or predict new attacks when some patterns are similar to the attack patterns learnt in the past.

Raj Kishore and Anamika Chauhan [5] proposes a system that is concerned with intrusion detection systems and its several types. An Intrusion detection system is a software to monitor and protect our network from any kind of intrusions. With rapid growth of internet, fear of cyber intrusion increases. This paper is concerned with intrusion detection systems and its several types. As we know that today's era is of computer networks which can be devastated to it, so intrusion detection system can help computer administrators to curb such activities and prevent our system. The experimental results of this paper is that to provide information about the need of intrusion detection systems in day to day life.

M.K Yadav and K.P Sharma [6] proposes a system mainly focuses on providing the analytical studies of such existing intrusion detection system. This research also examines the many available methods for building an efficient IDS utilizing single, hybrid, and ensemble machine learning techniques. Then, the approaches have been analyzed based on different data sets and compared with aiming to provide a simple path and guidance for effective future work.

III. METHODOLOGY

A. Recurrent Neural Networks

A network with layers and feedback loops called a recurrent neural network (RNN) may transmit information from the past. A RNN is composed loops that make the information persistent. The RNN's hidden layers serve as data capacity, similar to system memory. The strong type DNN called RNNs processes sequence data using loops and internal memory. The input and output of an RNN neuron are xt and ht, respectively. The loop, which is symbolised by the repeating arrow, information allow travel one stage of network to next stage.

B. Deep Learning

One of the many methods that make up Deep learning is part of machine learning (ML) (DL) Automated abstraction discovery is the fundamental idea behind all deep learning techniques. The supervised and unsupervised learning techniques used in DL are built on numerous ANN layers. Every processing layer in DL produces a non-linear output from the data it receives from the previous layer, which is made up of several processing layers. The human brain's neurons' capacity to analyse signals served as an inspiration for the DL functionality.

C. Long Short Term Memory

Deep learning uses an artificial recurrent neural network (RNN) architecture called long short-term memory (LSTM) (LSTM). Given that significant events in a time series may be separated by delays of variable lengths, LSTM networks are particularly well adapted to categorising, analysing, and making predictions based on time series data. Long Short Term Memory s are far better RNN and conventional feed-forward neural networks in many aspects. This is explained by their propensity to selectively remember patterns over long periods of time.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 7, July 2022

DOI: 10.17148/IJARCCE.2022.11743

D. System Architecture:

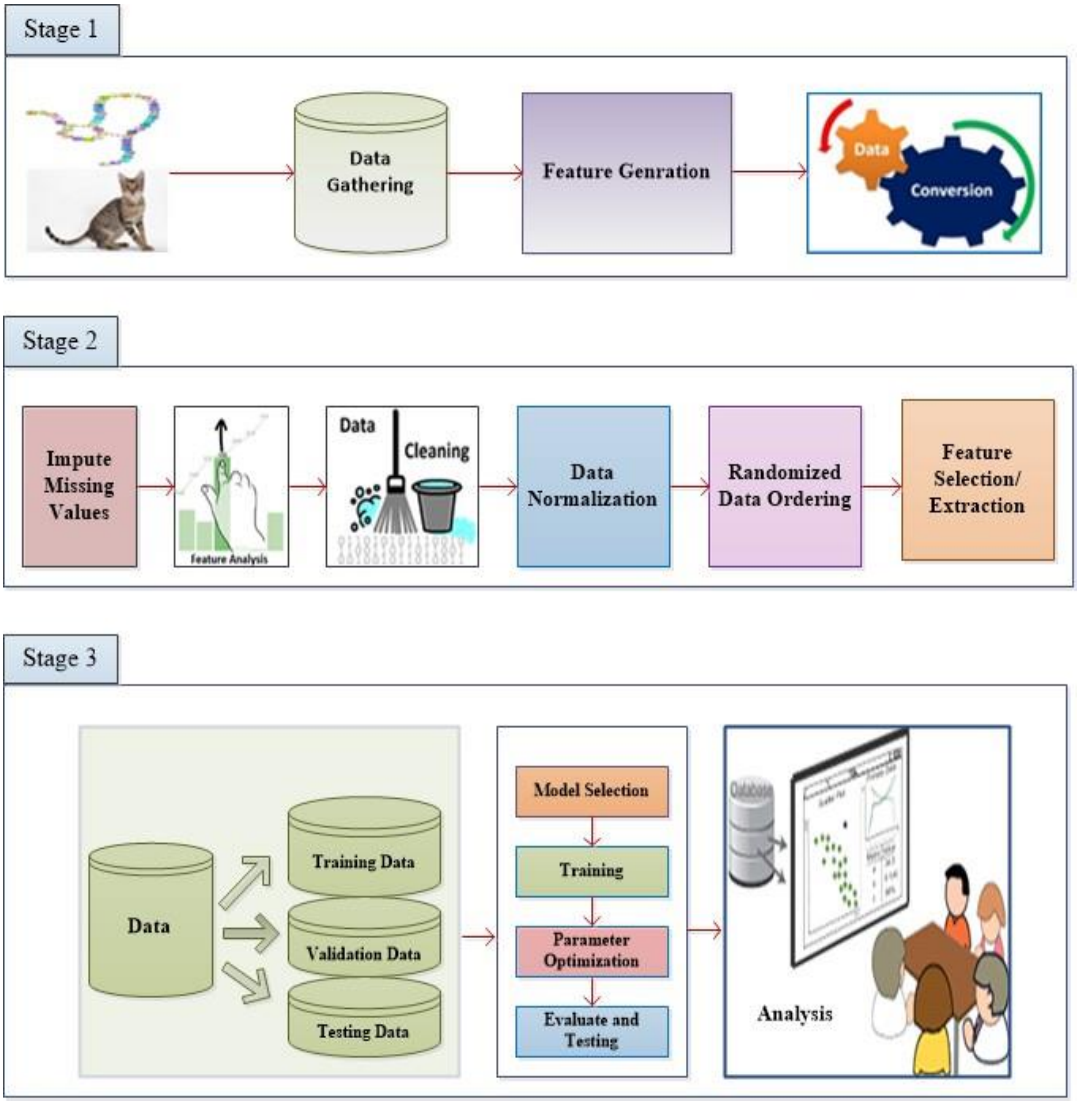

Fig Architecture diagram

Stage 1: Each traffic record has 41 features as well as a class designation. There are four different types of assaults in the dataset: R2L (Root to Local) assaults, DoS (Denial of Service), and U2R (User to Root)and problems (Probing attack).Testing set, which includes certain distinct attack types that disappear in the training set, can be used provided more realistic theory of foundation detection.

Stage 2: The data may have gaps and a lot of worthless information. To take care of this, data purification is finished. Dealing with inaccurate, noisy, etc., data is necessary.

Stage 3: After taking into account the data from Stage 2, the data is trained using an RNN neural network, and the resulting data is examined.

Data Absent:

This situation arises when there are some missing data from the data set. It may be dealt with in a number of ways.

Among them are:

1. Ignore the tuples: This strategy only works when the dataset at hand is sizable and a tuple has numerous missing values.
2. Complete the Missing Values: There are several methods for completing this assignment. You can opt to manually fill in the missing values, use the attribute mean, or use the value that is most likely.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 7, July 2022

DOI: 10.17148/IJARCCE.2022.11743

D. Flowchart of implementation

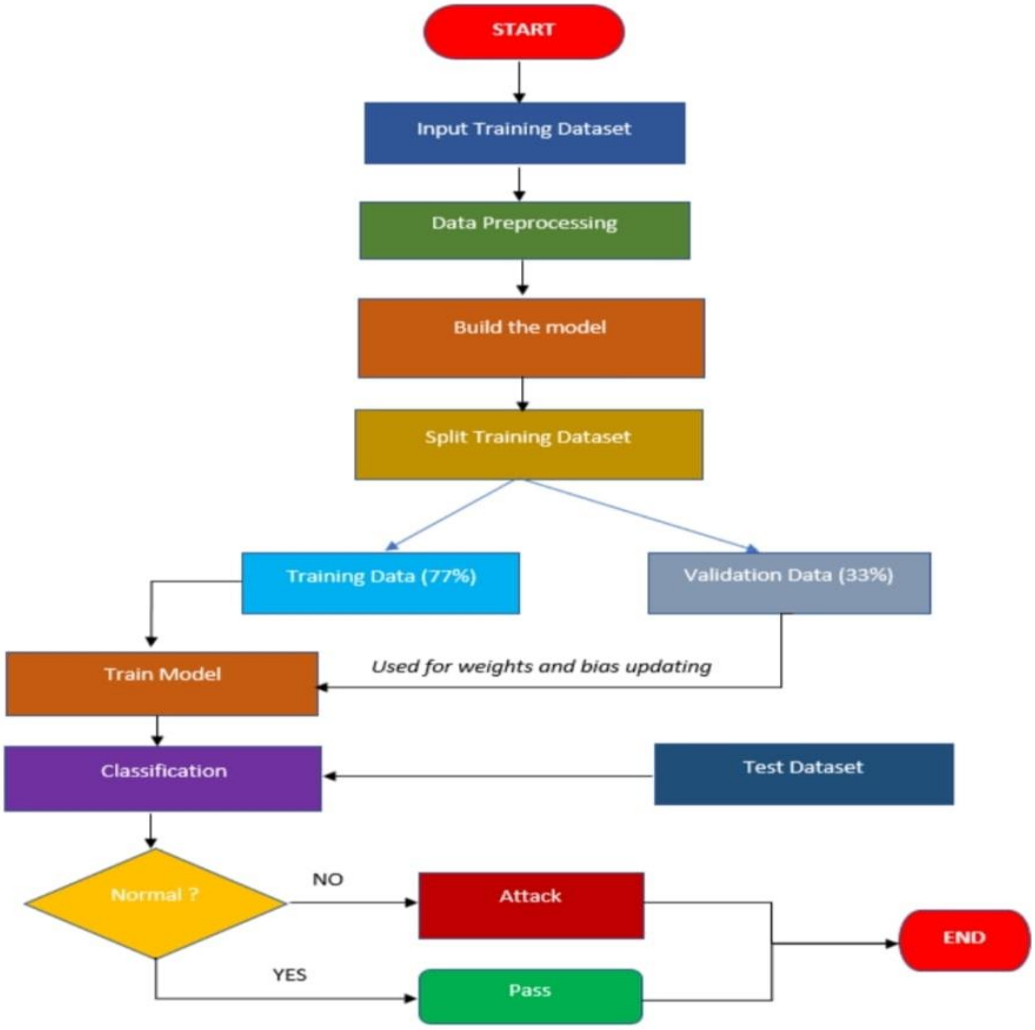

Fig: flow of implementation

Pre-processing involves encoding, normalising, and fitting the training set of samples in above figure into a data format that is appropriate for the IDP Flowchart 52 TensorFlow framework. A training subset and a validation subset will be created from the complete dataset after pre-processing. The BLSTM RNN network is trained using the first set, or the training subset, and validated using the second set, or the validation subset. subsequently the validation stage, a new testing dataset is used to evaluate the model's performance measures, including classification accuracy.

IV. IMPLEMENTATION

A python language for object oriented procedure-oriented programming, is used to implement the project. By establishing partitioned memory areas containing both data and function, which may be used as a model for building such modules on demand, object-oriented programming is a technique that offers a means to modularize programmes.
Programming in Python is used to carry out this project. Garbage collection and dynamic typing are features of Python. Including procedural, object-oriented, and functional programming, it supports a variety of programming paradigms. Python's extensive standard library has led to the moniker "batteries included" being applied to the language. This project utilises machine learning techniques.
The ultimate install the software package and it is intended the enjoyment of the environment, target audience and smooth running system off are all considered aspects of software implementation. The individuals are unsure of the software's intended purpose, which is to make their jobs simpler.

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 7, July 2022

DOI: 10.17148/IJARCCE.2022.11743

- The system's advantages must be understood by the active user.
- Their faith in the programme increased.
- The user receives appropriate instruction to ensure his comfort in the programme.

The user has to be informed that the server software needs to be operating on the server in order to obtain the findings before they can examine the system. If the server objects, the actual processes won't take place. Is not executing on the server.

Machine learning (ML), an area of artificial intelligence (AI), is frequently utilised in industry and academics to find solutions to a wide range of issues. But in computer science, artificial intelligence is not a novel idea. The notorious question, "Do machines have minds?," posed by Alan Turing in 1950 marked the beginning of the phenomenon. Later, AI began to concentrate on a variety of domains. The data-driven approach to machine learning is highly popular since there is a wealth of data available. As stated by Mitchell, ML is "With regard to that class of tasks T as well as performance metrics P, a machine program to gain knowledge via experience In case of performance at tasks in that class of tasks T increases with P as a metric of T. time. Predicting whether or not an email is spam is represented by the task "T." The "P," which refers for performance, is computed as the percentage of emails that were successfully identified. The "E," which stands for experience, is the training data set.

V. RESULT AND ANALYSIS

The proposed system uses one of the most current and broadest deep learning frameworks. Two experiments have been designed to study the performance of the RNN-IDS model for binary classification (Normal, anomaly) and five-category classification, such as Normal, DoS, R2L, U2R and Probe. In order to compare with other machine learning methods, contrast experiments are designed at the same time. In the binary classification experiments, we have compared the performance with an ANN, naive Bayesian, random forest, multi-layer perceptron, support vector machine and other machine learning methods. In the same way, we analyse the multi-classification of the RNN-IDS model based on the NSL-KDD dataset. By contrast, we study the performance of the ANN, naive Bayesian, random forest, multi-layer perceptron, support vector machine and other machine learning methods in the five-category classification. Finally, we compare the performance of the RNN-IDS model with traditional methods.

VI. CONCLUSION

We are investigating Intrusion detection attacks in this project, which is titled "Deep Learning Approach for Intrusion Detection Using Recurrent Neural Network." We have gathered raw datasets from kaggle.com in order to build this project. There are 42 parameters in these datasets, including things like duration, protocol type, and flags. The type of data and its class labels are being examined utilising recurrent neural network by gathering all of this dataset. For training and testing data, long short-term memory (LSTM) is employed. A total of 75% of the datasets in this project are for training, and 25% are for testing. Using a graph, we can display the accuracy of the results from our tests.

REFERENCES

[1] Network Intrusion Detection Using Deep Neural Networks M.Ponkarthika and Dr.V.R.Saraswathy (Open Access Quarterly International Journal) Volume 2, Issue 2, Pages 665-673, April-June 2018

[2] Host Based Intrusion Detection System with Combined CNN/RNN Model Ashima Chawla(B), Brian Lee, Sheila Fallon, and Paul Jacob

[3] On the Effectiveness of Machine and Deep Learning for Cyber Security 2018 10th International Conference on Cyber Conflict

[4] Collective Anomaly Detection Based on Long Short-Term Memory Recurrent Neural Networks Loci Bontemps, Van Loi Cao(B), James McDermott, and Nhien-An Le-Khan

[5] A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection Anna L. Buczak, Member, IEEE, and Erhan Guven, Member, IEEE

[6] Application of Neural Networks for Intrusion Detection in Tor Networks Taro Ishitaki, Donald Elma Zi, Yi Liu , Tetsuya Oda , Leonard Barolli and Kazunori Uchida‡2015 29th International Conference on Advanced Information Networking and Applications Workshops

[7] Application of Deep Recurrent Neural Networks for Prediction of User Behaviour in Tor Network 2017 31st International Conference on Advanced Information Networking and Applications Workshops.

[8] Deep Learning based Attribute Classification Insider Threat Detection for Data Security Fanzhi Meng, Fang Lou, Yun sheng Fu 2018 IEEE Third International Conference on Data Science in Cyber space.

