MACHINE LEARNING MODEL FOR PREDICTION OF POSTPARTUM DEPRESSION, A CASE OF MOMBASA COUNTY

GEORGE MONGARE KIMWOMI

A THESIS SUBMITTED TO THE INSTITUTE OF COMPUTING AND INFORMATICS IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF THE DEGREE OF MASTER OF SCIENCE IN INFORMATION TECHNOLOGY OF THE TECHNICAL UNIVERSITY OF MOMBASA

2023

DECLARATION

This thesis is my original work and has not been presented for academic award in any other university.

George Mongare Kimwomi

MSIT/0004/2020

Signature: _____ Date: _____

This thesis report has been submitted with our approval as University Supervisors.

Dr. Mvurya Mgala

Signature: _____ Date: _____

Dr. Fullgence Mwakondo

Signature: _____ Date: _____

Dr. Pamela Kimeto

Signature: _____ Date: JUNE 7, 2023

DEDICATION

I dedicate this thesis to my wife, children and dear parents for their motivation as I undertook the research.

ACKNOWLEDGEMENT

I wish to appreciate all the parties whose contribution and support enabled me to undertake this research. I acknowledge the guidance by Dr Mvurya Mgala, Dr Fullgence Mwakondo and Dr Pamela Kimeto right from shaping the research concept all through to the production of this thesis report which was a great learning experience for me. Much appreciation to this team for sacrificing their time and other resources to educate and guide me, and to check the document for errors to which I pray the almighty God to shower them with blessings.

I acknowledge the panel members of the various seminars I attended at the Technical University of Mombasa and Kabarak University for the positive critique and advice to the research work which contributed to the quality of this thesis report. I also recognize the support of Dr. Obadiah Musau, Dr. Kevin Tole and the staff of the Institute of Computing and Informatics at the Technical University of Mombasa for the immense support extended to me as I undertook the study.

I also acknowledge the role played by Dr Moses Thiga (Kabarak University) and Dr Mgala Mvurya (Technical University of Mombasa) in my selection as a beneficiary for research funding with the Kenya National Research Fund which facilitated my studies.

Many thanks to my wife Linet and children for the motivation and finally to the almighty God for the opportunity to undertake the research.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACRONYMS AND ABBREVIATIONS

AUC-ROC	Area under receiver operating characteristic curve
BDI	Beck Depression Inventory -II
CGTRH	Coast General Teaching and Referral Hospital
CRISP-DM	Cross-Industry Standard Process for Data Mining
EHR	Electronic Health Records
EPDS	Edinburgh Postnatal Depression Scale
KDD	Knowledge Discovery for Databases
KMO	Kaiser-Meyer-Olkin
ML	Machine Learning
MDD	Major Depressive Disorder
NYCCDRN	New York City Clinical Data Research Network data
PCA	principal component analysis
PHQ-9	Patient Health Questionnaire-9
PPD	Postpartum Depression
RF	Random Forest
ROC	**R**eceiver operating characteristic curve
SEMMA	Simple, Explore, modify, Model and Asses
SDG	Sustainable Development Goals
SFS	sequential forward selection

SPSS	Statistical Package for Social Sciences
SVM	Support Vector Machine
TSH	Tudor Sub-County Hospital
USPSTF	United States Preventive Services Task Force
UK	United Kingdom
UN	United Nations
WEKA	Waikato Environment for Knowledge Analysis
WCM	Weill Cornell Medicine
XRT	Extremely randomized trees

DEFINITION OF KEY TERMS

Area Under Receiver Characteristic Curve (AUC-ROC)	Performance matrix used to measure how well an algorithm can correctly discriminate output classes for all possible classifications
Binary Cross Entropy	the negative average of the log of corrected predicted probabilities for each class
Machine Learning:	Equipping computers with ability to learn from experience without directly reprogramming them
Model:	A conceptual representation of reality with a set of selected elements of the target system
Postpartum Depression:	An incapacitating but curable mental illness which affect mothers after delivery

ABSTRACT

Postpartum depression is a medical condition which affects many mothers. The condition exposes the mother and the newborn baby to illnesses that can lead to death. Management of the condition requires heavy expenditure incurred by the family, government, and stakeholders. The condition is also a source of many social problems. Manual systems are currently used to predict the condition which is slow and inconsistent. Machine learning technology which has reliably been used in prediction modelling in other domains can also be employed to build a model to predict mothers at risk of postpartum depression during pregnancy for primary prevention. In this study, perinatal records were collected from 324 mothers attending postnatal healthcare clinics at the Coast General Teaching and Referral Hospital and Tudor Sub-County Hospital in Mombasa County. The filter feature in WEKA was used to split the data into 70% and 30% for model training and testing respectively. Models were built on WEKA machine learning platform using logistic regression, support vector machine, extremely randomized trees, random forest and adaptive boosting algorithms which were identified from literature. A positive case of postpartum depression was defined as diagnosis or treatment of major depression within one year after delivery. Random forest model produced the best performance with a receiver operating characteristic (ROC) curve area of 0.863867 which is comparable within the bracket of high performing models. With this level of performance, the model can be used by healthcare staff to make quick and consistent prediction for early mitigation measures. Further research could be done with more data collected from other counties in Kenya.

Key words: Machine Learning, postpartum depression, prediction, mitigation measures.