

PERFORMANCE ANALYSIS OF DEEP LEARNING APPROACH FOR CLASSIFYING DDOS ATTACK FROM BENIGN NETWORK TRAFFIC



**HAFSA ABBAS
02-241192-012**

**BAHRIA UNIVERSITY ISLAMABAD
KARACHI CAMPUS**

Approval for Examination

“Scholar's Name: HAFSA ABBAS Registration No. 40917”

Programme of Study:

Master of Science (Software Engineering)

Thesis Title:

PERFORMANCE ANALYSIS OF DEEP LEARNING APPROACH FOR CLASSIFYING
DDOS ATTACK FROM BENIGN NETWORK TRAFFIC

It is to certify that the above scholar's thesis has been completed to my satisfaction and, to my belief, its standard is appropriate for submission for examination. I have also conducted plagiarism test of this thesis using HEC prescribed software and found similarity index 13% that is within the permissible limit set by the HEC for the MS degree thesis. I have also found the thesis in a format recognized by the BU for the MS thesis.”

Principal Supervisor's Signature: _____

Date: 29/AUG/2022

Name: DR. OSAMA REHMAN

Author's Declaration

"I, HAFSA ABBAS hereby state that my MS thesis titled

PERFORMANCE ANALYSIS OF DEEP LEARNING APPROACH FOR CLASSIFYING
DDOS ATTACK FROM BENIGN NETWORK TRAFFIC

is my own work and has not been submitted previously by me for taking any degree from
this university BAHRIA UNIVERSITY ISLAMABAD or anywhere else in the country/world.

At any time if my statement is found to be incorrect even after my graduation, the University
has the right to withdraw/cancel my MS degree."

Name of scholar: HAFSA ABBAS

Date: 29 August 2022

Plagiarism Undertaking

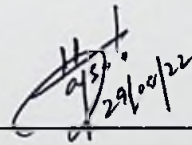
"I, solemnly declare that research work presented in the thesis titled

PERFORMANCE ANALYSIS OF DEEP LEARNING APPROACH FOR CLASSIFYING DDOS
ATTACK FROM BENIGN NETWORK TRAFFIC is solely my research work with no significant contribution from any other person. Small contribution / help wherever taken has been duly acknowledged and that complete thesis has been written by me.

I understand the zero-tolerance policy of the HEC and Bahria University towards plagiarism. Therefore, I as an Author of the above titled thesis declare that no portion of my thesis has been plagiarized and any material used as reference is properly referred / cited.

I undertake that if I am found guilty of any formal plagiarism in the above titled thesis even after award of MS degree, the university reserves the right to withdraw / revoke my MS degree and that HEC and the University has the right to publish my name on the HEC / University website on which names of scholars are placed who submitted plagiarized thesis."

Scholar / Author's Sign: _____



Name of the Scholar: _____

HAFSA ABBAS

Dedication

I dedicate this thesis to my parents and my sister who have supported me all the way as without their support this would not be possible. I would also like to dedicate this thesis to my respected teachers and especially my supervisor Dr. Osama Rehman for the guidance and countless meetings which led to the completion of this work. I dedicate my degree to my dearest Parents, Family, Friends, and respected Teachers who motivated, supported and encouraged me in every aspect of my life.

ABSTRACT

Cyber security has become a great issue in this technological world. There are several types of cyber-attacks that are present, where Distributed Denial-of-Service (DDoS) is one of the most common attack type in the cyber world. Researchers are doing their best to find a solution to get rid of DDoS attacks. With the advancement of technology day by day, millions of people across the world are relying on the internet. People are using internet in every field of life from the very basic home task to the academics' research. As the number of users are increasing day by day, security issues are also increasing. DDoS has grown more significantly than normal. DDoS attacks frequency is doubled in every year but due to COVID-19 pandemic, as everything is shifted on internet.

To identify and to take measures against DDoS attacks has become a necessary task. There is a need to make a system intelligent enough to detect the difference between the legitimate request and DDoS attack request. Blocking the traffic is not a solution. It is important to develop a technique which is intelligent enough to distinguish the normal and malicious traffic.

There are many solutions available up till now. Researchers are using different techniques to get rid of this problem. In this research, three different approaches are used to check which one is better for cyber security dataset. The dataset used is CICDDoS 2019 comprises of different DDoS attack types. The first approach is Machine Learning approach in which Random Forest algorithm are used. Second approach consists of ANN (Artificial Neural Network) and CNN (Convolutional Neural Network). The performance of CNN and RF is almost same. Accuracy obtained by using of all the three approaches are better. In some of the attack classification, the accuracy is increased up to 99.9%. Whereas ANN algorithm has an average performance for cyber security dataset. There are many anomalies occurred in the performance of ANN.

The performance parameters include Accuracy, Training Time, Testing Time and Confusion Matrix. CNN takes more time in training than RF but there is a very less chance of any

TABLE OF CONTENTS

CHAPTER 01	1
INTRODUCTION	1
1.1 BACKGROUND	2
1.2 PROBLEM STATEMENT	5
1.3 RESEARCH OBJECTIVE.....	6
1.4 RESEARCH CONTRIBUTION.....	6
1.5 THESIS ORGANIZATION.....	7
CHAPTER 02	8
RELATED WORK	8
CHAPTER 03	17
RESEARCH METHODOLOGY	17
3.1 DATASET DESCRIPTION.....	17
3.2 CLASS BALANCING.....	19
3.3 DATASET PREPROCESSING.....	20
3.4 MODEL SELECTION.....	21
3.4.1 RANDOM FOREST ALGORITHM	21
3.4.2 CNN (CONVOLUTIONAL NEURAL NETWORK)	23
3.4.3 ANN (ARTIFICIAL NEURAL NETWORK)	24
CHAPTER 04	28
RESULTS AND DISCUSSION	28
(RANDOM FOREST-MACHINE LEARNING)	28
4.1 ACCURACY	28
4.2 TRAINING TIME.....	29
4.3 TESTING TIME	30
4.4 CONFUSION MATRIX	31
4.4.1 True Positive	31
4.4.2 True Negative.....	32
4.4.3 False Positive	32
4.4.4 False Negative.....	33
CHAPTER 05	35
RESULTS AND DISCUSSION	35
(ANN AND CNN-DEEP LEARNING).....	35

5.1	ACCURACY	35
5.1.1	Layers = 5.....	35
5.1.2	Layers = 6.....	37
5.1.3	Layers = 7.....	38
5.2	TRAINING TIME.....	39
5.2.1	Layers = 5.....	39
5.2.2	Layers = 6.....	41
5.2.3	Layers = 7.....	42
5.3	TESTING TIME	44
5.3.1	Layers = 5.....	44
5.3.2	Layers = 6.....	45
5.3.3	Layers = 7.....	47
5.4	CONFUSION MATRIX	48
5.4.1	True Positive	48
5.4.1.1	Layers = 5.....	48
5.4.1.2	Layers = 6.....	50
5.4.1.3	Layers = 7.....	51
5.4.2	True Negative.....	52
5.4.2.1	Layers = 5.....	52
5.4.2.2	Layers = 6.....	54
5.4.2.3	Layers = 7.....	55
5.4.3	False Positive	57
5.4.3.1	Layer = 5	57
5.4.3.2	Layers = 6.....	58
5.4.3.3	Layers = 7.....	60
5.4.4	False Negative.....	61
5.4.4.1	Layers = 5.....	61
5.4.4.2	Layers = 6.....	63
5.4.4.3	Layers = 7.....	64
CHAPTER 06		66
Comparative Analysis.....		66
6.1	ACCURACY	67
6.1.1	UDP Attack Comparison.....	67
6.1.2	MSSQL Attack Comparison	68
6.2	TRAINING TIME.....	69

6.2.1	UDP Attack Comparison.....	69
6.2.2	MSSQL Attack Comparison	70
6.3	TESTING TIME	71
6.3.1	UDP Attack Comparison.....	71
6.3.2	MSSQL Attack Comparison	72
6.4	FALSE POSITIVE.....	73
6.4.1	UDP Attack Comparison.....	73
6.4.2	MSSQL Attack Comparison	74
6.5	FALSE NEGATIVE	75
6.5.1	UDP Attack Comparison.....	75
6.5.2	MSSQL Attack Comparison	76
	CONCLUSION.....	77
	REFERENCES.....	78