GENOMICS-BASED ARTIFICIAL NEURAL NETWORKS ENSEMBLE FOR MEDICAL DIAGNOSIS: A CASE STUDY OF NON-SMALL CELL LUNG CANCER

BY

ADETIBA EMMANUEL

(CUGP040050)

BEING

A Ph.D THESIS SUBMITTED TO THE DEPARTMENT OF ELECTRICAL AND INFORMATION ENGINEERING, COLLEGE OF SCIENCE AND TECHNOLOGY, COVENANT UNIVERSITY, CANAANLAND, OTA, NIGERIA, IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE AWARD OF THE DOCTOR OF PHILOSOPHY DEGREE (ELECTRICAL AND INFORMATION ENGINEERING)

OCTOBER 2013

DECLARATION

I hereby declare that I carried out the work reported in this thesis in the Department of Electrical and Information Engineering, College of Science and Technology, Covenant University, Ota, Nigeria under the supervision of Prof. F.A. Ibikunle and Prof. E.F. Adebiyi.

I also solemnly declare that no part of this report has been submitted here or elsewhere in a previous application for award of a degree. All sources of knowledge used have been duly acknowledged.

ADETIBA Emmanuel

CUGP040050

CERTIFICATION

This is to certify that the thesis titled "Genomics-based artificial neural networks ensemble for medical diagnosis: a case study of non-small cell lung cancer" by Adetiba Emmanuel meets the requirements and regulations governing the award of the Doctor of Philosophy degree in Electrical and Information Engineering of Covenant University and is approved for its contribution to knowledge and literary presentation.

Supervisor	Sign:----------	----------
	Name: Prof. F.A. Ibikunle	Date
Co-Supervisor	Sign:----------	----------
	Name: Prof. E.F. Adebiyi	Date
Head of Department	Sign:----------	----------
	Name: Dr. S.N. John	Date

DEDICATION

First and foremost, I dedicate this thesis to God the Father, Son and Holy Spirit. And also to Oluwatofunmi Joyce Adetiba (my precious daughter), Joy N. Adetiba (my priceless wife and best friend) and Theresa Wuraola Adetiba (my golden mum).

ACKNOWLEDGEMENTS

There is indeed a Spirit in man, but it is the inspiration of the Almighty that gives him understanding. On this note, I greatly appreciate and acknowledge my heavenly Father; the God of all inspirations for being the Alpha and Omega of this work.

I appreciate my Supervisor; Prof. F.A. Ibikunle for wise counsel, rare timeliness and sound technical guidance that made this work an outstanding success. He practically exemplified the fact that when the going gets tough, the tough gets going. Sir, you are a gem and I say a big thank you.

My Co-Supervisor; Prof. E.F. Adebiyi is an embodiment of possibility mentality. He stepped in and gave a unique direction to the course of this research work. I greatly appreciate the slot you carved for me in your ever tight schedule. God bless you sir.

I acknowledge my HOD; Dr. S.N. John for positive dispositions and remarks at all times which provided the nudge to stay focused. I appreciate Prof. C.O.A. Awosope and Dr. O.E. Agboje for encouragements, moral supports and fatherly dispositions at all times. Thanks to Dr. V.O. Matthews for always making his broad technical shoulder available for me to climb on and for endless encouragements. I appreciate Prof. A.A.A. Atayero (CU Deputy Vice-Chancellor, Academic) for modelling what it means to be a dogged and focused academician. I acknowledge Mr. Richard Awoseyin for being my first coach in the ever evolving field of ICT. Thank you to Dr. F.E. Idachaba, Dr. S.A. Daramola, Dr. A.U. Adoghe and Dr. C. Ndujiuba for relevant contributions to this work. To my friends and fans in the school of life; Engr. A.A. Awelewa, Engr. Isaac Samuel, Dr. Olushola James, Dr. S.A. Fatumo, Engr. (Mrs.) Badejo and all my colleagues in the department, I say a big thank you.

I specially acknowledge my brother, Barr. ‘Yemi Adetiba for his investments on my life (in cash and kind) at the time I needed it most. Thank you so much for always been a fatherly

brother! My heartfelt appreciations go to my other siblings; Mrs. Funmilayo Jimoh, Mrs. Funmilola Omotosho, Mr. Marcus Adetiba, ‘Gbenga Adetiba and ‘Wumi Jacob for unending supports in all my life endeavours.

To my darling wife; Joy N. Adetiba and precious daughter; Oluwatofunmi Joyce Adetiba, I say a big thank you. You have being part of my life in meaningful ways and having you as my God ordained core team members in the game of life gives me an assurance of being a perpetual winner.

I cannot but immensely appreciate the management of Covenant University under the leadership of our amiable Vice-Chancellor, Prof. C.K. Ayo, for providing a highly conducive environment for research and learning. Covenant University is indeed a modern cradle where the nobility of knowledge is always maintained and a shining light bearer on the African continent.

ABSTRACT

The adoption of information and communications technology to support cancer research has triggered the use of Deoxyribonucleic Acid (DNA) sequencing to generate complete genomes of different types of human cancer. This usually results in huge volume of genomics data which contain hidden mutations' profiles that cannot be accurately mined with molecular approach. However, using digital techniques to unravel the inherent mutations holds a lot of promises for early diagnosis of cancerous patients accurately and efficiently. In this research work, pre-processing and features extraction algorithms were developed to implement a genomics and bagged Artificial Neural Networks (ANNs) ensemble-based diagnostic system for non-small cell lung cancer. The pre-processing algorithm encodes and vectorises the genomics nucleotides using their relative molecular masses and normalizes the resultant vectors for any given patient. For the features extraction algorithm, two distance metrics were hybrided to extract eight features from the normalized genomics vectors. This helped to eliminate the "curse of dimensionality" for the bagged ANNs ensemble platform that was adopted as the classifier in the system. The performance of the system was evaluated using Mean Square Error (MSE) and confusion matrix when a single ANN was utilized and when bagged ANNs ensemble was employed. The ensembled ANNs option did not only achieve perfect output stability but also achieved a good MSE of 0.0139 and an accuracy of 100% which is by far better than the single ANN with MSE of 0.0817, accuracy of 83.3% and serious output instability. The 100% accuracy of the bagged ANNs ensemble platform is not unexpected because of the genomics data on which it is built. The result of the bagged ANNs ensemble-based system in this work is better in comparison with other similar systems in the literature.