

**DEVELOPMENT OF A FEDERATED LEARNING-BASED MALWARE  
DETECTION MODEL FOR INTERCONNECTED CLOUD  
INFRASTRUCTURES**

**MUGHOLE, KALIMUMBALO DANIELLA  
20PCK02091**

**MARCH, 2023**

**DEVELOPMENT OF A FEDERATED LEARNING-BASED MALWARE  
DETECTION MODEL FOR INTERCONNECTED CLOUD  
INFRASTRUCTURES**

**BY**

**MUGHOLE, KALIMUMBALO DANIELLA  
(20PCK02091)**

**B.Eng. Génie Electrique et Informatique,  
Université Libre des Pays des Grands Lacs, Goma, DR Congo**

**A DISSERTATION SUBMITTED TO THE SCHOOL OF POSTGRADUATE  
STUDIES IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR  
THE AWARD OF MASTER OF ENGINEERING DEGREE (M.ENG.) IN  
INFORMATION AND COMMUNICATION ENGINEERING,  
DEPARTMENT OF ELECTRICAL AND INFORMATION ENGINEERING,  
COVENANT UNIVERSITY, OTA, OGUN STATE, NIGERIA**

**MARCH, 2023**

## **ACCEPTANCE**

This is to attest that this dissertation is accepted in partial fulfilment of the requirements for the award of Master of Engineering degree (M.Eng) in Information and Communication Engineering, Department of Electrical and Information Engineering, College of Engineering, Covenant University, Ota, Ogun State, Nigeria.

**Miss Adefunke F. Oyinloye**  
(Secretary, School of Postgraduate Studies)

**Signature and Date**

**Prof. Akan B. Williams**  
(Dean, School of Postgraduate Studies)

**Signature and Date**

## **DECLARATION**

I, **MUGHOLE, KALIMUMBALO DANIELLA (20PCK02091)**, declare that this research was carried out by me under the supervision of Dr. Joke A. Badejo and Dr. Kennedy O. Okokpujie of the Department of Electrical and Information Engineering, College of Engineering, Covenant University, Ota, Ogun State, Nigeria. I attest that the dissertation has not been presented either wholly or partially for the award of any degree elsewhere. All sources of data and scholarly information used in this dissertation are duly acknowledged.

**MUGHOLE, KALIMUMBALO DANIELLA**

**Signature and Date**

## **CERTIFICATION**

We certify that this dissertation titled "**DEVELOPMENT OF A FEDERATED LEARNING-BASED MALWARE DETECTION MODEL FOR INTERCONNECTED CLOUD INFRASTRUCTURES**" is an original research work carried out by **MUGHOLE, KALIMUMBALO DANIELLA (20PCK02091)**, in the Department of Electrical and Information Engineering, College of Engineering, Covenant University, Ota, Ogun State, Nigeria, under the supervision of Dr. Joke A. Badejo and Dr. Kennedy O. Okokpujie. We have examined and found this work acceptable as part of the requirements for the award of Master of Engineering in Information and Communication Engineering.

**Dr. Joke A. Badejo**  
(Supervisor)

**Signature and Date**

**Dr. Kennedy O. Okokpujie**  
(Co-Supervisor)

**Signature and Date**

**Prof. Emmanuel Adetiba**  
(Head of Department)

**Signature and Date**

**Dr. Oluwumi Adetan**  
(External Examiner)

**Signature and Date**

**Prof. Akan B. Williams**  
(Dean, School of Postgraduate Studies)

**Signature and Date**

## **DEDICATION**

This dissertation is first and foremost dedicated to God Almighty, the source of all wisdom, knowledge, and understanding, for His grace and favour throughout this research. Then to my parents KAMBALE KALIMUMBALO and KABUO ISEGHUNDI for their endless support and love.

## ACKNOWLEDGMENTS

My profound gratitude goes firstly to God Almighty for granting me the willpower to pursue this study and for imparting the wisdom, strength, support, knowledge, and health to do so, most pertinent of which is life itself.

The expression of great gratitude is addressed to my supervisor Dr. Joke A. Badejo, for her support, comments, and encouragement. Also, my deep thanks to the management of Covenant Applied Informatics and Communications - African Centre of Excellence (CApIC-ACE) for the incredible support throughout my master's studies, particularly to the director Professor Ezekiel Adebisi and the deputy director Prof. Emmanuel Adetiba.

My deep thanks also go to Prof. Emmanuel Adetiba, Dr. Kennedy Okokpujie, Dr. Victoria Oguntosin, Prof. Francis Idachaba, and Dr. Oluwadamilola Oshin, for support, corrections, and above all encouragement. I especially appreciate all my lecturers for all the labor put into teaching and mentoring me and my colleagues through all the courses they handled. I also appreciate the effort put into developing and perfecting this research through constructive criticism and feedback.

My appreciation goes as well to Prof. Takenga Claude for his support, encouragement, and advice throughout my studies.

My sincere thanks remain meritorious to my parents, sisters, brothers, and friends who helped me achieve my goals; I quote Fleurette Vivuya, Raph-Arsène Kasongo, Inès Muheruki, Justine Kahindo, Emile Kalimumbalo, Josiane Ndaghane, and Elie Lubamba. Thank you for supporting me throughout my studies and especially praying and encouraging me on all those very hard days.

May my friends Molo Mbaso, Patrick Vingi, Myriam Baiguerel, John Chishugi, Sonia Nkongho, Mariam Kanonte, and Lumbani Bondera, find here the expression of my gratitude for their support, friendship, jokes, encouragement, and prayers.

# TABLE OF CONTENTS

<b>CONTENTS</b>	<b>PAGES</b>
<b>TITLE PAGE</b>	<b>ii</b>
<b>ACCEPTANCE</b>	<b>iii</b>
<b>DECLARATION</b>	<b>iv</b>
<b>CERTIFICATION</b>	<b>v</b>
<b>DEDICATION</b>	<b>vi</b>
<b>ACKNOWLEDGMENTS</b>	<b>vii</b>
<b>TABLE OF CONTENTS</b>	<b>viii</b>
<b>LIST OF FIGURES</b>	<b>x</b>
<b>LIST OF TABLES</b>	<b>xi</b>
<b>LIST OF ALGORITHMS</b>	<b>xii</b>
<b>LIST OF ABBREVIATIONS</b>	<b>xiii</b>
<b>ABSTRACT</b>	<b>xv</b>
<b>CHAPTER ONE: INTRODUCTION</b>	<b>1</b>
1.1 Background	1
1.2 Statement of the Problem	6
1.3 Aim and Objectives	7
1.4 Justification of the Study	7
1.5 Scope and Limitation of the Study	7
<b>CHAPTER TWO: LITERATURE REVIEW</b>	<b>8</b>
2.1 Preamble	8
2.2 Theoretical Background	8
2.2.1 Information Security	8
2.2.2 Malware Detection	9
2.2.3 Cloud Computing Infrastructure	16
2.3 Need for Artificial Intelligence	21
2.3.1 Machine Learning	22
2.3.2 Deep Learning	26
2.3.3 Federated Learning	27
2.4 Related Works	40
2.4.1 Review of Malware Detection	40
2.4.2 Gaps in the Existing Literature	42
2.5 Chapter Summary	45
<b>CHAPTER THREE: MATERIALS AND METHODS</b>	<b>46</b>
3.1 Preamble	46
3.2 Proposed Research Conceptual Framework	46



3.2.1 Problem Formulation	47
3.2.2 Data Acquisition	47
3.2.3 Data Pre-processing and Feature Selection	50
3.2.4 Model Development	52
3.2.5 Development Environment	61
3.3 Chapter Summary	61
<b>CHAPTER FOUR: RESULTS AND DISCUSSION</b>	<b>62</b>
4.1 Preamble	62
4.2 Model Development Results	62
4.2.1 Data Pre-processing Outputs	62
4.2.2 Model Training Results	63
4.3 Discussion	78
4.4 Chapter Summary	80
<b>CHAPTER FIVE: CONCLUSION AND RECOMMENDATIONS</b>	<b>81</b>
5.1 Preamble	81
5.2 Conclusion	81
5.3 Research Contribution	81
5.4 Recommendations	82
<b>REFERENCES</b>	<b>83</b>
<b>APPENDIX</b>	<b>92</b>

## LIST OF FIGURES

<b>FIGURES</b>	<b>TITLE OF FIGURES</b>	<b>PAGES</b>
Figure 1.1:	(a) Cloud Characteristics, (b) Cloud Service Models	3
Figure 2.1:	Cloud Computing Models	18
Figure 2.2:	General Federated Cloud Architecture	21
Figure 2.3:	Position of ML and DL in AI	22
Figure 2.4:	ML Taxonomy	24
Figure 2.5:	Structure of an ANN	26
Figure 2.6:	DL Taxonomy	27
Figure 2.7:	Taxonomy of Federated Learning System	29
Figure 2.8:	Types of Federated Learning based on Data partition	30
Figure 2.9:	MLP Architecture	31
Figure 2.10:	FFNN architecture	32
Figure 2.11:	LSTM architecture	33
Figure 2.12:	Types of Federated Learning based on the communication architecture	36
Figure 3.1:	Block Diagram of the FL-based Malware Detection	47
Figure 3.2:	Dataset Overview	50
Figure 3.3:	Classes in the dataset	52
Figure 3.4:	Number of Samples Per Class	53
Figure 3.5:	Flow Chart of the FL Environment	55
Figure 3.6:	MLP Network Architecture	58
Figure 3.7:	FFNN Network Architecture	59
Figure 3.8:	LSTM Network Architecture	59
Figure 4.1:	Dataset Information	62
Figure 4.2:	Features Normalised	63
Figure 4.3:	Classes One Hot Encoding	63
Figure 4.4:	Flower Framework Output	63
Figure 4.5:	Connection of the Client to the Server	64
Figure 4.6:	MLP Accuracy	66
Figure 4.7:	MLP Loss	67
Figure 4.8:	MLP-Confusion Matrix	69
Figure 4.9:	FFNN Accuracy	71
Figure 4.10:	FFNN Loss	72
Figure 4.11:	FFNN-Confusion Matrix	74
Figure 4.12:	LSTM Accuracy	76
Figure 4.13:	LSTM Loss	76
Figure 4.14:	LSTM-Confusion Matrix	78

## LIST OF TABLES

<b>TABLES</b>	<b>TITLE OF TABLES</b>	<b>PAGES</b>
	Table 2.1: Comparison Between Static and Dynamic Analysis	15
	Table 2.2: Gaps in the Existing Literature	42
	Table 3.1: Mapping of Objectives with Materials and Methods	46
	Table 3.2: Total Count of Malware Family in the Dataset	48
	Table 3.3: Feature Description	49
	Table 4.1: MLP Performance	65
	Table 4.2: MLP-Classification Report (4 clients)	68
	Table 4.3: MLP-Classification Report (8 clients)	68
	Table 4.4: FFNN Performance	70
	Table 4.5: FFNN-Classification Report (4 clients)	72
	Table 4.6: FFNN-Classification Report (8 clients)	73
	Table 4.7: LSTM Performance	74
	Table 4.8: LSTM-Classification Report (4 clients)	77
	Table 4.9: LSTM-Classification Report (8 clients)	77
	Table 4.10: Comparison with Other Works	80
	Table A.1: Other Performance Metrics of the MLP Model	92
	Table A.2: Other Performance Metrics of the FFNN Model	93
	Table A.3: Other Performance Metrics of the LSTM Model	94

## LIST OF ALGORITHMS

Algorithm 2.1: Federated Averaging	37
Algorithm 2.2: Federated Stochastic Variance Reduced Gradient	38
Algorithm 2.3: Federated Matched Averaging	39

## LIST OF ABBREVIATIONS

AI	Artificial Intelligence
ANN	Artificial Neural Networks
CApIC-ACE	Covenant Applied Informatics and Communication Africa Centre of Excellence
CC	Cloud Computing
CCMP	Cloud Computing Management Platforms
CSC	Cloud Service Consumer
CSP	Cloud Service Provider
CSU	Cloud Service User
CSV	Comma Separated Value
DL	Deep Learning
FedAvg	Federated Averaging
FEDGEN	Federated Genomic Cloud
FFNN	Feedforward Neural Network
FL	Federated Learning
FLOWER	Friendly Federated Learning Research Framework
FLS	Federated Learning System
FN	False Negative
FP	False Positive
HFL	Horizontal Federated Learning
IaaS	Infrastructure as a Service
IDS	Intrusion Detection System
LSTM	Long Short-Term Memory
MDS	Malware Detection System
ML	Machine Learning

MLP	Multi-Layer Perceptron
PaaS	Platform as a Service
PCA	Principal Component Analysis
ReLU	Rectified Linear Unit
SaaS	Software as a Service
SMOTE	Synthetic Minority Oversampling Technique
TN	True Negative
TP	True Positive
VFL	Vertical Federated Learning
VM	Virtual Machine

## ABSTRACT

Due to the large number of heterogeneous applications using the same infrastructure, enforcing security and reliability in the cloud is a difficult but crucial task. A security analysis system that detects threats for example malicious software (malware) should exist within the cloud infrastructure. Different malware techniques that bypass network-based and host-based security protections have led to the development of new methods for analysing and detecting malware, which have evolved over the past decades. Due to the complexity of learning the ever-changing configurations of malware, it is challenging for forensics investigators to keep up with the exponential rise in the number and variety of malware species. In this research work, a malware detection model was developed for interconnected cloud infrastructures based on federated learning. This technique enables collaboration between multiple devices in the training of machine learning models without exchanging data, thereby preserving the privacy of individual users. Three different deep-learning algorithms were selected and used in the training, validation, and testing of the models. By the model training with eight clients and twenty-five federation rounds, the FeedForward Neural Networks(FFNN) model provided the best performance with a precision of 84%, an F1-score of 84%, and an accuracy of 84% whereas the Multi-Layer Perceptron(MLP) model provided 83% of precision, 83% of F1-score, and 83% of accuracy and the Long Short-Term Memory(LSTM) model provided a performance with 80% of precision, 80% of F1-score, and 80% of accuracy as well.

***Keywords: Federated learning, malware detection, federated cloud, machine learning***