MAP REDUCE SECURITY MODEL FOR ASTHMA PREDICTION IN CHILDREN USING FEDERATED XGBOOST

EKPO, RAPHAEL HENSHAW
(CUGP100277)

APRIL, 2024

MAP REDUCE SECURITY MODEL FOR ASTHMA PREDICTION IN CHILDREN USING FEDERATED XGBOOST

BY

EKPO, RAPHAEL HENSHAW
(CUGP100277)
B.Sc Computer Science, University of Uyo, Uyo
M.Sc Computer Science, Covenant University, Ota

A THESIS SUBMITTED TO THE SCHOOL OF POSTGRADUATE STUDIES IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE AWARD OF DOCTOR OF PHILOSOPHY (Ph.D) DEGREE IN COMPUTER SCIENCE, DEPARTMENT OF COMPUTER AND INFORMATION SCIENCES, COLLEGE OF SCIENCE AND TECHNOLOGY, COVENANT UNIVERSITY, OTA, OGUN STATE, NIGERIA

APRIL, 2024

ACCEPTANCE

This is to attest that this thesis is accepted in partial fulfillment of the requirements for the award of the degree of Doctor of Philosophy in Computer Science in the Department of Computer and Information Sciences, College of Science and Technology, 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, **EKPO, RAPHAEL HENSHAW (CUGP100277)**, declare that this research work was carried out by me under the supervision of Prof. Victor C. Osamor of the Department of Computer Science and Prof. Ambrose A. Azeta of the Department of Computer and Information Sciences, Covenant University, Ota, Nigeria. I attest that this thesis has not been presented either wholly or partly for the award of any degree elsewhere. All the sources of data and scholarly information used in this thesis are duly acknowledged.

EKPO, RAPHAEL HENSHAW

Signature and Date

CERTIFICATION

This is to certify that the research work titled "**MAP REDUCE SECURITY MODEL FOR ASTHMA PREDICTION IN CHILDREN USING FEDERATED XGBOOST"** is an original research work carried out by **EKPO, RAPHAEL HENSHAW (CUGP100277)** in the Department of Computer and Information Sciences, College of Science and Technology, Covenant University, Ota, Ogun State, Nigeria, under the supervision of Prof. Victor C. Osamor and Prof. Ambrose A. Azeta. We have examined and found this work acceptable as part of the requirements for the award of a degree of Doctor of Philosophy in Computer Science.

Prof. Victor C. Osamor
(Supervisor) **Signature and Date**

Prof. Ambrose A. Azeta
(Co-Supervisor) **Signature and Date**

Prof. Olufunke O. Oladipupo
(Head of Department) **Signature and Date**

Prof. Adebukola S. Onashoga
(External Examiner) **Signature and Date**

Prof. Akan B. Williams
(Dean, School of Post-Graduate Studies) **Signature and Date**

DEDICATION

This thesis is dedicated to God almighty for His strength, and help throughout this study up to its successful completion.

ACKNOWLEDGEMENTS

I want to acknowledge the Almighty God, my strength, my protector, my helper, my provider and way maker who made this Ph.D a reality. Despite all the challenges, He gave me victory at last. To Him be all the glory, honour and praises.

I sincerely appreciate the presiding Bishop of Living Faith Church Worldwide and Chancellor of Covenant University, Dr. David Oyedepo. You are a blessing to our generation. I also express my profound gratitude to the Vice-Chancellor, Prof. Abiodun H. Adebayo, the Registrar, Mrs Regina A. Tobi-David, the Dean, School of Postgraduate Studies, Prof. Akan B. Williams and Dean, College of Science and Technology, Prof. Timothy Anake for their leadership role and support in this great citadel of learning.

I also wish to appreciate the effort of my supervisors, Prof. Victor C. Osamor and Prof. Ambrose A. Azeta, for their guidance, patience, and support throughout this PhD programme. I appreciate the head of the department, Prof. Olufunke O. Oladipupo. You have been a mother. My gratitude goes to the Departmental Post Graduate Coordinator, Dr Mrs. Itunu Isewon for her constant support. My special regards to my colleagues and all members of staff of the college of Natural and Applied Sciences, Crawford University. Thank you for standing by me.

I like to acknowledge, the Head of Department of Computer and Mathematical Sciences, Crawford University, Dr. Tayo Adefokun, for his understanding and encouragement. My heartfelt gratitude goes to my wife, Temitope, Raphael Henshaw, who has always been there for me, supported and encouraged me throughout the research work. My children Abundance Henshaw, Enoch Henshaw, and Asher Henshaw, who endured my absence for many days during this research work, you are the best. I am thankful to my brothers and sister in Christ – Dare Runsewe, Yusuf Kolawale and Yemisi Olabisi – for their constant prayers and support towards the success of this programme. I cannot forget the likes of Blessing Odede, Dr Adewole Adewummi, amongst others. These are the people who sacrificed their time to make this program possible. Finally, I am extremely grateful to my spiritual fathers, Pastor Uchechukwu Osinobi, and Pastor Matthew Ashimolowo for your prayers and teachings that were helpful to me in completing this programme.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ACC	Accuracy
ADB	Adaboost
ANN	Artificial Neural Network
AUC	Area Under the Receiver Operating Curve
BP	Back Propagation
BPANN	Back Propagation Neural Network
CAPE	Children Asthma Prediction in Early Life
CAPP	Children Asthma Prediction at Preschool
CPDS	Cluster Primal Dual Splitting
CPX	Cox Proportional Hazard
CSAMM	Context sensitive Auto Association Memory Neural Network Model
CVAE	Conditional Variational Auto Encoder
DNN	Deep Neural Network
DT	Decision Tree
HER	Electronic Health Record
FCB	Fast Correlation- Based Filter
FENO	Fractional Exhaled Nitric Oxide
FEV1	Forced Exploratory Volume in one seconds
GA	Genetic Algorithm
GBDT	Gradient Boosting Decision Tree
GBM	Gradient Boost Machine
GOOGLE COLAB	Google Collaboratory
HRS	Hospital Record System
HSCRP	High Sensitivity Creative Protein
IDE	Integrated Development Environment
KNN	K- Nearest Neighbour
LASSO	Least Absolute Shrinkage and Selection Logistic Regression
LOAC	Level of Asthma Control
LR	Linear Regression
LR	Logistic Regression
LSSVM	Least Square Support Vector Machine
LSTM	Long Short erm Memory

MCC:	Matthews Correlation Coefficient
MDBS:	Multi Boost with Decision Support
MEF50:	Maximum Exploratory Flow of 50% of the vital flow capacity
MLP:	Multi Layer Perceptron
MLPNN:	Multi Layer Perceptron Neural Network
MNL:	Multinomial Logistic Regression
NB:	Naïve Bayes
NSCH:	National Survey of Children Health
PR:	Poission Regression
PRE:	Precision
PSO:	Particle Swarm Optimization
RBF:	Radial Basis Function
RCC:	Recall
RDSVM:	Radial Bias kernel function Support Vector Machine
RF:	Random Forest
RFE:	Recursive Feature Elimination
SEN:	Sensitivity
SME:	Security of Multiparty Communication
SPE:	Specificity
SSVM:	Sparse Support Vector Machine
SVM:	Support Vector Machine
TEE:	Trusted Executive Environment
XGB:	Extreme Gradient boosting

ABSTRACT

A substantial number of asthma development prediction models in children, such as conventional methods involving risk factors, logistic regression, the hybrid of statistical methods, and machine learning based approaches, exist. The problem associated with conventional methods of asthma prediction in children is low predictive accuracy of the model. However, using centralised machine learning approaches in healthcare requires training the learning models on large datasets. Besides cost, data privacy and security represent the main problems with centralised machine learning. The objective of this study is to develop a Map Reduce Security for asthma prediction in children using a federated XGBoost as a response to the aforementioned limitations associated with the existing asthma prediction model for children. This study leveraged two diverse datasets: the Nigerian Hospital Asthma dataset and the National Survey of Children's Health dataset for benchmarking purposes. After preprocessing the dataset, the symmetrical uncertainty and normalization interaction gain, as well as the undersampling approach, were employed for feature selection and dataset balancing. The system was trained, and tested using Federated Artificial Intelligence (AI) Technology Enabler (FATE) and extended to XGBoost model with one central server for federated algorithm averaging. The map reduce security measure was employed for the input data during training to avoid data leakage. The study was implemented with the Python programming language on Google Collaboratory (Collab) environment. The results of the analysis showed considerable high accuracy of 0.98, precision (0.98%), recall (0.98%), and F1-score (0.99%) for the asthmatic class and precision of 0.98%, as well as F1-score of 0.99% for the non-asthmatic class. The implemented model was benchmarked with the existing study on asthma prediction model in children using the same NSCH dataset of 23 features and 50212 samples. This study achieved a prediction accuracy of 93.8%, outperforming the existing models. The simulated attack on the implemented model results show that the model correctly identified whether a data point was used in the training process with predictive accuracy of 97.94%, membership inference attack, and data loss of 0.014%, respectively. In conclusion, the study finds that the model developed by the researchers performed better in predicting childhood asthma than some of the state-of-the-art machine learning algorithms.

Keywords: Asthma prediction model, Asthma in Children, Encryption, Federated Learning, Performance Metrics and Machine Learning.