USE OF MACHINE LEARNING FOR THE PREDICTION OF DIABETES FROM PHOTOPLETHYSMOGRAPHY (PPG) MEASUREMENTS & PHYSIOLOGICAL CHARACTERISTICS

Chirath Yudara Hettiarachchi

188098R

Thesis/Dissertation submitted in partial fulfilment of the requirements for the degree Master of Science in Computer Science and Engineering.

Department of Computer Science & Engineering

University of Moratuwa

Sri Lanka

March 2020

DECLARATION

I declare that this is my own work and this dissertation does not incorporate without acknowledgement any material previously submitted for a Degree or Diploma in any other University or institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

Also, I hereby grant to University of Moratuwa the non-exclusive right to reproduce and distribute my dissertation, in whole or in part in print, electronic or other medium. I retain the right to use this content in whole or part in future works (such as articles or books).

Signature: *UOM Verified Signature* Date: 2020 / 04 / 21

The above candidate has carried out research for the Masters thesis/Dissertation under my supervision.

Signature of the Supervisor: Date:

ACKNOWLEDGEMENTS

I am sincerely grateful to my supervisor Dr Charith Chitraranjan for having me as a student to pursue my research interests in the area of biomedical signal processing and machine learning. The constant supervision and guidance provided, enabled me to successfully complete my research. I would also like to thank the Head of the Department of Computer Science & Engineering Dr Amal Shehan Perera and the non-academic staff for providing me an excellent working environment during the period of my research. The progress review committee comprising of Dr Charith Chitraranjan, Dr Anjula De Silva and Dr Uthayasanker Thayasivam extended valuable feedback and advice towards the development of this research.

I would also like to highlight the financial grant provided by the Senate Research Council (SRC/LT/2019/33) of the University of Moratuwa, which ensured the success of the research. The financial support extended by the Prof V.K Samaranayake Research Grant of the LK Domain Registry assisted me in presenting my research at the Artificial Intelligence in Medicine (AIME 2019) conference held in Poland and enabled me to obtain important exposure in the global research arena. I would also like to thank the AIME 2019 organizing committee for the Student Travel Grant provided for me to attend the conference.

I thank my fellow research colleagues for their support, and for creating a conducive research environment through knowledge and experience sharing filled with refreshing activities and entertainment. I would like to thank my family for their support, and understanding regarding my passion towards research. Finally, I would like to thank all my previous mentors and teachers especially from the Department of Electronics & Telecommunication Engineering for grooming me towards research and helping, advising me at every step of the way. All these efforts and support extended to me will be crucial milestones toward the success of my future research.

ABSTRACT

Prediction of Diabetes Using Non-Invasive Photoplethysmography (PPG) Measurements & Physiological Characteristics

Type 2 Diabetes (T2D) is a chronic disease affecting millions of people worldwide. It is a result of impaired glucose regulation, leading to abnormally high levels of glucose causing microvascular and macrovascular problems. The failure to timely identify and treat, results in complications such as limb amputations, blindness and heart disease. Busy unhealthy lifestyles are a root cause and not much effort undertaken to obtain regular health checkups for early T2D detection.

Photoplethysmography (PPG) is a non-invasive, optic technique mostly used towards disease estimation in clinical environments. Recent technological advancements have integrated PPG sensors within smartphones and wearables. However, these signals suffer from various noise components, which is intensified in signals acquired in routine everyday environments. The research analysed the feasibility of short (~2.1s) PPG segments in order to address these limitations and identify biomarkers related to T2D. The identified biomarkers mainly relate to the vascular system of the body. Several classification algorithms were evaluated using cross validation to estimate T2D, focussing on a public PPG dataset. Linear Discriminant Analysis (LDA) achieved the highest area under the ROC curve of 79% for the estimation of T2D in a setting where healthy individuals, T2D only, T2D subjects with hypertension and prehypertension were present.

It is important to identify relationships between standard medical measures such as Fasting Blood Glucose (FBG) and PPG features, for better understanding T2D estimation. FBG measurements were collected, and several regression algorithms evaluated using leave-one-out cross validation to assess the suitability of predicting FBG using PPG features. The results were examined using the Clarke's Error Grid, where 75% & 22.5% of predictions were distributed in regions A & B respectively for both ElasticNet and Lasso Regression. The results were comparable with long PPG signal based approaches. The suitability of the method in practical environments was evaluated using simulated PPG signals with noise and motion artifacts. The ElasticNet Regression achieved 70% and 27.5% in regions A & B respectively. The analysis of short PPG segments shows promise towards the development of an early T2D estimation system in a routine everyday environment.

Keywords: Type 2 Diabetes, Photoplethysmography, Machine Learning, Classification, Regression

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

PPG	Photoplethysmography
APG	Accelerated Photoplethysmography
T2D	Type 2 Diabetes
FBG	Fasting Blood Glucose
ML	Machine Learning
CNN	Convolution Neural Networks
LSTM	Long Short-Term Memory
LDA	Linear Discriminant Analysis
AI	Augmentation Index
SVD	Singular Value Decomposition
BMI	Body Mass Index

TABLE OF CONTENTS