

**Performance Evaluation of Some Selected Learning Algorithms for Software Defect  
Prediction Using Harmony Search Algorithm**

**Folorunsho BAMISAYE  
LCU/PG/002134**

**Being a MSc Thesis Submitted to the Department of Computer Science, Faculty of  
Natural and Applied Science, Lead City University, Ibadan, Oyo State, Nigeria**

**In Partial Fulfillment of the Requirements for the Award of Master of Science Degree  
(MSc) in Computer Science**

**2022**

## Certification

This is to certify that Folorunsho Bamisaye with matriculation number PG/LCU/002134 carried out this research work titled “Performance Evaluation of Some Selected Learning Algorithms for Software Defect Prediction Using Harmony Search Algorithm”, in the Department of Computer Science, Faculty of Natural and Applied Sciences, Lead City University, Ibadan, Oyo State, for the award of Master of Science degree (MSc) in Computer Science and that this has not been previously submitted.

---

**Mrs. V. B. Oyekunle**  
(Supervisor)

---

Date

---

**Dr. Wilson Sakpere**  
(Head of Department)

---

Date

DO NOT COPY. LEAD CITY UNIVERSITY, NIGERIA

## **Dedication**

This project work is dedicated to God Almighty.

DO NOT COPY. LEAD CITY UNIVERSITY, NIGERIA

## **Acknowledgment**

I am very grateful to this great institution for creating enabling environment for research work during my academic year.

My appreciation goes to my indefatigable supervisor, Mrs. V. B. Oyekunle for always creating time out of her busy schedule to guide, advise, teach, and support me to accomplish this research work. I am grateful to our able PG Coordinator, Dr. A. A. Waheed for his support, guide and commitment to the success of this great work. I am grateful to my Head of Department Dr. Wilson Sakpere for his leadership role.

My gratitude goes to my beautiful wife (Funmilayo) and children (OreofeJesupo and AnuJesupo) for their understanding, prayers, and support while leaving them alone for days for research work until the completion of this project. I am also very grateful to my parents, elder and younger ones for their prayers and support.

“Though the above-mentioned institutions and persons have assisted in the process of this research work, I alone stand responsible for the errors, if any, found in the work.”

## Abstract

Software defects (SD) is a significant area of software development that has called for the attention of both academics and professionals in the last few decades. The more software evolves, the more there is a need to produce testing measures by which the reliability, dependability, and efficiency of the software can be verified and ascertained. The quest to predict defects or better still produce error-free software is commendable, however, other factors such as an accurate prediction rate and time interval for prediction call for imperative attention. The prolonged processing of prediction can lead to misclassification, most especially when a large dataset is used for classification. In other words, misclassification and prolonged processing is inevitable when a large dataset is used for prediction.

That is why this study has applied a meta-heuristic optimization algorithm for feature selection (FS) to reduce high dimensionality which may lead to prolonged classification time and misclassification. The HSA was employed for feature selection, and five learning algorithms- Support Vector Machine (SVM), Artificial Neural Network, K-Nearest Neighbour, Naïve Baye, and C4.5 were applied for classification. Also, Eclipse dataset was used for prediction. The operational output of the model was achieved using some evaluation metrics such as precision, accuracy, recall, classification time, and F1 score.

The recorded results with HSA revealed that the ANN algorithm achieved the lowest classification time of 24.09s in the Eclipse Dataset which shows that the predictive rate of ANN outperformed other classifiers used for defect classification. The highest accuracy of 86.44% was obtained in SVM which shows that KNN outperformed other learning algorithms used for prediction in terms of correctness.

Based on the output of KNN with harmony search in the SDP, it is therefore recommended for model development.

**Keywords:** Software Defeat, Defect Prediction, Dimensionality Reduction, Feature Selection, Misclassification

**Word Count: 279**

## Table of Contents

	Page
Title Page	
Certification	ii
Dedication	iii
Acknowledgment	iv
Abstract	v
Table of Contents	vi
List of Tables	x
List of Figures	xii
List of Algorithms	xiv
List of Acronyms	xv
<b>Chapter One: Introduction</b>	
1.1 Background to the Study	1
1.2 Statement of the Problem	4
1.3 Aim and Objectives of the Study	5
1.4 Research Questions	6
1.5 Significance of the Study	7
1.6 Scope of the Study	7
1.7 Limitations of the Study	7
1.8 Operational Definition of Terms	8
1.9 Endnotes	10
<b>Chapter Two: Literature Review</b>	
2.1 Conceptual Review	12
2.2 Other Method of Software Defect Classification	14
2.2.1 Defect Prediction by Code Metrics	14
2.2.2 Some Selected Metaheuristic Algorithms	21

2.2.2.1 Deep Learning Algorithm	22
2.2.2.2 Particle Swarm Optimization (PSO)	24
2.2.2.3 Ant Colony Optimization (ACO)	29
2.2.2.4 Artificial Bee Colony Optimization (ABCO)	31
2.2.2.5 Firefly Optimization Algorithm (FOA)	37
2.2.2.6 Harmony Search Algorithm (HSA)	38
2.2.3 Example of Feature Extraction Algorithm	41
2.2.3.1 Principal Component Analysis	42
2.2.3.2 Principal Component Analysis for Feature Extraction	45
2.2.4 Dimensionality Reduction	46
2.2.5 Overview of Machine Learning Algorithms applied	48
2.2.5.1 Artificial Neural Network (ANN)	48
2.2.5.2 Support Vector Machine (SVM)	49
2.2.5.3 Naïve Bayes (NB)	50
2.2.5.4 K-Nearest Neighbour (KNN)	51
2.2.5.5 C4.5	52
2.2.6 Features	53
2.2.3.1 Process Metrics	54
2.2.3.2 Code Metrics	55
2.2.6 Feature Selection	56
2.2.7 Software Defect Management	64
2.3 Review of Empirical Studies	72
2.4 Conceptual Framework	76
2.5 Summary of Literature Reviewed	76
2.6 Endnotes	77

## **Chapter Three: Methodology**

3.1	Research Approach	82
3.2	System Design	83
3.2.1	Data Preprocessing	84
3.2.2	Datasets Overview	85
3.3	Requirements Specification	86
3.4	Dataset Acquisition	86
3.5	Evaluation Metrics	87
3.5.1	Accuracy	88
3.5.2	Precision	88
3.5.3	Recall	89
3.5.4	F1-Score	89
3.5.4	Classification Time	90
3.6	Endnotes	91

## **Chapter Four: Results and Discussion of Findings**

4.1	Results	92
4.2	Discussion of Findings	105
4.3	Comparative Analysis Results	107
4.4	Endnotes	114

## **Chapter Five: Conclusion**

5.1	Summary of Findings	115
5.2	Conclusion	116
5.3	Recommendation	117
5.4	Contribution to Knowledge	117
5.5	Suggestion for Further Studies	118

Bibliography	119
Conference Proceedings	119
E-Books	119
Journals	119
Magazines	122
Appendix	123
Bio-data	144
University Compliance Form	154

DO NOT COPY. LEAD CITY UNIVERSITY, NIGERIA

## List of Table

<b>Title</b>	<b>Page</b>
2.0: Research Result	73
3.1: Information on Datasets	85
4.1: The Result of Performance Metrics for Eclipse Dataset using ANN for Prediction (without HSA)	93
4.2: The Result of Performance Metrics for Eclipse Dataset using ANN for Prediction (with HSA)	94
4.3: The Result of Performance Metrics for Eclipse Dataset using C4.5 for Prediction (without HSA)	95
4.4: The Result of Performance Metrics for Eclipse Dataset using C4.5 for Prediction (with HSA)	96
4.5: The Result of Performance Metrics for Eclipse Dataset using KNN for Prediction (without HSA)	97
4.6: The Result of Performance Metrics for Eclipse Dataset using KNN for Prediction (with HSA)	98
4.7: The Result of Performance Metrics for Eclipse Dataset using NB for Prediction (without HSA)	99
4.8: The Result of Performance Metrics for Eclipse Dataset using NB for Prediction (with HSA)	100
4.9: The Result of Performance Metrics for Eclipse Dataset using SVM for Prediction (without HSA)	101
4.10: The Result of Performance Metrics for Eclipse Dataset using SVM for Prediction (without HSA)	102
4.11: The Summary of the Result when Relevant Features were selected for Prediction	103

DO NOT COPY. LEAD CITY UNIVERSITY, NIGERIA

## List of Figures

Figure	Title	Page
2.1:	Filter Method of Software Defect Management	23
2.2:	Ant Bee Colony Optimization	33
2.3:	Artificial Neural Network	49
2.4:	Support Vector Machine	50
2.5:	Naïve Bayes	51
2.6:	K-Nearest Neighbour	52
2.6:	C4.5	53
2.8:	Software Defect Management	56
2.9:	Filter Method of Software Defect Management	56
2.10:	Wrapper Method of Software Defect Management	59
2.11:	Intrinsic Method of Software Defect Management	60
2.12:	Filter Method of Software Defect Management	61
2.13:	Software Defect Management	65
2.14:	Defect Life Cycle	68
2.14:	Defect Management reports	69
3.1:	Software Defect Predictive Model	83
4.1:	Interface for Experimental Setup	92
4.2:	Comparative Analysis for Accuracy when Relevant Features were Selected for Prediction	107
4.3:	Comparative Analysis for Accuracy when Relevant Features were not Selected for Prediction	108

4.4: Comparative Analysis for Precision when Relevant Features were Selected for Prediction	108
4.5: Comparative Analysis for Precision when Relevant Features were not Selected for Prediction	109
4.6: Comparative Analysis for Recall when Relevant Features were Selected for Prediction	109
4.7: Comparative Analysis for Recall when Relevant Features were not Selected for Prediction	110
4.8: Comparative Analysis for F1-Score when Relevant Features were Selected for Prediction	111
4.9: Comparative Analysis for F1-Score when Relevant Features were not Selected for Prediction	111
4.10: Comparative Analysis for Classification Time when Relevant Features were Selected for Prediction	112
4.11: Comparative Analysis for Classification Time when Relevant Features not were Selected for Prediction	112

## List of Algorithms

Title	Page
2.1 Algorithm Particles Swarm Optimization	25
2.2 Algorithm Ant Colony Optimization	31
2.3 Artificial Bee Colony Optimization	32
2.4 Algorithm Firefly	38
2.5 Algorithm Harmony Search	38

DO NOT COPY. LEAD CITY UNIVERSITY, NIGERIA

## List of Acronyms

<b>Abbreviation</b>	<b>Meaning</b>
AI	Artificial Intelligence
OP	Open Source
DT	Decision Tree
DP	Defect prediction
F1	F1-score
KNN	K-nearest Neighbour
LR	Linear Regression
ML	Machine Learning
RF	Random Forest
SDP	Software Defect Prediction
SVM	Support Vector Machine
KNN	K-Nearest Neighbour
ANN	Artificial Neural Network
NB	Naïve Baive
HSA	Harmony Search Algorithm
DLA	Deep Learning Algorithm
PSO	Particle Swarm Optimization
ACO	Ant Colony Optimization
ABCO	Artificial Bee Colony Optimization
FOA	Firefly Optimization Algorithm
RNG	Relative Neighbours Graph
CNN	Condensed Nearest Neighbour
MLR	Multinomial Logistic Regression

NFNO	Neural Fuzzy Network Optimization
ANNO	Artificial Neural Network Optimization
CB	Computational Biology,
GIP	General Image Processing
PGO	Power Generation Optimization
NR	Network Routing,
FF	Financial Forecasting
TS	Tabu Search
AIS	Artificial Immune System
SA	Simulated Annealing
FS	Feature Selection
FE	Feature Extraction

DO NOT COPY. LEAD CITY UNIVERSITY, NIGERIA