

**GENAPP: A WEB APPLICATION FOR PREDICTING  
PLASMODIUM FALCIPARUM RESISTANCE TO SELECTED  
ANTIMALARIA DRUGS**

**AKINWALE, MERCY OJOCHENWUMI  
(22PBF02392)  
B.Sc Physiology, University of Ibadan, Ibadan**

**AUGUST, 2024**

**GENAPP: A WEB APPLICATION FOR PREDICTING  
PLASMODIUM FALCIPARUM RESISTANCE TO SELECTED  
ANTIMALARIA DRUGS**

**BY**

**AKINWALE, MERCY OJOCHENWUMI  
(22PBF02392)  
B.Sc Physiology, University of Ibadan, Ibadan**

**A DISSERTATION SUBMITTED TO THE SCHOOL OF  
POSTGRADUATE STUDIES IN PARTIAL FULFILMENT OF THE  
REQUIREMENTS FOR THE AWARD OF THE MASTER OF  
SCIENCE (M.Sc) DEGREE IN BIOINFORMATICS IN THE  
DEPARTMENT OF COMPUTER AND INFORMATION SCIENCES,  
COLLEGE OF SCIENCE AND TECHNOLOGY, COVENANT  
UNIVERSITY, OTA, OGUN STATE, NIGERIA**

**AUGUST, 2024**

## **ACCEPTANCE**

This is to attest that this dissertation is accepted in partial fulfilment of the requirements for the award of the degree of Master of Sciences in Bioinformatics in the Department of Computer and Information Sciences, College of Science and Technology, Covenant University, Ota, 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, AKINWALE MERCY OJOCHENWUMI (22PBF02392)**, declare conducted this research entitled “**GENAPP: A WEB APPLICATION FOR PREDICTING *PLASMODIUM FACLIPARUM* RESISTANCE TO SELECTED ANTIMALARIA DRUGS**”. It was carried out under the supervision of Prof. Jelili O. Oyelade. Concepts of this research project are the results of the research carried out by Akinwale Mercy Ojochenwumi, and other researchers' ideas have been fully recognized.

**AKINWALE, MERCY OJOCHENWUMI**

**Signature and Date**

## **DEDICATION**

I dedicate this project to God Almighty for His grace, wisdom and knowledge given to me throughout my Master's Degree Programme. Also, I thank my family and friends for their unending love and support.

## **CERTIFICATION**

This is to certify that this dissertation titled “**A WEB APPLICATION FOR PREDICTING PLASMODIUM FACLIPARUM RESISTANCE TO SELECTED ANTIMALARIA DRUGS**” is original research carried out by **AKINWALE, MERCY OJOCHENWUMI (2PBF02392)** in the Department of Computer and Information Sciences, College of Science and Technology, Covenant University, Ota, Ogun State, Nigeria under the supervision of Prof. Jelili O. Oyelade. We have examined and found this work acceptable as part of the requirements for the award of Master of Science (M.Sc.) in Computer Science.

**Prof. Jelili O. Oyelade**  
(Supervisor)

**Signature and Date**

**Prof. Olufunke Oladipupo**  
(Head of Department)

**Signature and Date**

**Prof. Zacchaeus O. Omogbadegun**  
(External Examiner)

**Signature and Date**

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

**Signature and Date**

## **ACKNOWLEDGEMENTS**

I give all thanks to the almighty God for the supply of grace, strength, wisdom and helps through different stages of the project. I want to express my heartfelt gratitude to my supervisor Prof. Jelili O. Oyelade, my research supervisor, for his guidance and support through the project, thank you for being a father. Thank you to the PG coordinator Dr. Itunuola Isewon for your selfless commitment in giving invaluable guidance during the research, grateful for the investment of your time, resource and every input you had in this project. I thank the management of Covenant University, Capic-ACE for the funding and support for this work. A profound thank you to the HOD, Professor Oladipupo Grace and other faculty of computer and information Science department for their academic supports and contributions.

I am incredibly grateful to my lovely husband – Mr Zacchaeus Akinwale for his love, understanding, unwavering support, prayers and the sacrifices he had to make through the course of my studies. I deeply appreciate my parents Pastor and Pastor Mrs Akowe and my siblings- Salem Akowe and Prince Akowe for their love, prayers, support and encouragements through this journey. Thank you to my father, mother, sisters and brothers In-laws for their support and understanding through this phase of my life.

Finally, I want to extend my gratitude to my friends, Racheal O, Chidimma Moghalu, senior colleagues Dr Emmanuel Jerry Dauda and Toba Emmanuel, and co-research assistants -Ashuza Kuderha Rifin, Onyemeachi Promise, Ndifon Naomi for their support and helps at different stages of the project. I am deeply grateful for the collective contributions of everyone who has helped me complete this research work, whether directly or indirectly. God bless you all.

# TABLE OF CONTENTS

CONTENTS	PAGES
<b>TITLE PAGE</b>	<b>i</b>
<b>COVER PAGE</b>	<b>ii</b>
<b>ACCEPTANCE</b>	<b>iii</b>
<b>DECLARATION</b>	<b>iv</b>
<b>DEDICATION</b>	<b>v</b>
<b>CERTIFICATION</b>	<b>vi</b>
<b>ACKNOWLEDGEMENT</b>	<b>vii</b>
<b>TABLE OF CONTENT</b>	<b>viii</b>
<b>LIST OF FIGURES</b>	<b>xi</b>
<b>LIST OF TABLES</b>	<b>xii</b>
<b>LIST OF ABBREVIATIONS</b>	<b>xiii</b>
<b>ABSTRACT</b>	<b>xiv</b>
<b>CHAPTER ONE</b>	<b>1</b>
1.1 Background Of Study	1
1.2 Statement of Research Problem	3
1.3 Aim and Objectives of the Study	4
1.4 Research Methodology	4
1.5 Significance of the Study	5
1.6 Scope of the study	5
1.7 Contribution to Knowledge.	6
1.8 Organization of the dissertation	6
<b>CHAPTER TWO</b>	<b>6</b>
2.1 Introduction	7
2.2 Antimalaria Drugs and development of drug resistance	7
2.3 Parasite lifecycle	8
2.3.1 The liver stage	9
2.3.2 Erythrocytic (Blood) stage	9
2.3.3 Mosquito Sexual Stage	10
2.3.4 Plasmodium Drug Resistance	10
2.4 Antimalaria Drugs and their Mechanisms of Action	11
2.5 Antimalaria drug Resistance Mechanism in <i>P.falciparum</i>	20
2.6 Artemisinin-based combination therapies drugs	21
2.7 Genomic Approaches to Drug Resistance in Malaria	22
2.7.1 Genome-wide association studies (GWAS)	23



2.7.2	Genetic crosses and linkage analyses	23
2.7.3	In vitro drug selection and whole-genome sequencing	23
2.8	Machine learning	24
2.8.1	Ensemble learning algorithms	25
2.8.2	Random Forest	25
2.8.3	Gradient Boost	25
2.8.4	Deep Learning	26
2.8.5	Support Vector Machine (SVM)	26
2.9	Related Works	27
<b>CHAPTER THREE</b>		<b>30</b>
3.1	Introduction	30
3.2	Work Flow	30
3.3	Data	31
3.3.1	Data Pre-processing	31
3.3.2	Quality control and Assessment	31
3.3.3	SNIPPY: Variant calling	32
3.3.4	Genome Assembly	32
3.4	Data Encoding and class label curation	33
3.5	Model Selection	34
3.6	Methods	34
3.6.1	Ensemble Machine Learning Methods	34
3.6.2	Bagging Approach	35
3.6.3	Boosting Approach	35
3.6.4	Random Forest Models	36
3.6.5	Gradient Boosting Machine	37
3.7	Support vector machine (SVM)	38
3.8	Cross-Validation	38
3.9	Variable Importance	39
3.10	Performance Evaluation	39
3.11	Deployment of the classifiers on a Web Application	40
3.12	Conclusion	41
<b>CHAPTER FOUR</b>		<b>41</b>
4.1	Data Exploration and Preprocessing	41
4.2	Quality Control Results	43

4.3.1.	Chloroquine Resistance	46
4.3.2.	Halofantrine Resistance	47
4.3.3.	Dihydroartemisin Resistance	48
4.3.4.	Lumefantrine Resistance	49
4.3.5.	Quinine Resistance	49
4.3.6.	Piperaquine (PIQ)	50
4.4.	Variable Importance	51
4.4.1.	Important Features for Chloroquine Resistance	51
4.4.2.	Important Features for DihydroArtemisinin (DHA) Resistance	52
4.3.2.	Important Features for Lumefantrine Resistance	53
4.3.3.	Important Features for Halofantrine Resistance	53
4.2.4	Important Features for Quinine Resistance	61
4.2.5	Important Features for Piperaquine Resistance	62
4.3	Web Application deployment	63
4.4	Benchmarking of GenApp best performing Algorithm with existing works	64
<b>CHAPTER FIVE</b>		<b>66</b>
5.1.	Summary	66
5.2.	Conclusion	67
5.3.	Contribution to Knowledge	66
5.4.	Limitation of study	67
5.5.	Recommendation	67
<b>REFERENCES</b>		<b>68</b>

## LIST OF FIGURES

FIGURES	TITLE OF FIGURE	PAGES
Figure 2. 1:	The life cycle of malaria and available treatment	8
Figure 2. 2:	Antimalaria resistance mechanism of P.falciparum to some antimalaria drug	21
Figure 2. 3:	Deep learning architecture	26
Figure 3. 1:	Workflow diagram	30
Figure 3. 2:	Illustration of (A)Bagging and (B) Boosting ensemble algorithms	36
Figure 3. 3:	Schematic diagram of a Random Forest Algorithm	37
Figure 4. 2:	Bar plots of Phenotype expressions of the selected drugs	42
Figure 4. 3:	Base sequence quality report visualization before trimming and cleanup	43
Figure 4. 4:	Base sequence quality report visualization before trimming and cleanup	44
Figure 4. 5:	Base sequence quality report visualization before trimming and cleanup	45
Figure 4. 6:	Mean square quality	46
Figure 4. 7:	Performance of the selected algorithm on Plasmodium falciparum resistance to Chloroquine	50
Figure 4. 8:	Performance of the selected algorithm on Plasmodium falciparum resistance to Halofantrine.	50
Figure 4. 9:	Bar charts of performance metrics scores of models	51
Figure 4. 10:	Performance of the selected algorithm on Plasmodium falciparum resistance to Lumefrantine	52
Figure 4. 11:	Performance of the selected algorithm on Plasmodium falciparum resistance to Quinine	53
Figure 4. 12:	Performance of the selected algorithm on Plasmodium falciparum resistance to Piperaquine	54
Figure 4. 13:	Identified coding gene regions on Chromosome 4	65
Figure 4. 14:	Novel gene identified with yet unknown function on Chromosome 11	65
Figure 4. 15:	Novel gene identified with yet unknown function on Chromosome 11	65
Figure 4. 16:	Identified genes with unknown function Associated with resistance	63
Figure 4. 17:	Frontend view of GenApp	67

## LIST OF TABLES

<b>TABLES</b>	<b>TITLE OF TABLE</b>	<b>PAGES</b>
Table 2. 1:	antimalarial drugs used as monotherapy	14
Table 2. 2:	artemisinin-based combination treatments (acts)	22
Table 3. 1:	organism information	32
Table 3. 2:	resistance threshold used for conversion of continuous to binary resistance phenotype	35
Table 3. 3:	selected parameters for the gbm model	39
Table 4. 1:	summary table of identified novel genes driving antimalaria drugs resistance	56
Table 4. 2:	summary table of benchmarked results	68

## ABBREVIATIONS

ABBREVIATION	MEANING
ACT	Artemisinin-based Combination Therapy
AMR	Antimalaria Resistance
CNV	Copy-Number Variation
CQ	Chloroquine
DHA	Dihydroartemisinin
LUM	Lumefantrine
PIQ	Piperaquine
HLF	Halofantrine
IC50	50% Inhibitory Concentration
GWAS	Genome- Wide Association Study
ENA	European Nucleotides Archive
DNA	Deoxyribonucleic Acid
ML	Machine Learning
NCBI	National Center for Biotechnology Information
EMBL-EBI	European Molecular Biology Laboratory- European Bioinformatics Institute
PFCRT	Plasmodium falciparum Chloroquine Resistance
WHO	World Health Organization
PFK13	Plasmodium falciparum K13-propeller domain.

## ABSTRACT

Antimalarial drug resistance poses a significant challenge to global malaria control efforts, particularly in regions burdened by *Plasmodium falciparum*, the deadliest malaria parasite. The development and spread of resistance to widely used antimalarial drugs, such as chloroquine, Lumefantrine, Halofantrine, Quinine, Piperaquine and Dihydroartemisinin, have greatly impacted treatment efficacy and disease outcomes. This resistance is driven by various genetic mutations in *P. falciparum*, which confer the ability to survive drug exposure. This study explores the prediction of antimalarial drug resistance using machine learning algorithms Random Forest, Gradient Boosting Machine (GBM), and Support Vector Machine (SVM). Focusing on six key antimalarial drugs Chloroquine, Dihydroartemisinin, Lumefantrine, Quinine, Halofantrine, and Piperaquine the research aims to identify genetic markers that contribute to resistance and develop predictive models to enhance treatment strategies. To avoid model overfitting, 5-fold cross-validation was conducted on the training set to choose the optimal hyperparameter values. Regardless of the resistance mechanism, whether acquired resistance or point mutations in the chromosome, the accuracy (mean cross-validation score) of Random Forest had an average of 83% across all drugs. The model significantly classified the resistant isolates from the sensitive isolates of the parasite and could be used as potential tools in antimalarial resistance surveillance and clinical studies. A number of genes associated with antimalaria drug resistance were identified. Novel genes and loci were also discovered, of interest are genes on chromosomes 1, 4, 7, 8, 9, 10, 11, 17 and 19.

***Keywords: Machine learning, Antimalarial drug resistance, Plasmodium falciparum, genomic studies, phenotype prediction, malaria eradication***