Bannari Amman Institute of Technology
Sathyamangalam 638 401

Stay Ahead

Final Year Students
Project Handbook

Regulations 2018

NOVEL CYBER ATTACK DETECTION USING HYBRID DEEP LEARNING MODEL

PROJECT REPORT

Submitted by

PREETHIKA N C
SWETHA SRIDEVI N
SUBIKSHA T

In partial fulfillment for the award of the degree

of

BACHELOR OF ENGINEERING

in

ELECTRONICS AND COMMUNICATION ENGINEERING

BANNARI AMMAN INSTITUTE OF TECHNOLOGY
(An Autonomous Institution Affiliated to Anna University, Chennai)
SATHYAMANGALAM-638401

ANNA UNIVERSITY: CHENNAI 600 025

MAY 2022

BONAFIDE CERTIFICATE

Certified that this project report "**NOVEL CYBER ATTACK DETECTION USING HYBRID DEEP LEARNING MODEL**" is the bonafide work of "**PREETHIKA N C, SWETHA SRIDEVI N, SUBIKSHA T**" who carried out the project work under my supervision.

SIGNATURE	**SIGNATURE**
POONGODI C	NIRMALAKUMARI K
HEAD OF THE DEPARTMENT	**SUPERVISOR**
Dr. C.Poongodi B.E., M.E., Ph.D.,	K Nirmalakumari
Professor and Head,	Assistant Professor
Department of ECE	Department of ECE
Sathyamangalam	Sathyamangalam

Submitted for Project Viva Voce examination held on ………………

Internal Examiner	External Examiner

DECLARATION

We affirm that the project work titled “**Novel Cyber Attack Detection Using Hybrid Deep Learning Model**” being submitted in partial fulfillment for the award of the degree of **Bachelor of Engineering** in **Electronics and Communication Engineering** is the record of original work done by us under the guidance of **Mrs.NIRMALAKUMARI K**, Assistant Professor, Department of ECE. It has not formed a part of any other project work(s) submitted for the award of any degree or diploma, either in this or any other University.

(Signature of candidate)	(Signature of candidate)	(Signature of candidate)
PREETHIKA N C	**SWETHA SRIDEVI N**	**SUBIKSHA T**
(181EC216)	**(181EC280)**	**(181EC272)**

I certify that the declaration made above by the candidates is true.

(Signature of the Guide)

NIRMALAKUMARI K

ACKNOWLEDGEMENT

We would like to enunciate heartfelt thanks to our esteemed Chairman **DR.S.V.BALASUBRAMANIAM**, and the respected Director **DR.M.P.VIJAYKUMAR**, for providing excellent facilities and support during the course of study in this institute.

We are grateful to **POONGODI C, HOD, ECE,** for her valuable suggestions to carry out the project work successfully.

We wish to express our sincere thanks to **MR. BARANIDHARAN, ASSISTANT PROFESSOR,** for his constructive ideas, inspirations, encouragement and much needed technical support extended to complete our project work.

We wish to express our sincere thanks to the Faculty **MRS.NIRMALAKUAMRI K, ASSISTANT PROFESSOR, ECE,** for her constructive ideas, inspirations, encouragement, excellent guidance and much needed technical support extended to complete our project work.

We would like to thank our friends, faculty and non-teaching staff who have directly and indirectly contributed to the success of this project.

PREETHIKA N C
SWETHA SRIDEVI N
SUBIKSHA T

ABSTRACT

Nowadays, 5 billion people utilize the internet worldwide, which accounts for about 63 percent of the global population. Web clients proceed to develop as well, with the most recent information showing that the world's associated population developed by nearly 200 million in the one year to April 2022. The utilization and demand of the internet is increasing swiftly. Therefore, sensitive data is increasing day by day. As every minute passes by, the information created increases exponentially. The created information must be secure or it might lead to the divulgence of sensitive information of the users. The digital world connects everything and everybody to apps, data, purchases, services, and communication. The truth that nearly everybody in this world is currently more dependent on data and communication technology means that cybercriminals have a booming criminal opportunity. With the number of internet users increasing day by day, the number of cybercriminals is also increasing. When it comes to privacy, cyber security plays a crucial role here. Data breaches, Denial of Service attacks, Credential breaches are arduous to predict and businesses lose millions of dollars each day due to cyber attacks and leads to a bad reputation and credibility. Nobody is safe from the threat of cyber-attacks in this digitalized world. If Cybercriminals can access our computer, then they can and will easily steal our sensitive information. We need to possess some knowledge about the attacks if we wish to withstand a chance against these kinds of threats. For this reason, we have proposed the novel cyber attack detection system using a hybrid deep learning model which identifies the cyber attacks based on the various features extracted from the dataset (NSL-KDD) which contains four different attack classes. The cyber attack identification is done by a hybrid deep learning model which concatenates three individual algorithms for greater accuracy. CNN, MLP and LSTM are the three individual algorithms and then concatenated into a hybrid one which is used to detect and classify the type of attacks with greater accuracy. The web application has been developed to show the final results which are attack classes based on the input given by the user.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

CNN	Convolutional Neural Network
LSTM	Long Short Term Memory
MLP	MultiLayer Perceptron
EDA	Data Exploratory Analysis
TCP	Transmission Control Protocol
UDP	User Datagram Protocol
ICMP	Internet Control Message Protocol
DOS	Denial-of-Service
U2R	User to Root
R2L	Remote to user
DL	Deep Learning
ML	Machine Learning

CHAPTER 1
INTRODUCTION

1.1. Topic Introduction

The evolving technology and everything being available online, the attack surface has increased substantially. Cyber security is still on the rise and attackers use sophisticated tools and attacks with ease and gain access to systems and networks. Cybersecurity is essential to deal with all the threats that we face in this digital era. Its crucial that we stay safe from the attackers and hackers who may be a threat to our sensitive information and to overcome these sorts of issues we propose the project "novel cyber attack detection using a hybrid deep learning model".

1.2. Motivation

With the advent of the digital era it's becoming difficult to shield our personnel and confidential information which may be of a high revenue for the cyber attackers. With the intention of creating awareness of the Cyber attacks which are confusing to the public we decided to help them detect possible harmful attacks through our system.By using our system people can detect threats from their device and can make necessary precautions to secure their sensitive information.

1.3. Task Undertaken

Our team undertook the task to build a cyber attack detection system to help people to secure the sensitive datas from the hackers. The tasks undertaken were to build a hybrid model by combining the layers of three algorithms such as CNN, LSTM and MLP with the respective activation function to trigger the neuron. Finally, the web application has developed by using flask to show the results which are the attack classes such as DOS, Probing, U2R, R2L and Normal based on the input given by the user.

1.4. Aim & Objective

The main purpose of the proposed system is to detect network intrusions and to protect a computer network from unauthorized users, including insiders. The intrusion detector learning task is to build a predictive model (i.e. a hybrid model) capable of distinguishing between “bad” connections, called intrusions or attacks, and “good” normal connections by processing various features from the dataset. The hybrid deep learning model is used to identify the cyber attacks based on the feature extracted from the dataset.

1.5. Scope of the work

- The scope of our system is to identify the most frequent cyber attacks to ensure the security of a computer network.
- Data protection is essential because it safeguards all types of data against unauthorised access. Critical information, personally identifiable information (PII), protected health information (PHI), personal information, intellectual property, data, and government and industry information systems are all included.
- Unless we properly secure any network we will be vulnerable to malicious use and accidental damage. Private data, including trade secrets and customer information, can be exposed by hackers, malicious insiders, or poor security methods within an organisation (PII). As a result, implementing our identification system allows us to learn more about attacks as they occur.

1.6. Intended Approach

We plan to design the system in the following order:

- Finding the dataset for the hybrid model.
- Pre-process the dataset mainly to avoid missing values and for feature selection.
- Analyze the data by using the EDA process.
- Splitting the dataset for training and testing purposes.

1. Process the data for training.
2. Creation of a hybrid deep learning model by combining layers of three different models. (i.e. MLP, CNN, LSTM).
3. Predict the model by using testing data.
4. Finally, built the web app to show the results.

CHAPTER 2
LITERATURE REVIEW

- In the survey report by Yuling Cheni and Yanmiao Li, they described key literature surveys on machine learning (ML) and deep learning (DL) methods for network analysis of intrusion detection and gave a brief tutorial on each ML/DL method.
- In this paper by R.Vinayakumar, Mamoun Alazab, Firstly, they compared traditional MLAs and deep learning architectures to detect viruses, classification, and categorization using various public and private datasets. Secondly, they removed all of the dataset bias removed in the experimental analysis by using disjoint splits of the public and private datasets to train and test the model on different timescales. Thirdly, they made a significant contribution by proposing a novel image processing technique with optimal parameters.
- In this research paper by Shiji Zheng , they used Deep learning's convolutional neural network for attack detection, and a predictive analysis model that can actively learn is created. The experiment on the KDD99 dataset demonstrates that it can effectively improve the accuracy and adaptive ability of threat detection, as well as have some efficiency and improvement.
- In this paper by Atsushi Takeda, Daichi Nagasawa , they show the experimental results of evaluating the proposed method's performance using the KDD Cup 99 Dataset. The experimental results show that the proposed method is more accurate than previous studies in detecting U2R or R2L attacks.
- In this review paper by Ali Bou Nassif, Manar Abu Talib, Qassim Nasir, Fatima Mohamad Dakalbab , they present 43 different anomaly detection applications found in the selected research articles. Furthermore, they identify 29 different ML models used in anomaly detection. Then they present 22 different datasets used in anomaly detection trials, as well as many other general datasets.

After that the unsupervised anomaly detection has been implemented more widely by researchers than other classification anomaly detection systems.
Anomaly detection using ML models is rapidly developing , and many ML models have been applied by many researchers.Finally they provide recommendations and guidelines to researchers based on their analysis.

CHAPTER 3
CYBER ATTACK DETECTION USING HYBRID DEEP LEARNING MODEL

3.1. Problem Statement

Nowadays, 5 billion people utilize the internet worldwide, which accounts for about 63 percent of the global population. Web clients proceed to develop as well, with the most recent information showing that the world's associated population developed by nearly 200 million in the one year to April 2022. The utilization and demand of the internet is increasing swiftly. Therefore, sensitive data is increasing day by day. As every minute passes by, the information created increases exponentially. The created information must be secure or it might lead to the divulgence of sensitive information of the users. The digital world connects everything and everybody to apps, data, purchases, services, and communication. The truth that nearly everybody in this world is currently more dependent on data and communication technology means that cybercriminals have a booming criminal opportunity. So, not only the number of internet users are increasing day by day, the number of cybercriminals are also increasing with them. It seems there is no such protection for private data. When it comes to privacy, cyber security plays a crucial role here. Data breaches, Denial of Service attacks, Credential breaches are arduous to predict and businesses lose millions of dollars each day due to cyber attacks and leads to a bad reputation and credibility. As technology has developed, so has the dark web fortified its sophistication. It has provided a haven for cybercriminals and resulted in an increased threat on the surface Internet usage. These security threats have increased the importance of cyber security. Securing this world is essential for protecting people, organizations, habitats, and infrastructure. Cybercrime rate is increasing; hence, without cyber security, we could lose sensitive information, money, or reputation. Cyber security is as important as the current need for technology. Nobody is safe from the threat of cyber-attacks in this digitalized world. If Cybercriminals can access our computer, then they can and will easily steal our sensitive information.

We need to possess some knowledge about the attacks if we wish to withstand a chance against these kinds of threats. For this reason, we have proposed the novel cyber attack detection system using a hybrid deep learning model which identifies the cyber attacks based on the various features extracted from the dataset (NSL-KDD) which contains four different attack classes. The cyber attack identification is done by a hybrid deep learning model which concatenates three individual algorithms for greater accuracy. CNN, MLP and LSTM are the three individual algorithms and then concatenated into a hybrid one which is used to detect and classify the type of attacks with greater accuracy. The web application has been developed to show the final results which are attack classes based on the input features given by the user.

3.2. Methods

The Hybrid model has been created using tensorflow (open source library) and keras (Interface for the tensorflow) by combining the layers of three models. Activation functions such as Relu, sigmoid and softmax are used in different layers of the neural network. We train the data using the hybrid model for correlation, prediction and classification of the dataset. The Newly trained model will be ready to take in new data and feed us predictions. The result will be good or bad connections based on the features of individual TCP, UDP and ICMP connections, content features within a connection suggested by domain knowledge and the traffic features computed using a two-second time window. The combined layers of three models (i.e.MLP, CNN, LSTM) with appropriate activation function gives us greater accuracy than a single model.

3.3. Design

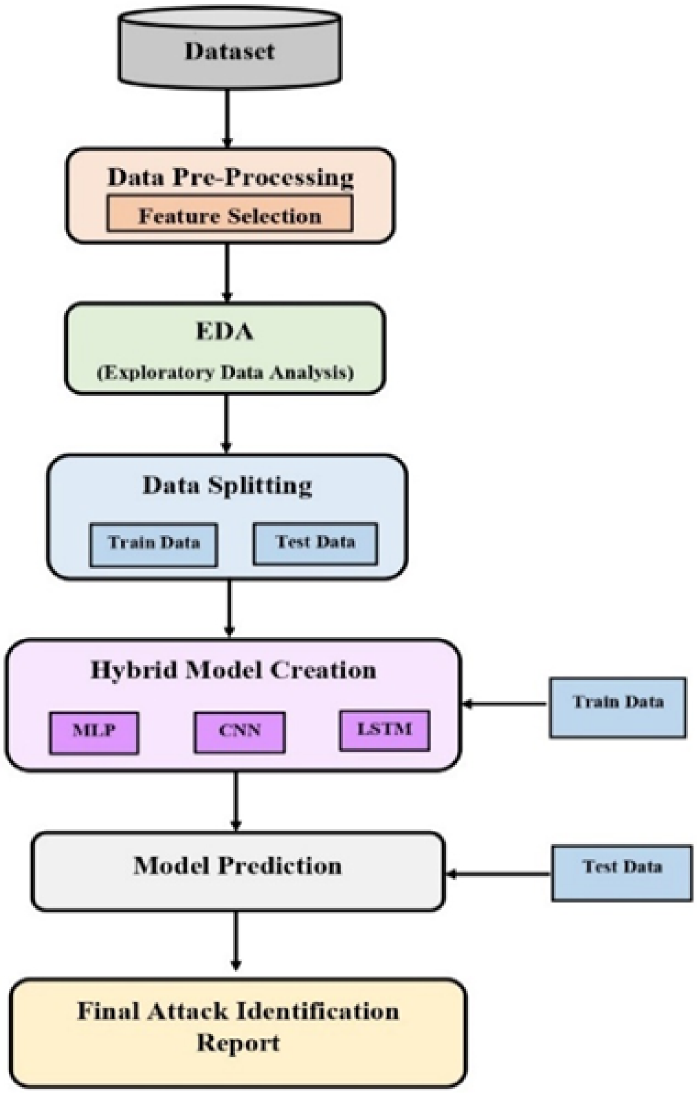


Figure (a)

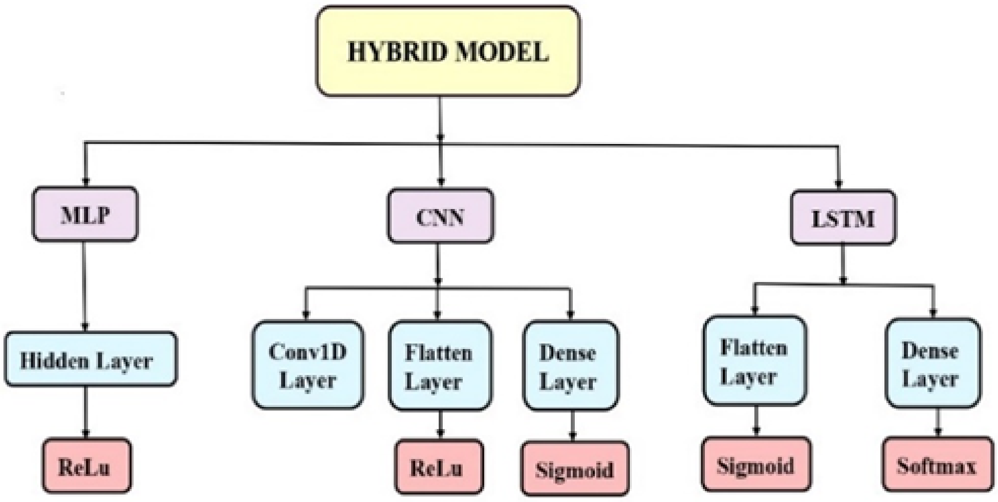


Figure (b)

The overall design of the detection system's figures are included above.

The Cyber attack identification has done by hybrid deep learning models (i.e. CNN, MLP, LSTM). NSL dataset has been taken for the project. It is a redefined version of the KDD'cup99 dataset. Hybrid model has been created for the selected dataset. Dataset has been taken for the data pre-processing. After organizing the raw data, it will be made suitable for our hybrid models. Then, Data analysis is done by the EDA process. It is the process of investigating a dataset to discover patterns, anomalies to form hypotheses based on our understanding of the dataset. EDA is used for seeing what the data can tell us before the modeling task. After analysis, data splitting algorithm has been implemented to split data into a train and test set. This approach allows us to find the model's hyper-parameter and also estimate the generalization performance. The Hybrid model was created by using tensorflow (open source library) and keras (Interface for the tensorflow). We train the data using the hybrid model for correlation, prediction and classification of the dataset. The Newly trained model will be ready to take in new data and feed us predictions..

The result will be good or bad based on the features of individual TCP, UDP and ICMP connections, content features within a connection suggested by domain knowledge and the traffic features computed using a two-second time window.

3.4. Experimental setup

The following software tools/libraries are required for the Novel Cyber Attack Detection System Using Hybrid Deep Learning Model:

- ➔ Python IDLE: Integrated development environment for Python that comes with the language's default implementation. It is employed in the construction of the hybrid deep learning model.
- ➔ TensorFlow is an open-source end-to-end machine learning platform. It is a symbolic math library that performs deep neural network training and inference tasks using dataflow and differentiable programming.
- ➔ Keras: Keras is an open-source Python interface for artificial neural networks. Keras is a user interface for the TensorFlow library.
- ➔ Pandas: Pandas is an open source Python package. It is built on Numpy, another multidimensional array support package. It's a useful data cleaning and analysis tool. It provides fast, flexible, and expressive data structures.
- ➔ NumPy: NumPy is a Python library that includes a variety of mathematical functions, random number generators, linear algebra routines, Fourier transforms, and other features.
- ➔ Matplotlib: Matplotlib is a Python library for creating static, animated, and interactive visualisations.
- ➔ Seaborn: Seaborn is a matplotlib-based Python data visualisation library. It provides an advanced interface for creating visually appealing and informative statistical graphics.

- ➔ Statsmodels: Statsmodels is a Python module that provides classes and functions for estimating various statistical models, running statistical tests, and exploring statistical data.
- ➔ Scikit-learn: Python's most useful and robust machine learning library is Scikit-learn (Sklearn). Through a consistent Python interface, it provides a set of efficient tools for machine learning and statistical modelling, such as classification, regression, clustering, and dimensionality reduction
- ➔ Flask: Flask is a microweb framework written in Python. Because it does not require the use of any specific tools or libraries, it is classified as a microframework. It is used to build the web application that displays the results.

3.5. Modeling

First, Imported all the required libraries for the project such as numpy, pandas, matplotlib etc. Selected dataset has been imported in the python program and splitted into the train and test set. Top features have been selected which helps in detecting irrelevant features, and helps in reducing overfitting and will help in the enhancement of the performance. Data analysis has been done for the selected top features such as protocol type, service, flag and attack distributions by using the EDA process. Single models have been created for the accuracy comparison with Hybrid model. Then, Hybrid Model was built by combining the layers of single models with respective activation functions. Finally, the web application has been developed by using flask to show the results which are attack classes such as DOS, Probing, U2R, R2L and Normal based on the input given by the user.

3.6. Methodology

A. Dataset

For the project, the NSL dataset was obtained. It's an enhanced version of the KDD'cup99 dataset. Each record has 41 attributes that describe various aspects of the flow and are labelled as attack type or normal.

The 42nd attribute describes the various 5 classes of network connection vectors, which are divided into one normal and four attack classes. The four attack classes are further classified as DoS, Probe, R2L, and U2R, as shown in Tables I and II. The NSL-KDD data set divides attack classes into four types:

DOS: Denial of service is an attack category, which depletes the victim‟s resources thereby making it unable to handle legitimate requests – e.g. syn flooding. Relevant features: "source bytes" and "percentage of packets with errors".
Attack Types: Back, Land, Neptune, Pod, Smurf,Teardrop,Apache2, Udp storm, Processtable, Worm.

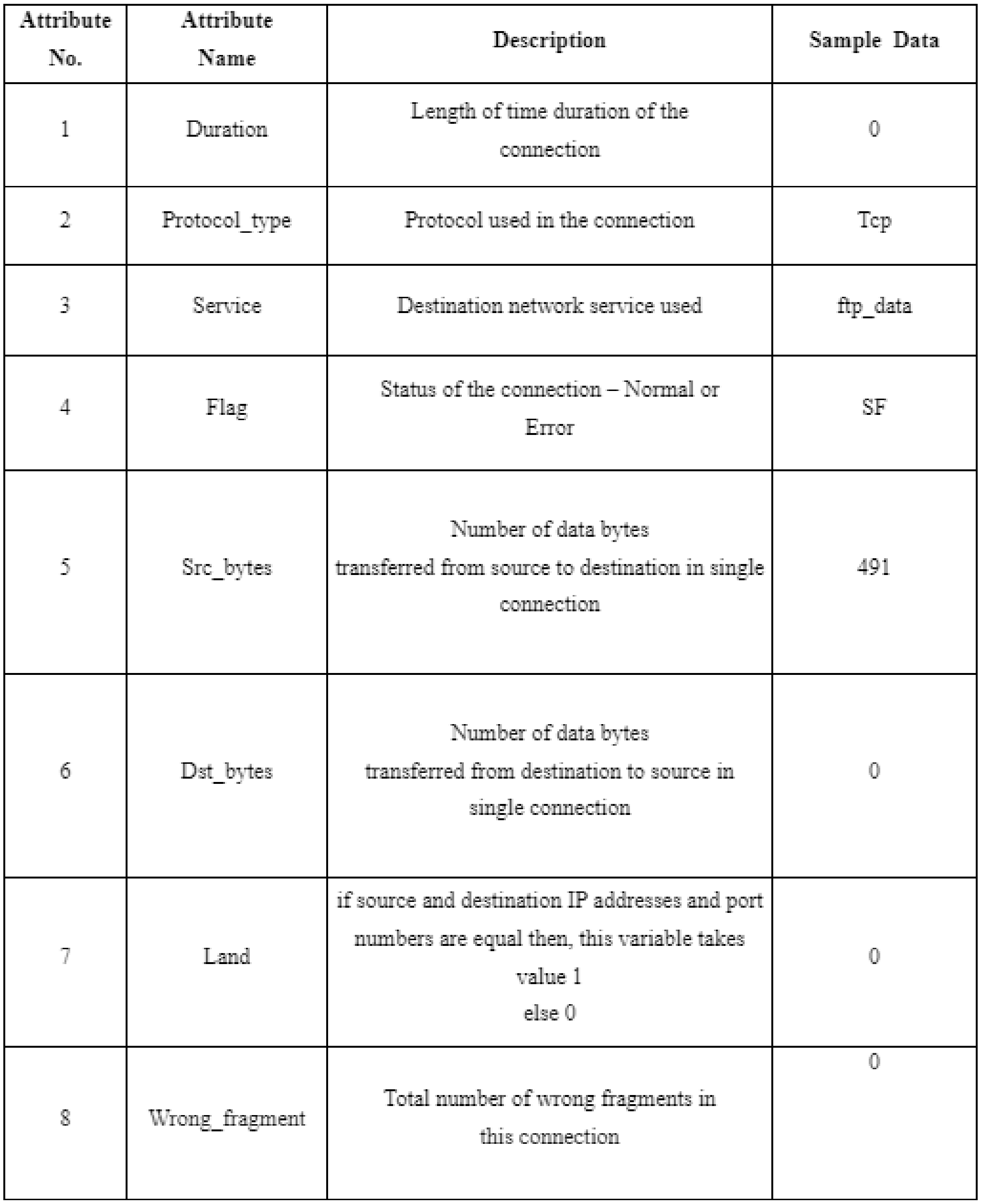
Probing: Surveillance and other probing attack‟s objective is to gain information about the remote victim e.g. port scanning. Relevant features: "duration of connection" and "source bytes".
Attack Types: Satan, Ipsweep, Nmap, Portsweep, Mscan, Saint.

U2R: Unauthorized access to local super user (root) privileges is an attack type, by which an attacker uses a normal account to login into a victim system and tries to gain root/administrator privileges by exploiting some vulnerability in the victim e.g. buffer overflow attacks. Relevant features: "number of file creations" and "number of shell prompts invoked,".
Attack Types: Buffer_overflow, Loadmodule, Rootkit, Perl, Sql attack, Xterm, Ps.

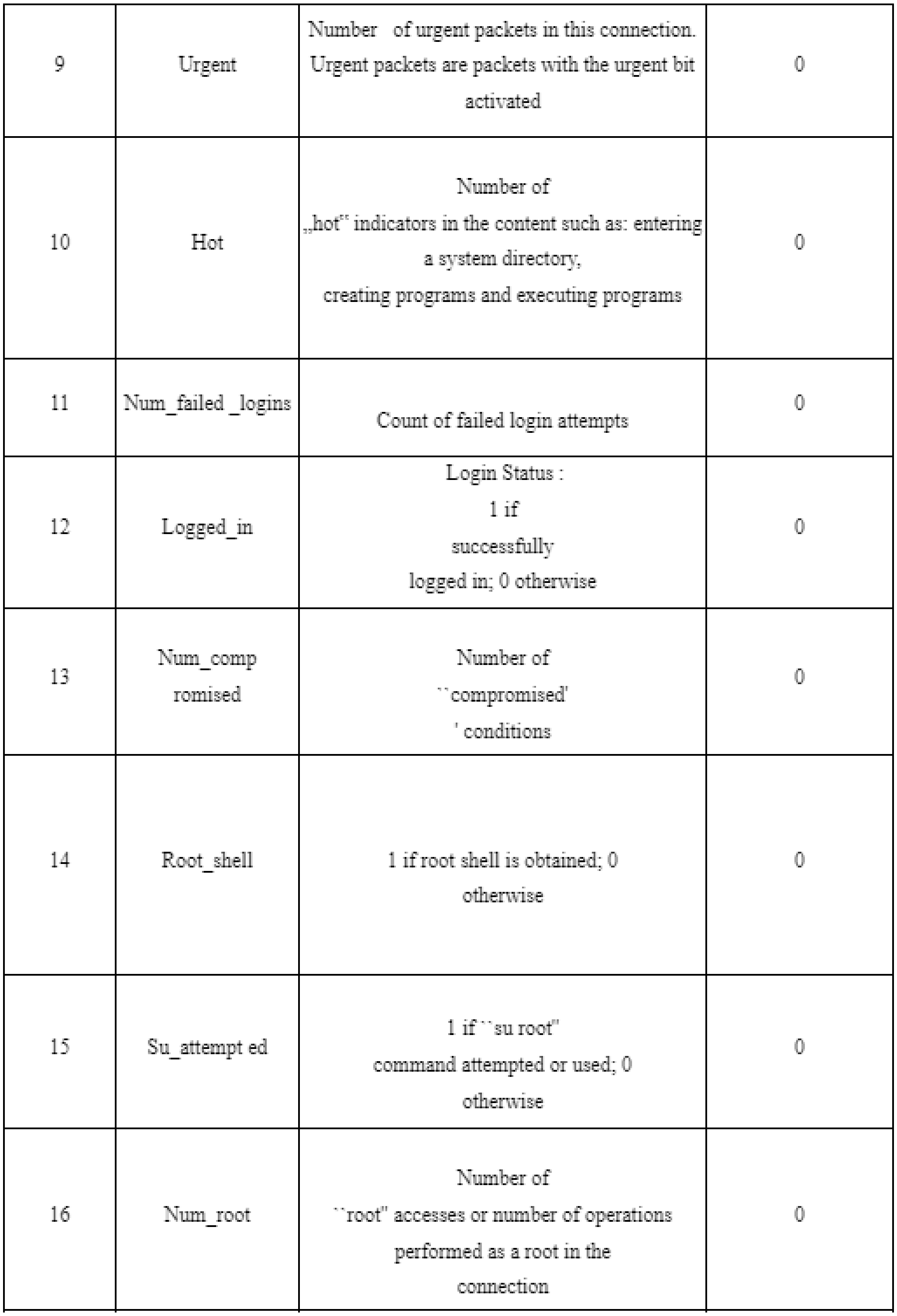
R2L: unauthorized access from a remote machine, the attacker intrudes into a remote machine and gains local access to the victim machine. E.g. password guessing Relevant features: Network level features – "duration of connection" and "service requested" and host level features - "number of failed login attempts".

Attack Types: Guess_Password, Ftp_write, Imap, Phf, Multihop, Warezmaster, Warezclient, Spy, Xlock, Xsnoop, Snmp guess, Snmp getattack, Httptunnel, Sendmail, Named.

Table I

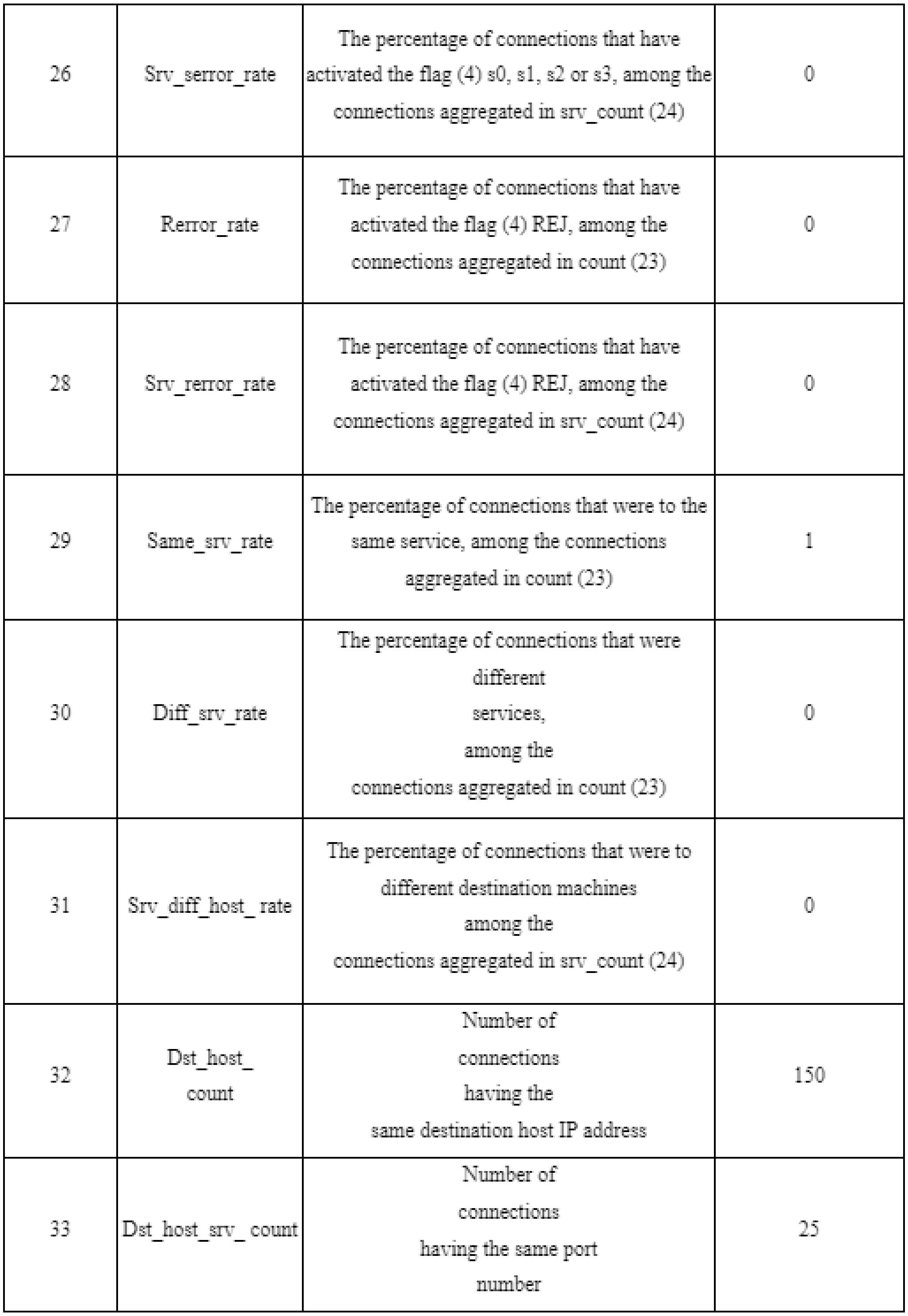


Attribute No.	Attribute Name	Description	Sample Data
1	Duration	Length of time duration of the connection	0
2	Protocol_type	Protocol used in the connection	Tcp
3	Service	Destination network service used	ftp_data
4	Flag	Status of the connection – Normal or Error	SF
5	Src_bytes	Number of data bytes transferred from source to destination in single connection	491
6	Dst_bytes	Number of data bytes transferred from destination to source in single connection	0
7	Land	if source and destination IP addresses and port numbers are equal then, this variable takes value 1 else 0	0
8	Wrong_fragment	Total number of wrong fragments in this connection	0

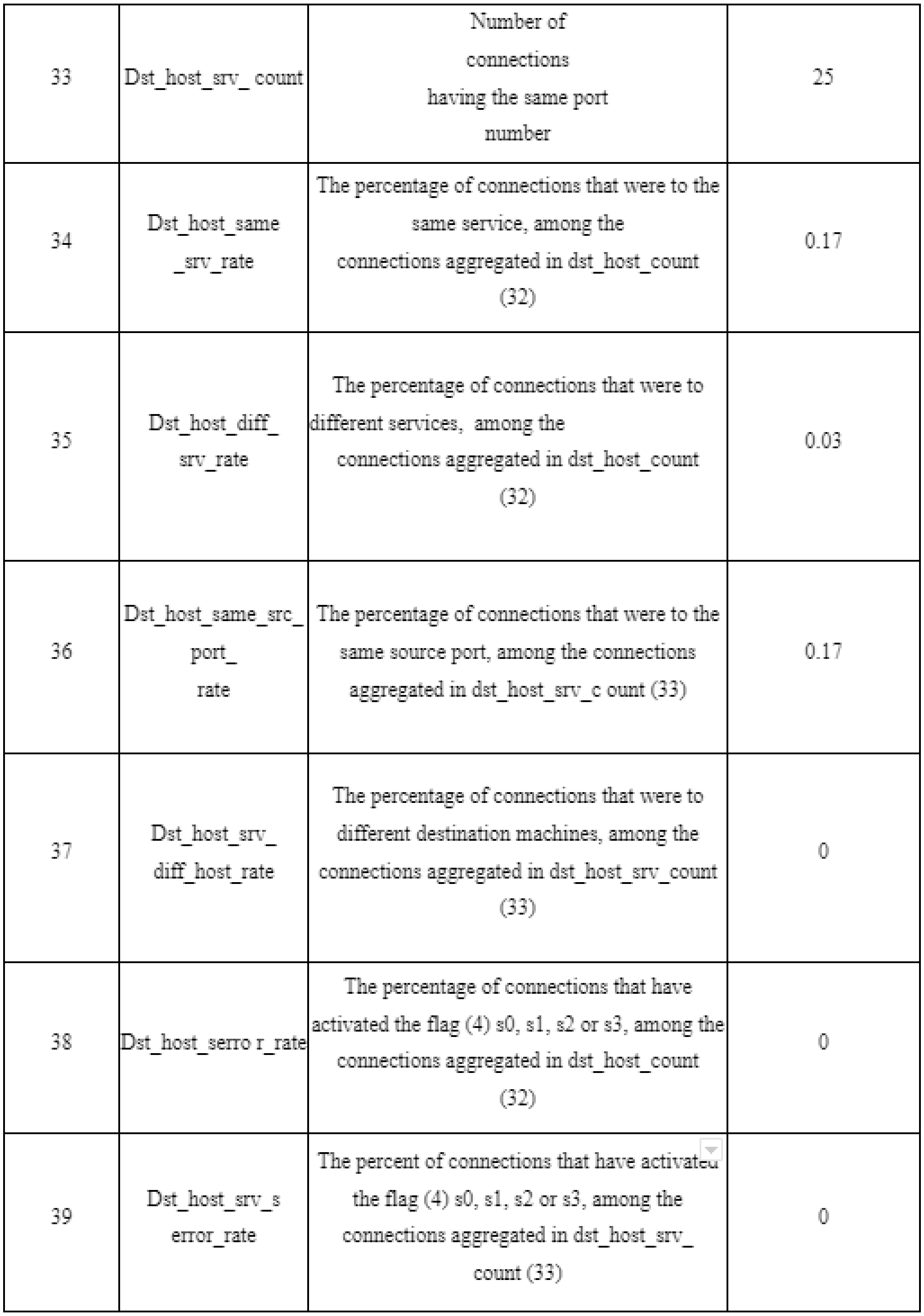


9	Urgent	Number of urgent packets in this connection. Urgent packets are packets with the urgent bit activated	0
10	Hot	Number of „hot" indicators in the content such as: entering a system directory, creating programs and executing programs	0
11	Num_failed _logins	Count of failed login attempts	0
12	Logged_in	Login Status : 1 if successfully logged in; 0 otherwise	0
13	Num_comp romised	Number of ``compromised' ' conditions	0
14	Root_shell	1 if root shell is obtained; 0 otherwise	0
15	Su_attempt ed	1 if ``su root'' command attempted or used; 0 otherwise	0
16	Num_root	Number of ``root'' accesses or number of operations performed as a root in the connection	0

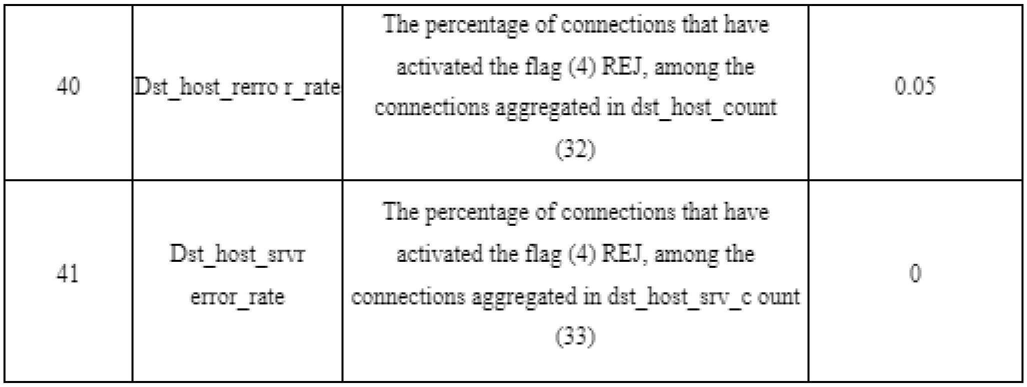
17	Num_file_creations	Number of file creation operations in the connection	0
18	Num_shells	Number of shell prompts	0
19	Num_access_files	Number of operations on access control files	0
20	Num_outbound_cmds	Number of outbound commands in an ftp session	0
21	Is_hot_logi n	1 if the login belongs to the ``hot'' list i.e., root or admin; else 0	0
22	Is_guest_lo gin	1 if the login is a ``guest'' login; 0 otherwise	0
23	Count	Number of connections to the same destination host as the current connection in the past two seconds	2
24	Srv_count	Number of connections to the same service (port number) as the current connection in the past two seconds	2
25	Serror_rate	The percentage of connections that have activated the flag (4) s0, s1, s2 or s3, among the connections aggregated in count (23)	0



26	Srv_serror_rate	The percentage of connections that have activated the flag (4) s0, s1, s2 or s3, among the connections aggregated in srv_count (24)	0
27	Rerror_rate	The percentage of connections that have activated the flag (4) REJ, among the connections aggregated in count (23)	0
28	Srv_rerror_rate	The percentage of connections that have activated the flag (4) REJ, among the connections aggregated in srv_count (24)	0
29	Same_srv_rate	The percentage of connections that were to the same service, among the connections aggregated in count (23)	1
30	Diff_srv_rate	The percentage of connections that were different services, among the connections aggregated in count (23)	0
31	Srv_diff_host_ rate	The percentage of connections that were to different destination machines among the connections aggregated in srv_count (24)	0
32	Dst_host_ count	Number of connections having the same destination host IP address	150
33	Dst_host_srv_ count	Number of connections having the same port number	25

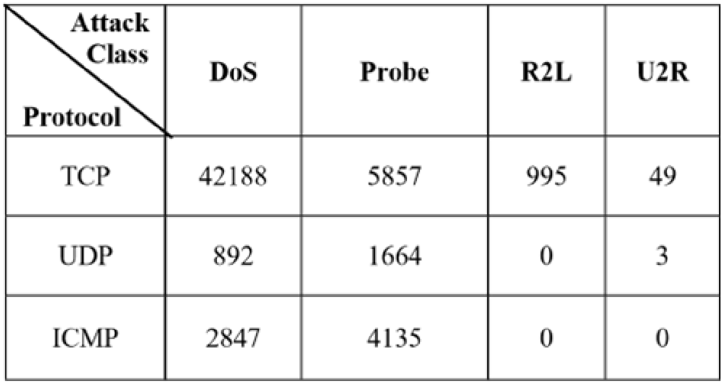


33	Dst_host_srv_ count	Number of connections having the same port number	25
34	Dst_host_same_srv_rate	The percentage of connections that were to the same service, among the connections aggregated in dst_host_count (32)	0.17
35	Dst_host_diff_srv_rate	The percentage of connections that were to different services, among the connections aggregated in dst_host_count (32)	0.03
36	Dst_host_same_src_port_rate	The percentage of connections that were to the same source port, among the connections aggregated in dst_host_srv_c ount (33)	0.17
37	Dst_host_srv_diff_host_rate	The percentage of connections that were to different destination machines, among the connections aggregated in dst_host_srv_count (33)	0
38	Dst_host_serro r_rate	The percentage of connections that have activated the flag (4) s0, s1, s2 or s3, among the connections aggregated in dst_host_count (32)	0
39	Dst_host_srv_serror_rate	The percent of connections that have activated the flag (4) s0, s1, s2 or s3, among the connections aggregated in dst_host_srv_ count (33)	0



40	Dst_host_rerro r_rate	The percentage of connections that have activated the flag (4) REJ, among the connections aggregated in dst_host_count (32)	0.05
41	Dst_host_srvr error_rate	The percentage of connections that have activated the flag (4) REJ, among the connections aggregated in dst_host_srv_c ount (33)	0

Table II



Attack Class / Protocol	DoS	Probe	R2L	U2R
TCP	42188	5857	995	49
UDP	892	1664	0	3
ICMP	2847	4135	0	0

B. Data Pre-Processing

Dataset has been taken for the data pre-processing. Top features are used here to build the model. Feature selection is used to select the most relevant features to feed the model. It also helps in detecting irrelevant features, which reduces overfitting and may lead to an improvement in performance. Furthermore, a model becomes easier to comprehend when it has less variables. After organizing the data, it will be more suitable for our hybrid models.

C. Exploratory Data Analysis

Data analysis has been done by the EDA process. It is the process of investigating a dataset to discover patterns, anomalies to form hypotheses based on our understanding of the dataset. EDA is used for seeing what the data can tell us before the modeling task. Here, we have used EDA for protocol type, service, flag and attack distributions.

D. Data Splitting

After analysis, a data splitting algorithm has been implemented to split data into a training and testing set. This approach allows us to find the model's hyper-parameter and also estimate the generalization performance.

E. Hybrid Model Creation

MLP, CNN and LSTM algorithms are used for the hybrid model creation. Multi-Layer Perceptron (MLP) is used for attack detection here, which is a better solution for cyber attack detection. This algorithm uses a number of layers so it is more secure from the hacker. Multi-layer Perceptron classifier which in the name itself connects to a Neural Network. MLP Classifier relies on an underlying Neural Network to perform the task of classification. We have used Convolutional Neural Network (CNN) here which provides better classification because it automatically detects the important features without any human supervision. CNN is a neural network that extracts input dataset features and another neural network classifies the input features. The input dataset is used by the feature extraction network. The extracted feature signals are utilized by the neural network for classification. The network consists of an input layer, followed by three convolutional and average pooling layers and followed by a soft max fully connected output layer to extract features. LSTM is used here which is capable of successfully learning the features extracted from the dataset in the training period. This capability allows the models to distinguish effectively in the normal traffic from the network attacks. Long short-term memory (LSTM) is an artificial recurrent neural network (RNN) architecture used in the field of deep learning (DL).

Unlike standard feedforward neural networks, LSTM has feedback connections. LSTM network enables input sequence data into a network, and makes predictions based on the individual time steps of the sequence data. Recurrent Neural Networks (RNN) are a type of Neural Network where the output from the previous step is fed as input to the current step. We have given more importance for feature extraction and classification to build the model. The Hybrid model was created using tensorflow (open-source library) and keras (Interface for the tensorflow) by combining the hidden layers of MLP, Cov1D, Flatten & Dense layers of CNN and the flatten & dense layers of LSTM. The fully connected layers are given to the activation function. A layer consists of small individual units called neurons. A neuron in a neural network can be better understood with the help of biological neurons. An artificial neuron is similar to a biological neuron. It receives input from the other neurons, performs processing, and produces an output. Different layers perform different transformations on the inputs. In MLP, hidden layers are used to perform nonlinear transformations of the inputs entered into the network. In CNN, Conv1D layer is used here to create a convolution kernel that is convolved with the layer input over a single spatial dimension to produce a tensor of outputs. Next, the Flatten layer is used to convert the data into a 1-dimensional array for inputting it to the next layer which is a dense layer here. Dense Layer is used to classify a dataset based on output from convolutional layers. Each Layer in the Neural Network contains neurons, which compute the weighted average of its input and this weighted average is passed through a nonlinear function. In LSTM, the same flatten and dense layers are used with different activation functions. An activation function is a function that is added into an artificial neural network in order to help the network learn complex patterns in the data. When compared with a neuron-based model that is in our brains, the activation function is at the end deciding what is to be fired to the next neuron. Activation functions such as Relu are used in the Hidden Layers of MLP, ReLu & sigmoid are used in the different layers of CNN and sigmoid & softmax are used in the various layers of LSTM.

The rectified linear activation function or ReLU for short is a piecewise linear function that will output the input directly if it is positive, otherwise, it will output zero. The sigmoid activation function is also called the logistic function. It is the same function used in the logistic regression classification algorithm. The function takes any real value as input and outputs values in the range 0 to 1. The softmax function is used as the activation function in the output layer of neural network models that predict a multinomial probability distribution. That is, softmax is used as the activation function for multi-class classification problems where class membership is required on more than two class labels. Here, we have used 5 classes. The data has been trained by using the hybrid model for correlation, prediction and classification of the dataset.

The NSL-KDD dataset contains 125972 training records and 22543 testing records which makes a total of 148515 records. And the epoch size taken for the model is 20. Here, epoch indicates the number of passes of the entire training dataset. We cannot pass the entire dataset into the neural network at once. So, we divide the training dataset into batches. 3500 samples are taken for processing in a single batch. The iteration is calculated as 35. Iterations are the number of batches needed to complete one epoch. The Newly trained model will be ready to take in new data and feed us predictions. The result will be good or bad connections based on the features of individual TCP, UDP and ICMP connections, content features within a connection suggested by domain knowledge and the traffic features computed using a two-second time window. The combined layers of three models (i.e.MLP, CNN, LSTM) with appropriate activation function gives us greater accuracy than a single model. Finally, the web application has developed by using flask to show the results which is attack classes based on the input features given by the user.

3.7. Operational Algorithm

System operation is executed with the following steps:

Step-1: Imported all the dependent libraries such as numpy, pandas, matplotlib etc. for the further implementation.

Step-2: Imported the NSL-KDD dataset for the purpose of splitting.

Step-3: Splitted the train and test dataset to build the model.

Step-4: Selected the top features from the dataset inorder to select the most relevant features to feed the model.

Step-5: Analysed the top features from the dataset such as protocol type, service, flag and attack distributions by using EDA process.

Step-6: Created the single models such as LSTM, MLP and CNN with different layers and its respective activation function for classification.

Step-7: Built the Hybrid model by combining the different layers of a single model for greater accuracy.

Step-8: Built the web application by using flask to show the output results which are attack classes based on the input features given by the user.

3.8. Feasibility analysis

Hybrid methods combine two or more DL and/or soft computing methods for higher performance and optimum results. In fact, hybrid methods benefit from the advantage of two or more methods that achieve better performance. Here, hybrid methods contain one unit for classification and one unit for the optimization and another unit for reaching an accurate output. Therefore, it can be claimed that hybrid methods contain different single methods and form a method with higher flexibility with a high capability compared with single methods. Hybrid methods have become more popular due to their high potential and capability.

For Example: Hybrid methods are the same as a company with different employees with different expertise to achieve a single goal. This shows the feasibility of our proposed system.

CHAPTER 4

RESULTS AND DISCUSSION

First, Exploratory Data Analysis results are taken for the discussion.

(i.e. protocol_type, service, flag, attack).

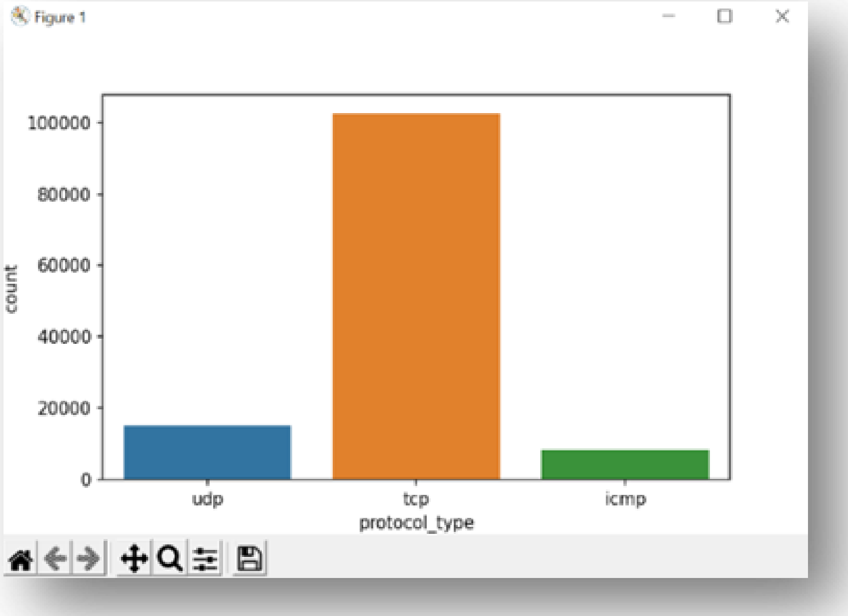


Figure (c)

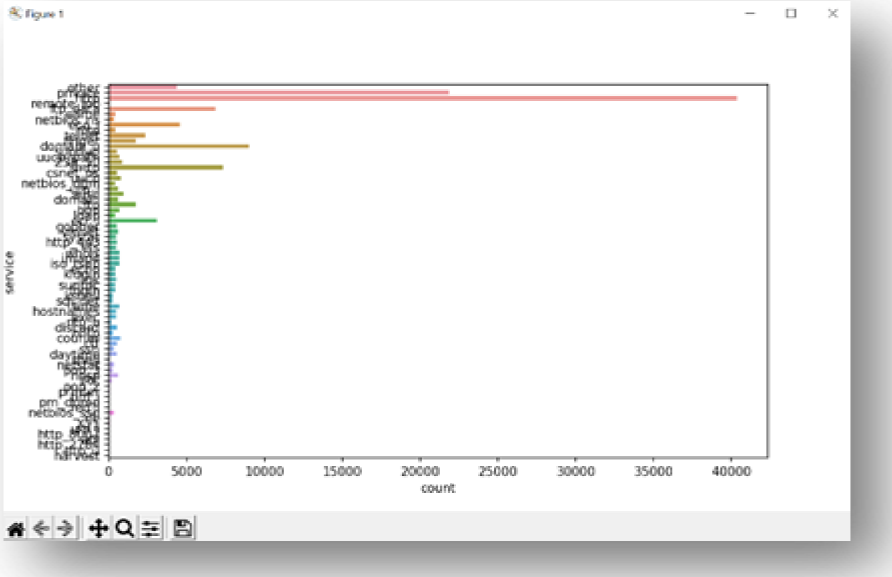


Figure (d)

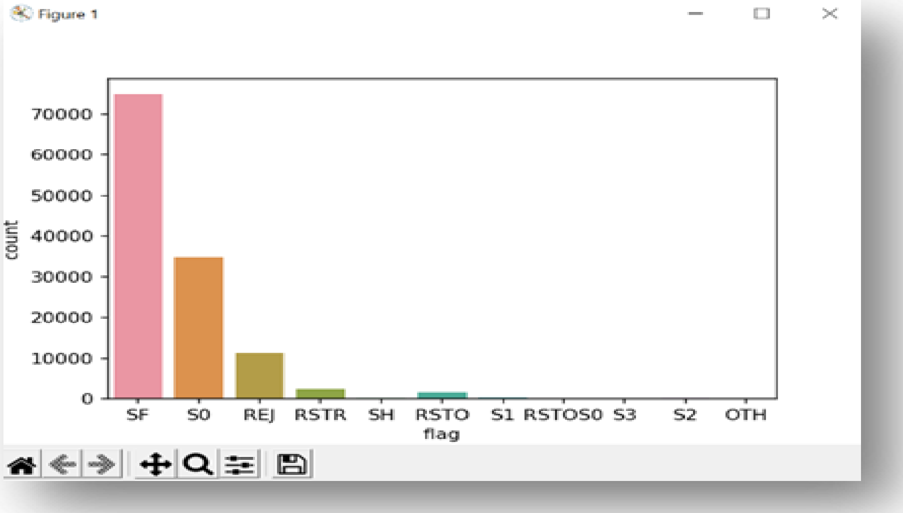


Figure (e)

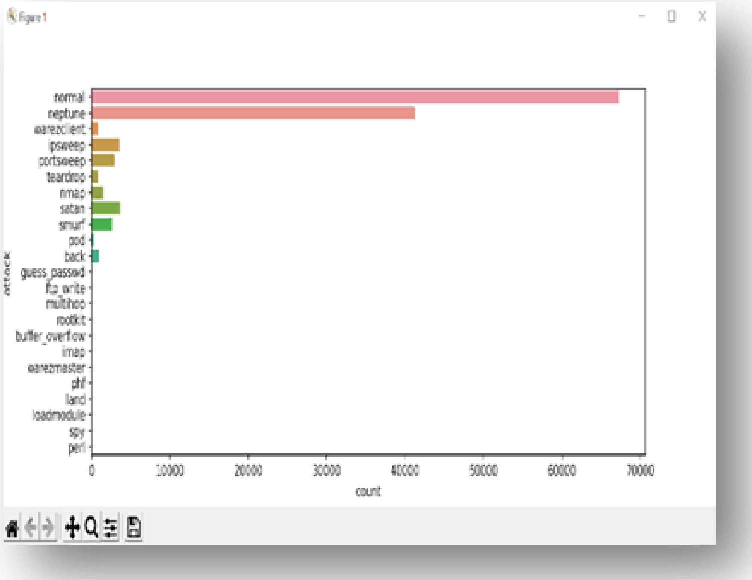


Figure (f)

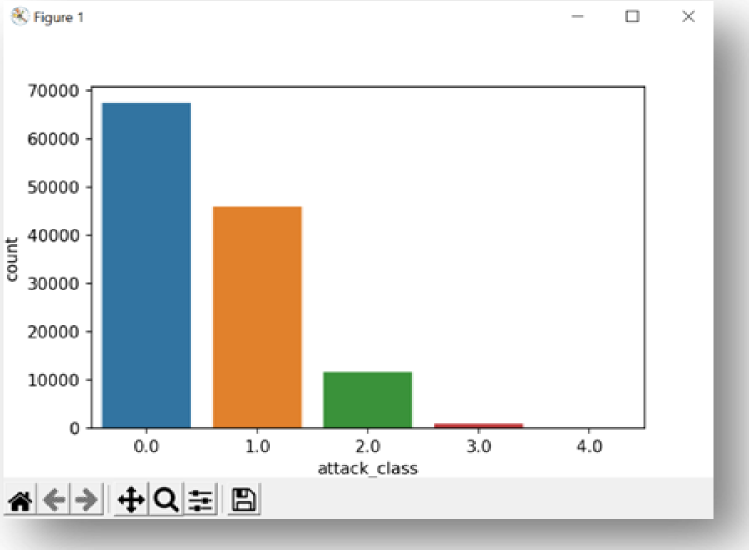


Figure (g)

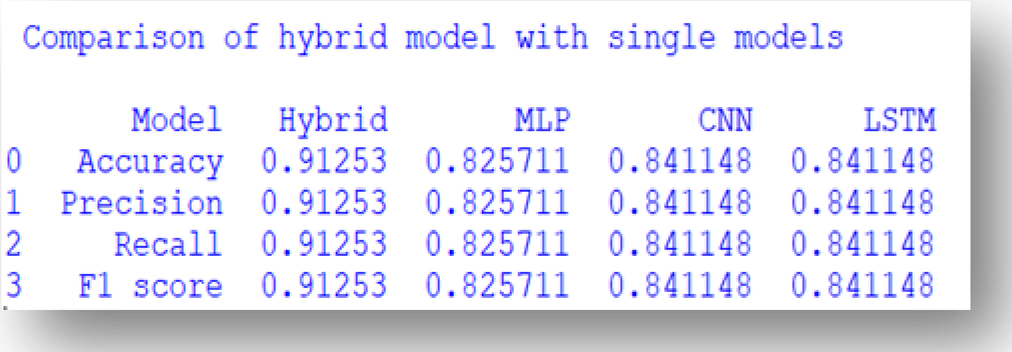
Figure(g) shows that 0.0 indicates normal connection,

1.0 indicates Dos attack,

2.0 indicates probing,

3.0 indicates U2R and

4.0 indicates R2L.



```
Comparison of hybrid model with single models

       Model   Hybrid       MLP       CNN      LSTM
0   Accuracy  0.91253  0.825711  0.841148  0.841148
1  Precision  0.91253  0.825711  0.841148  0.841148
2     Recall  0.91253  0.825711  0.841148  0.841148
3   F1 score  0.91253  0.825711  0.841148  0.841148
```

Figure (h)

The comparison image Figure(h) shows that the hybrid model performs better than a single model because of a combined deep neural network. Here Accuracy, Precision, Recall and F1 Score values are the same because of the weighted average.

A. Performance Metrics:

i. Confusion Matrix:

It is the easiest way to measure the performance of a classification problem where the output can be of two or more types of classes.

ii. Classification Accuracy:

It is the most common performance metric for classification algorithms. It may be defined as the number of correct predictions made as a ratio of all predictions made.

iii. Classification Report:

This report consists of the scores of Precisions, Recall, and F1. They are explained as follows −

a) Precision: Precision, used in classifications, defined as the number of correct records classified in the particular class.

b) Recall: Recall is defined as the number of positives returned by our DL model.

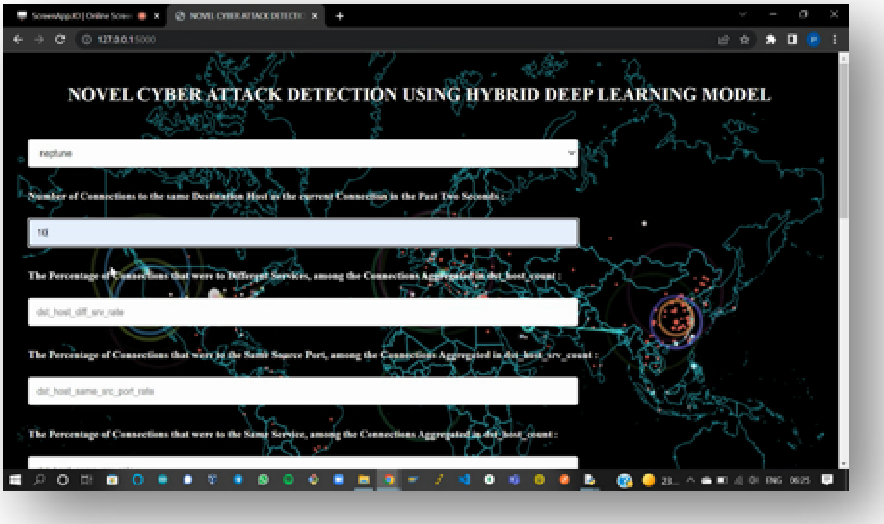
c) F1 Score: This score will give us the harmonic mean of precision and recall. Mathematically, F1 score is the weighted average of the precision and recall.

A Web Application has been developed by using flask which predicts the attack classes based on the features given by the user as shown in the Figure(i) and Figure(j).

The input features used in the web application are listed below:

- attack type
- count
- dst_host_diff_srv_rate
- dst_host_same_src_port_rate
- dst_host_same_srv_rate
- dst_host_srv_count
- flag
- logged_in
- same_srv_rate
- serror_rate
- Protocol_type

The final prediction will be the attack classes such as normal or DOS or Probe or U2R or R2L as shown in the figure(j).



NOVEL CYBER ATTACK DETECTION USING HYBRID DEEP LEARNING MODEL

neptune

Number of Connections to the same Destination Host as the current Connection in the Past Two Seconds :

10

The Percentage of Connections that were to Different Services, among the Connections Aggregated in dst_host_count :

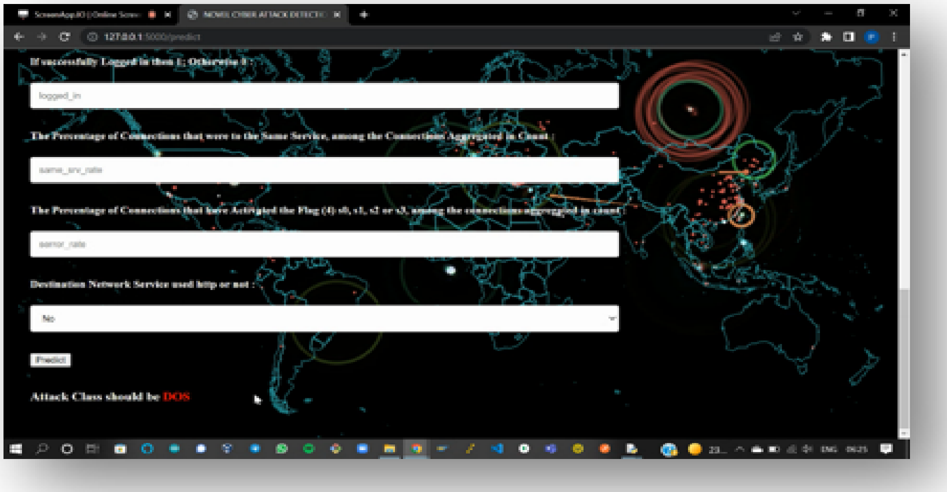
dst_host_diff_srv_rate

The Percentage of Connections that were to the Same Source Port, among the Connections Aggregated in dst_host_srv_count :

dst_host_same_src_port_rate

The Percentage of Connections that were to the Same Service, among the Connections Aggregated in dst_host_count :

Figure (i)



If successfully Logged in then 1; Otherwise 0 :

logged_in

The Percentage of Connections that were to the Same Service, among the Connections Aggregated in Count :

same_srv_rate

The Percentage of Connections that have Activated the Flag (4) s0, s1, s2 or s3, among the connections aggregated in count :

serror_rate

Destination Network Service used http or not :

No

Predict

Attack Class should be DOS

Figure (j)

CHAPTER 5
CONCLUSION

The main purpose of the proposed system is to detect network intrusions and to protect a computer network from unauthorized users, including insiders. The intrusion detector learning task is to build a predictive model (i.e. a hybrid model) capable of distinguishing between “bad” connections, called intrusions or attacks, and “good” normal connections by processing various features from the dataset. The hybrid deep learning model is used to identify the cyber attacks based on the feature extracted from the dataset.

The analysis results on the NSL-KDD dataset show that it is the best data set to simulate and test the performance of cyber attack detection. The Hybrid approach for cyber attack detection increases the accuracy rate and the web application predicts the attack classes directly through browser based on the features given by the user. This analysis conducted on the NSL-KDD dataset with the help of figures and tables helps any researcher to have a clear understanding of the dataset. It also brings to light that most of the attacks are launched using the inherent drawbacks of the TCP protocol. Our research also provides a logging system so that it will be easier to protect against network based attacks in the future. In future, it is proposed to conduct an exploration on the possibility of employing optimizing techniques to develop an attack detection model having a better accuracy rate.

REFERENCES

[1] Nebhen Jamel, Sidra Abbas,” A Hybrid Approach for Network Intrusion Detection” in Computers, Materials & Continua, CMC, 2022, vol.70, no.1.

[2] A. B. Nassif, M. A. Talib, Q. Nasir and F. M. Dakalbab, "Machine Learning for Anomaly Detection: A Systematic Review" in IEEE Access, vol. 9, pp. 78658-78700,2021,doi:10.1109/ACCESS.2021.3083060.

[3] F. M. M. Mokbal, W. Dan, A. Imran, L. Jiuchuan, F. Akhtar and W. Xiaoxi, "MLPXSS: An Integrated XSS-Based Attack Detection Scheme in Web Applications Using Multilayer Perceptron Technique" in IEEE Access, vol. 7, pp. 100567-100580, 2019, doi: 10.1109/ACCESS.2019.2927417.

[4] C. Chen, K. Zhang, K. Yuan, L. Zhu and M. Qian, "Novel Detection Scheme Design Considering Cyber Attacks on Load Frequency Control" in IEEE Transactions on Industrial Informatics, vol. 14, no. 5, pp. 1932-1941, May 2018, doi: 10.1109/TII.2017.2765313.

[5] B. Nugraha and R. N. Murthy, "Deep Learning-based Slow DDoS Attack Detection in SDN-based Networks," 2020 IEEE Conference on Network Function Virtualization and Software Defined Networks (NFV-SDN), 2020, pp.51-56,doi:10.1109/NFV-SDN50289.2020.9289894.

[6] N. Ahuja, G. Singal and D. Mukhopadhyay, "DLSDN: Deep Learning for DDOS attack detection in Software Defined Networking," 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence), 2021, pp. 683-688,doi:10.1109/Confluence51648.2021.9376879.

[7] S. Al-Emadi, A. Al-Mohannadi and F. Al-Senaid, "Using Deep Learning Techniques for Network Intrusion Detection," 2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT), 2020, pp.171-176,doi:10.1109/ICIoT48696.2020.9089524.

[8] N. Chockwanich and V. Visoottiviseth, "Intrusion Detection by Deep Learning with TensorFlow" 2019 21st International Conference on Advanced Communication Technology (ICACT), 2019,pp.654-659,doi: 10.23919/ICACT.2019.8701969.

[9] B. Patel, Z. Somani, S. A. Ajila and C. -H. Lung, "Hybrid Relabeled Model for Network Intrusion Detection," 2018 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData), 2018, pp. 872-877, doi: 10.1109/Cybermatics_2018.2018.00167.

[10] R. Vinayakumar, M. Alazab, K. P. Soman, P. Poornachandran and S. Venkatraman, "Robust Intelligent Malware Detection Using Deep Learning," in IEEE Access, vol. 7, pp. 46717-46738, 2019, doi: 10.1109/ACCESS.2019.2906934.

[11] R. Vinayakumar, M. Alazab, K. P. Soman, P. Poornachandran, A. Al-Nemrat and S. Venkatraman, "Deep Learning Approach for Intelligent Intrusion Detection System," in IEEE Access, vol. 7, pp. 41525-41550, 2019, doi: 10.1109/ACCESS.2019.2895334.

[12] J. Kim, Y. Shin, and E. Choi, “An Intrusion Detection Model based on a Convolutional Neural in Network” Journal of Multimedia Information System, vol. 6, no. 4. Korea Multimedia Society - English Version Journal, pp. 165–172, 31-Dec-2019.

[13] A. A. Elsaeidy, A. Jamalipour and K. S. Munasinghe, "A Hybrid Deep Learning Approach for Replay and DDoS Attack Detection in a Smart City," in IEEE Access, vol. 9, pp. 154864-154875, 2021, doi: 10.1109/ACCESS.2021.3128701.

[14] P. Jisna, T. Jarin and P. N. Praveen, "Advanced Intrusion Detection Using Deep Learning-LSTM Network On Cloud Environment," 2021 Fourth International Conference on Microelectronics, Signals & Systems (ICMSS), 2021, pp. 1-6, doi: 10.1109/ICMSS53060.2021.9673607.

[15] M. Ebrahimian and R. Kashef, "A CNN-based Hybrid Model and Architecture for Shilling Attack Detection," 2021 IEEE Canadian Conference on Electrical and Computer Engineering(CCECE),2021,pp.1-7, doi:10.1109/CCECE53047.2021.9569048.

[16] M. Anwer, G. Ahmed, A. Akhunzada and S. Siddiqui, "Intrusion Detection Using Deep Learning," 2021 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME), 2021, pp. 1-6, doi: 10.1109/ICECCME52200.2021.9590852.

[17] P. Shettar, A. V. Kachavimath, M. M. Mulla, N. D. G and G. Hanchinmani, "Intrusion Detection System using MLP and Chaotic Neural Networks" 2021 International Conference on Computer Communication and Informatics (ICCCI), 2021, pp. 1-4, doi: 10.1109/ICCCI50826.2021.9457024.

[18] Y. Peng, "Application of Convolutional Neural Network in Intrusion Detection," 2020 International Conference on Advance in Ambient Computing and Intelligence (ICAACI),2020,pp.169-172, doi: 10.1109/ICAACI50733.2020.00043.

[19] J. Malik, A. Akhunzada, I. Bibi, M. Imran, A. Musaddiq and S. W. Kim, "Hybrid Deep Learning: An Efficient Reconnaissance and Surveillance Detection Mechanism in SDN," in IEEE Access, vol. 8, pp. 134695-134706, 2020 doi: 10.1109/ACCESS.2020.3009849

[20] .S.M. D. Hossain, H. Ochiai, D. Fall and Y. Kadobayashi, "LSTM-based Network Attack Detection: Performance Comparison by Hyper-parameter Values Tuning," 2020 7th IEEE International Conference on Cyber Security and Cloud Computing (CSCloud)/2020 6th IEEE International Conference on Edge Computing ,2020,pp.62-69,doi:10.1109/CSCloud-EdgeCom49738.2020.00020.

[21] Garg, K. Kaur, N. Kumar, G. Kaddoum, A. Y. Zomaya and R. Ranjan, "A Hybrid Deep Learning-Based Model for Anomaly Detection in Cloud Datacenter Networks," in IEEE Transactions on Network and Service Management, vol. 16, no. 3, pp. 924-935, Sept. 2019, doi: 10.1109/TNSM.2019.2927886.

[22] I. Jemal, M. A. Haddar, O. Cheikhrouhou and A. Mahfoudhi, "M-CNN: A New Hybrid Deep Learning Model for Web Security," 2020 IEEE/ACS 17th International Conference on Computer Systems and Applications (AICCSA), 2020, pp. 1-7, doi: 10.1109/AICCSA50499.2020.9316508.

[23] Ö. Aslan and A. A. Yilmaz, "A New Malware Classification Framework Based on Deep Learning Algorithms", in IEEE Access, vol. 9, pp. 87936-87951, 2021, doi: 10.1109/ACCESS.2021.3089586.

[24] A. R. K. Kowsik, R. K. Pateriya and P. Verma, "A Deep Learning based Hybrid Approach for DDoS Detection in Cloud Computing Environment," 2021 IEEE 4th International Conference on Computing, Power and Communication Technologies (GUCON), 2021, pp. 1-6, doi: 10.1109/GUCON50781.2021.9573817.

[25] S. N. Pakanzad and H. Monkaresi, "Providing a Hybrid Approach for Detecting Malicious Traffic on the Computer Networks Using Convolutional Neural Networks," 2020 28th Iranian Conference on Electrical Engineering (ICEE), 2020, pp. 1-6, doi: 10.1109/ICEE50131.2020.9260686.

[26] Naser, Shaymaa, Abdulwahab, Aalaa, Ali, Yossra, “Deep learning model for cyber-attacks detection method in wireless sensor networks”in periodical of Engineering and natural sciences, 2022, pp.251-259,doi:10.21533/pen.v10i2.2838.

[27] Zhou, Yiyun, Han, Meng, Liu, Liyuan,He, Jing, Wang, Yan, “Deep Learning Approach for Cyberattack Detection,”in 2018 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS): 2018 IEEE Infocom MiseNet Workshop, 2018, doi:10.1109/INFCOMW.2018.8407032.

[28] Wu, Yirui, Wei, Dabao, Feng, Jun, “ Network Attacks Detection Methods Based Deep Learning Techniques: A Survey ”in Security and Communication Networks,2020(1):1-17,doi:10.1155/2020/8872923.

[29] Bapiyev, I.M., Aitchanov, B.H.,Tereikovskyi, Ihor., Tereikovska, L.A., Korchenko, A.A, “ Deep neural networks in cyber attack detection systems”in International Journal of Civil Engineering and Technology,2017, 8(11):1086-1092.

[30] N., Jayapandian , “Cyber Secure Man-in-the-Middle Attack Intrusion Detection Using Machine Learning Algorithms,”in AI and Big Data’s Potential for Distruptive Innovation,2020,doi10.4018/978-1-5225-9687-5.ch011,pp.291-316.

Reviewer Comments and queries from first review

PMC 1 Comments:

1. Recent trends in cyber attack need to be considered for data sets.
2. Data set quantity and quality needs to be explored for different types of attacks.
3. Selection of ML algorithms and parameters needs to be justified.

PMC 2 Comments:

1. Recent trends in cyber-attack and cyber-security needs to be considered for making the data sets.
2. Data set quantity and quality for different forms of attacks needs to be carried out.
3. Selection of the ML algorithm and parameters needs to be presented with reasons.
4. Good presentation.

Reviewer Comments and queries from second review

1. Attach output video in the presentation instead of output screenshot.

Reviewer Comments and queries from third review

1. Good Presentation.
2. Reduce the Content of the Presentation.

Individual Contribution of Student 1: (PREETHIKA N C)

- Proposed an idea to create a hybrid model for detecting different cyber attacks.
- Researched the theoretical information on the hybrid models and different phases of DL algorithm to build the model more suitable for our project.
- Conducted research on the different types of deep learning models.
- Synthesized other analyst's thoughts.
- Compared and differentiated analysts' works.
- Evaluated existing exploration.
- Gained an understanding of the model designing process.
- Distinguished the proposed model comparative with others' work.
- Got a basic understanding of all the fundamental layers to build the neural network.
- Researched a lot of papers and found an effective way to implement the idea.
- Explored a lot of libraries related to our project which is available in python.
- Implemented the single models for comparison.
- Different datasets have been taken to find the suitable dataset for the hybrid model.
- Built the hybrid deep learning model by using the python language.
- Worked on the conference and presented it.

Individual Contribution of Student 2: (SWETHA SRIDEVI N)

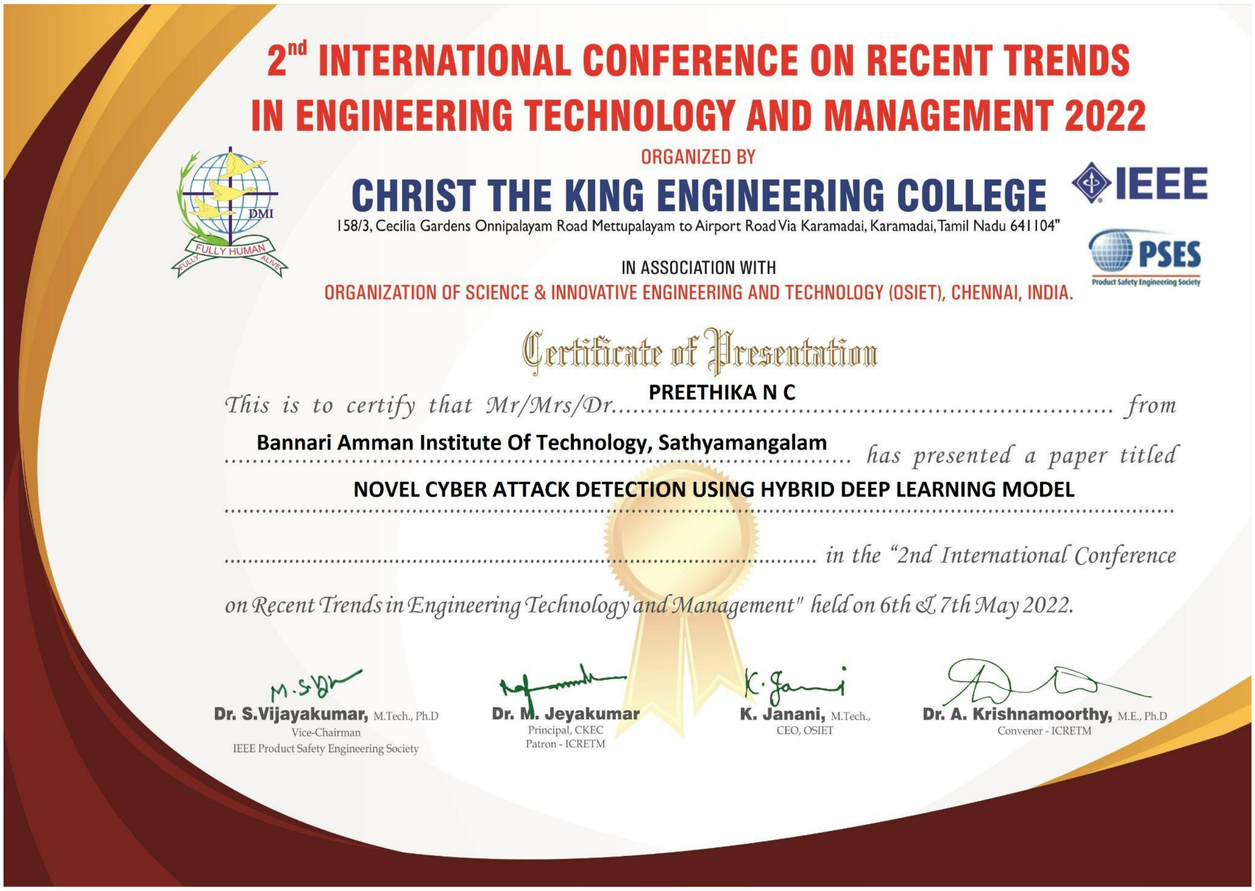
- Proposed the idea to create a hybrid model for identifying different kinds of frequently occurring cyber attacks.
- Researched the theoretical information on the different types of datasets and attack types and chose the high performing dataset that is suitable for our project.
- Explored lots of different intrusion detection datasets and attack classes.
- Found the most effective dataset and researched to create a project work plan.
- Conducted research on the different types of hybrid models.
- Gained an understanding of pre-existing intrusion detection systems.
- Curated a list of previous models with deep learning algorithms using different datasets.
- Performed the literature survey, researched a lot of papers and found a unique and effective way to implement the idea as a project.
- Contributed to the project by taking care of the networking and security aspect of the project.
- Followed the work plan and implemented the plan as a project.
- Worked on the content for publishing a paper on “Novel cyber attack detection using a hybrid deep learning model”.
- Presented the project as a paper in the ‘2nd International Conference on Recent trends in engineering technology and management 2022’.
- Finally, published the paper successfully.

Individual Contribution of Student 3: (SUBIKSHA T)

- Researched the theoretical information on the cyber attacks to get suitable dataset.
- Researched about hybrid models and to understand the project in detail.
- Researched on creating a different hybrid model to differentiate from different works.
- Proposed an idea to create a Web Application to display the accuracy,prediction etc.
- Utilized Flask to create a web application to display the results
- Researched a lot of survey papers on how cyber attacks occur and what are the frequent attacks in recent times.
- Researched some datasets and found the suitable one for our project and compared it with other datasets.
- Researched lot of papers and found a different way to implement our project.
- Distinguished the proposed model compared with other's work.
- Got an understanding on Deep Learning models.

Originality Score (Turnitin Report)

Conference Certificates & Journal publication



2nd INTERNATIONAL CONFERENCE ON RECENT TRENDS IN ENGINEERING TECHNOLOGY AND MANAGEMENT 2022

ORGANIZED BY

CHRIST THE KING ENGINEERING COLLEGE

158/3, Cecilia Gardens Onnipalayam Road Mettupalayam to Airport Road Via Karamadai, Karamadai, Tamil Nadu 641104"

IN ASSOCIATION WITH

ORGANIZATION OF SCIENCE & INNOVATIVE ENGINEERING AND TECHNOLOGY (OSIET), CHENNAI, INDIA.

Certificate of Presentation

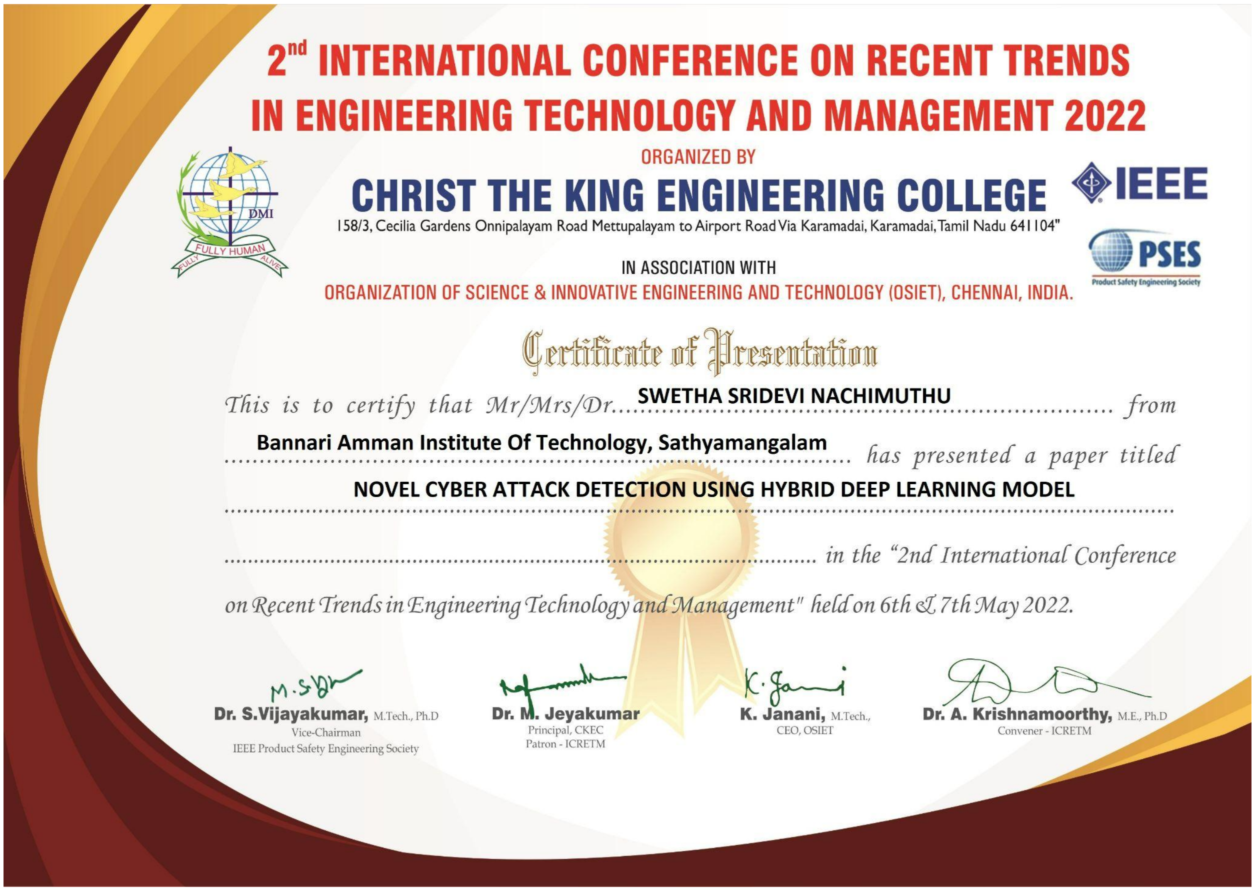
This is to certify that Mr/Mrs/Dr. PREETHIKA N C *from* Bannari Amman Institute Of Technology, Sathyamangalam *has presented a paper titled* NOVEL CYBER ATTACK DETECTION USING HYBRID DEEP LEARNING MODEL *in the "2nd International Conference on Recent Trends in Engineering Technology and Management" held on 6th & 7th May 2022.*

Dr. S.Vijayakumar, M.Tech., Ph.D
Vice-Chairman
IEEE Product Safety Engineering Society

Dr. M. Jeyakumar
Principal, CKEC
Patron - ICRETM

K. Janani, M.Tech.,
CEO, OSIET

Dr. A. Krishnamoorthy, M.E., Ph.D
Convener - ICRETM



2nd INTERNATIONAL CONFERENCE ON RECENT TRENDS IN ENGINEERING TECHNOLOGY AND MANAGEMENT 2022

ORGANIZED BY

CHRIST THE KING ENGINEERING COLLEGE

158/3, Cecilia Gardens Onnipalayam Road Mettupalayam to Airport Road Via Karamadai, Karamadai, Tamil Nadu 641104"

IN ASSOCIATION WITH

ORGANIZATION OF SCIENCE & INNOVATIVE ENGINEERING AND TECHNOLOGY (OSIET), CHENNAI, INDIA.

Certificate of Presentation

This is to certify that Mr/Mrs/Dr.......**SWETHA SRIDEVI NACHIMUTHU**...... *from*

......**Bannari Amman Institute Of Technology, Sathyamangalam**...... *has presented a paper titled*

......**NOVEL CYBER ATTACK DETECTION USING HYBRID DEEP LEARNING MODEL**......

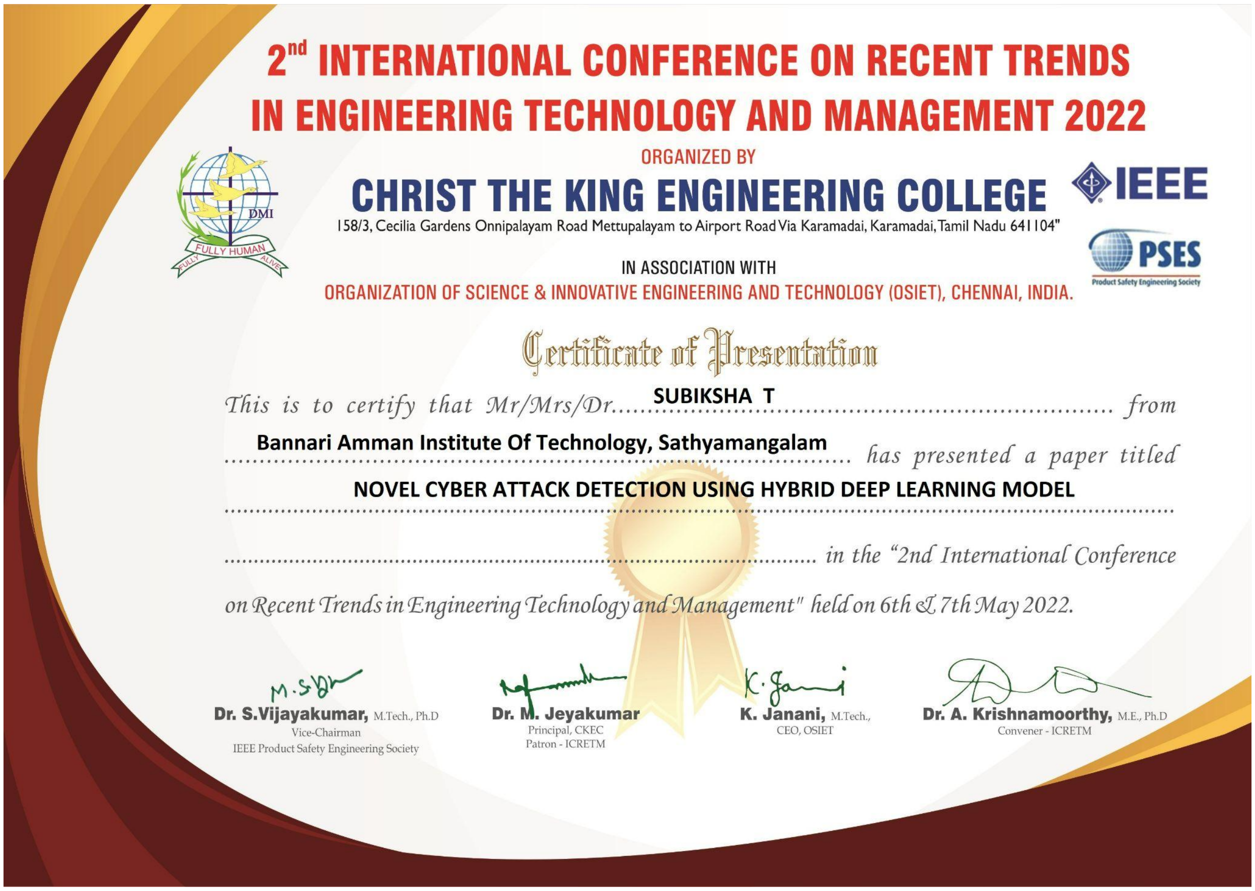
...... *in the "2nd International Conference on Recent Trends in Engineering Technology and Management" held on 6th & 7th May 2022.*

Dr. S.Vijayakumar, M.Tech., Ph.D
Vice-Chairman
IEEE Product Safety Engineering Society

Dr. M. Jeyakumar
Principal, CKEC
Patron - ICRETM

K. Janani, M.Tech.,
CEO, OSIET

Dr. A. Krishnamoorthy, M.E., Ph.D
Convener - ICRETM



2nd INTERNATIONAL CONFERENCE ON RECENT TRENDS IN ENGINEERING TECHNOLOGY AND MANAGEMENT 2022

ORGANIZED BY

CHRIST THE KING ENGINEERING COLLEGE

158/3, Cecilia Gardens Onnipalayam Road Mettupalayam to Airport Road Via Karamadai, Karamadai, Tamil Nadu 641104"

IN ASSOCIATION WITH

ORGANIZATION OF SCIENCE & INNOVATIVE ENGINEERING AND TECHNOLOGY (OSIET), CHENNAI, INDIA.

Certificate of Presentation

This is to certify that Mr/Mrs/Dr. SUBIKSHA T *from* Bannari Amman Institute Of Technology, Sathyamangalam *has presented a paper titled* NOVEL CYBER ATTACK DETECTION USING HYBRID DEEP LEARNING MODEL *in the "2nd International Conference on Recent Trends in Engineering Technology and Management" held on 6th & 7th May 2022.*

Dr. S.Vijayakumar, M.Tech., Ph.D Vice-Chairman IEEE Product Safety Engineering Society	**Dr. M. Jeyakumar** Principal, CKEC Patron - ICRETM	**K. Janani,** M.Tech., CEO, OSIET	**Dr. A. Krishnamoorthy,** M.E., Ph.D Convener - ICRETM