A MULTI-OMICS CLASSIFIER FOR PREDICTION OF ANDROGEN DEPRIVATION TREATMENT RESPONSE IN PROSTATE CANCER DATASET

ALAGBE, EMMANUEL OLUWATOBA

(20PBF02172)

B. Tech (Hons) Microbiology, Ladoke Akintola University of Technology, Ogbomosho.

JANUARY, 2023

A MULTI-OMICS CLASSIFIER FOR PREDICTION OF ANDROGEN DEPRIVATION TREATMENT RESPONSE IN PROSTATE CANCER DATASET

BY

ALAGBE, EMMANUEL OLUWATOBA

(20PBF02172)

B. Tech (Hons) Microbiology, Ladoke Akintola University of Technology, Ogbomosho.

A DISSERTATION SUBMITTED TO THE SCHOOL OF POSTGRADUATE STUDIES IN PARTIAL FULFILMENT OF THE REQUIREMENT FOR THE AWARD OF MASTER OF SCIENCE (M.Sc.) DEGREE IN BIOINFORMATICS DEPARTMENTMENT OF COMPUTER AND INFORMATION SCIENCES, COVENANT UNIVERSITY.

JANUARY, 2023

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, ALAGBE, EMMANUEL OLUWATOBA (20PBF02172) declare that this dissertation entitled “A Multi-Omics Classifier for Prediction of Treatment Response in Prostate Cancer Dataset” is a representation of my work, and is written and implemented by me under the supervision of Dr. Itunuoluwa Isewon of the Department of Computer and information sciences, Covenant University, Ota, Nigeria. I attest that this dissertation has in no way been submitted either wholly or partially to any other university or institution of higher learning for the award of a masters’ degree. All information cited from published and unpublished literature has been duly referenced.

Signature: Date:

CERTIFICATION

This is to certify that this dissertation titled "**A MULTI-OMICS CLASSIFIER FOR PREDICTION OF ANDROGEN DEPRIVATION TREATMENT RESPONSE IN PROSTATE CANCER DATASET**" is original research carried out by **ALAGBE, EMMANUEL OLUWATOBA (20PBF02172)** in the Department of Computer and Information Sciences, College of Science and Technology, Covenant University, Ota, Ogun State, Nigeria under the supervision of Dr. Itunuoluwa Isewon. We have examined and found this work acceptable as part of the requirements for the award of Master of Science (M.Sc.) in Bioinformatics.

Dr. Itunuoluwa M. Isewon
Supervisor **Signature and Date**

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

Prof. Olusegun Folorunsho
External Examiner **Signature and Date**

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

DEDICATION

I dedicate this project to God Almighty for His unending supply of grace, and strength given to me during my master's degree Programme. Furthermore, to biomedical and computational researchers who paved the way, providing platforms for insightful biomedical investigations.

ACKNOWLEDGEMENTS

I am eternally grateful to God who has always caused me to triumph and has once again granted me the strength, wisdom, and grace to complete this Programme.

I acknowledge the Chancellor, Dr. David O. Oyedepo for his commitment to the vision of this exemplary university and for remaining a viable role model to us. I am grateful to the Vice-chancellor, Prof. Abiodun H. Adebayo for his visionary leadership and partnership with the chancellor to see the vision of the university become a reality. Furthermore, I acknowledge all members of management and the board of regents for their commitment to the university.

I appreciate the Covenant Applied Informatics and Communication Africa Centre of Excellence (CApIC-ACE) for sponsoring my studentship. I truly appreciate the head of department Prof. Olufunke O. Oladipupo for being accessible and consistently insistent on comprehensive knowledge of the concepts taught in this Programme. My profound gratitude goes to my supervisor Dr. Itunuoluwa M. Isewon for believing in me, supporting me through this Programme, for her ever-present willingness to pay the sacrifice to see me grow, and ultimately for insisting on quality research.

I am eternally grateful to my amazing parents: Pastor and Evangelist Alagbe, for their encouragement, prayers, and selfless sacrifices which continues to serve as a robust foundation for whatever achievement I make. To my wonderful sisters Ayo and Moyinoluwa Alagbe, thanks for believing in your brother. Thanks for your immense financial and emotional support.

I am appreciative of my dear partner and muse: Ogungbesan Temitope, thanks for tutoring me, inspiring me, and brainstorming with me. I'm thankful to my amazing coursemates: Adebola precious, Erika, Shekari, Itunu, Tobi, and Miracle for their amazing support as we journeyed together through the MSc programme. I acknowledge every member of the CUPGF executives, thanks for carrying the burden with me which enabled me to balance managing the fellowship and my academics. To my beloved brothers: Olumikki, Dolad, and Sogo, I'm grateful for your encouragement and your being great examples to me. I am also grateful to my friends without whom my CU story would be largely incomplete: Kaki joy, Ify, Daniel Olusegun, Daniel (Black), Lois, Tomi, Nonye, Tomi (rommie), Folo, Stephen, Samuel, Faith, Adegbesin, Akinbola David, Esther salami, Sonia, Daniella to name a few. You guys have been awesome.

Finally, I'm overflowing with gratitude to my dear friend: Abimbola (mo)Bowofoluwa Sharon, whose persistence resulted in me applying for the ACE scholarship thus starting the cascade leading up to this success. Thanks for your encouraging words and prayers, thanks for being my cheerleader from the very beginning, and thanks for continually inspiring me to be a better researcher and person in general. You are greatly cherished and acknowledged.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Prostate cancer (PCa) is estimated to cause over 375,000 deaths and nearly 1.4 million new cases globally. Several factors contribute to PCa heterogeneity, consequently, the stage of the disease decides the strategy employed in combating the disease. The problem of missing data frequently plagues clinical research. The primary treatment outcomes in the TCGA prostate cancer phenotypic dataset had 120 (19.26%) missing values. Treatment strategies could be negatively impacted by limited care giver experience, "trial and error" approaches to treatment, and the genetic makeup of an individual. To the best of our knowledge, using Machine Learning (ML) to forecast treatment response among PCa patients had not been investigated. The aim of this study is to develop a classifier (that acts as a decision support system) from multi-omics datasets for predicting treatment response in PCa patients. RNAseq, miRNAseq, reverse phase protein array (RPPA), copy number variation (CNV) were used in the study. This study employed R programming to preprocess the data. Differential expression analysis for the RNAseq and miRNAseq conducted using the DESeq2 library. Python programming was used to implement the ML algorithms which include XGBoost, Adaboost, multilayer perceptron, decision tree, logistic regression, support vector machine, gradient boosting classifier, Random forests, naive bayes, and K -nearest neighbors. The performance metrics used include macro f1 score, macro recall, macro precision, weighted f1 score, weighted recall, weighted precision, specificity, sensitivity, accuracy, and area under the receiver operator curve. It was discovered that tree-based models were better for the task than probability and kernel-based models. This study computationally demonstrated that muti-omics strategies are generally superior to single-omics strategies, but the adoption of such strategy isn't a foolproof solution. A classifier capable of predicting treatment outcomes amongst PCa patients was built and the predicted labels for patients with missing phenotypic values in the TCGA dataset was provided.

Keywords: Prostate cancer, Precision Oncology, Multi-omics, Machine Learning Treatment response.