A Predictive Analysis of Electronic Healthcare Records for Stroke Symptoms

Pattanapong Chantamit-o-pas

Faculty of Engineering and Information Technology
University of Technology Sydney

A thesis submitted in fulfilment of the requirement for
the degree of Doctor of Philosophy

November 2019

CERTIFICATE OF ORIGINAL AUTHORSHIP

I, Pattanapong Chantamit-o-pas, declare that the thesis is submitted in fulfilment of the requirements for the award of Doctoral of Philosophy, in the School of Computer Science, Faculty of Engineering and Information Technology at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise referenced or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This document has not been submitted for qualifications at any other academic institution.

This research is supported by the Australian Government Research Training Program.

Production Note:
Signature removed prior to publication.

Signature: .....

Date: ……..21st..November..2019.....

ACKNOWLEGEMENT

First of all, I would like to express my sincere gratitude to my principal supervisor, Dr. Madhu Goyal. My priceless journey started when she accepted me to be her PhD student in December 2016. She has given me not only excellent guidance in my research, but also great support and encouragement. This really helps me to finish my research and learn how to be a good researcher.

Most importantly, my journey could not have happened without the financial support given by the Royal Thai government scholarship, offered by the Ministry of Science and Technology of Thailand. It has been an excellent opportunity to have the opportunity to gain more knowledge and experiences aboard in Australia. Furthermore, without support the secondary data from the Department of Medical Services (affiliated with Ministry of Public Health of Thailand), and Office of Educational Affairs (affiliated with Royal Thai Embassy to Australia), the research would not have been accomplished. With this acknowledgement, we would like to express our sincere appreciation to them.

Additionally, I would like to thank my parents who always stand beside me and walk through all bad situations with me. I am who I am because of them, so it is undoubtedly that this success also belongs to my parents. I cannot be here without their endless care, support, and encouragement. Likewise, I would like to thank my family, Professor Tony Moon, my friends, and my colleagues who always help me to get over all the bad situations.

Finally, I dedicate this thesis to Dr. Madhu Goyal, who is a key person of this work, my parents, and my family.

LIST OF PUBLICATIONS

1. Chantamit-o-pas, P. & Goyal, M. 2018, 'A Case-Based Reasoning Framework for Prediction of Stroke', in D.K. Mishra, A.T. Azar & A. Joshi (eds), *Information and Communication Technology: Proceedings of ICICT 2016*, Springer Singapore, Singapore, pp. 219-27.
2. Chantamit-o-pas, P. & Goyal, M. 2017, 'Prediction of Stroke Using Deep Learning Model', *International Conference on Neural Information Processing*, Springer, Guangzhou, China, pp. 774-81.
3. Chantamit-o-pas, P. & Goyal, M. 2018, 'Long Short-Term Memory Recurrent Neural Network for Stroke Prediction', *International Conference on Machine Learning and Data Mining in Pattern Recognition*, Springer International Publishing, Cham, pp. 312-23.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Abstract

Cerebrovascular symptoms, commonly known as stroke, can affect different parts of the human body depending on the area of brain affected. The patients who survive usually have a poor quality of life because of serious illness, long-term disability and become a burden to their families and the health care system. There is a strong demand for the management focused on prevention and early treatment of disease by analysing different factors. However, a high volume of medical data, heterogeneity, and complexity have become the biggest challenges in stroke symptoms prediction. Algorithms with very high level of accuracy are, therefore, vital for medical diagnosis. The development of such algorithms nevertheless still remains obscure despite its importance and necessity for healthcare. Electronic Healthcare Records (EHRs) describe the details about patient's physical and mental health, diagnosis, lab results, treatments or patient care plan and so forth. The huge amount of information in these records provides insights about the diagnosis and prediction of various diseases. Currently, the International Classification of Diseases, 10^{th} Revision or ICD-10^{th} codes is used for representing each patient record. The huge amount of information in these records provides insights about the diagnosis and prediction of various diseases. Various machine learning techniques are used for the analysis of data derived from these patient records. The predictive techniques have been widely applied in clinical decision making such as predicting occurrence of a disease or diagnosis, evaluating prognosis or outcome of diseases and assisting clinicians to recommend treatment of diseases. However, the conventional predictive models or techniques are still not effective enough in capturing the underlying knowledge because it is incapable of simulating the complexity on feature representation of the medical problem domains. This research used aggregated files of Electronic Healthcare Records

(EHRs) from Department of Medical Services, The Ministry of Public Health of Thailand between 2015 and 2016. The empirical research is intended to evaluate the ability of machine learning and deep learning to recognize patterns in multi-label classification of stroke. This research aims at the investigation of five techniques: Support Vector Machine (SVM); k-Nearest Neighbours (k-NN); Backpropagation; Recurrent Neural Network (RNN); and Long Short-Term Memory - Recurrent Neural Network (LSTM-RNN). These are powerful and widely used techniques in machine learning and bioinformatics. First, we decoded ICD-10th codes into the health records, as well as other potential risk factors within EHRs into the pattern and model for prediction. Second, we purposed a conceptual Case Based Reasoning (CBR) framework for stroke disease prediction that uses previous case-based knowledge. A conceptual case-based reasoning framework to predict from patients' health risk factors and to recognize a particular case that probably develop stroke and prepare or warn patients to handle disease burden outcome. It describes the design, implementation and evaluation of a novel system to facilitate stroke prediction, which relies on data collected from EHRs. Finally, the effectiveness of Backpropagation; RNN; and LSTM-RNN for prediction of stroke based on healthcare records is modelled. The results show several strong baselines that include accuracy, recall, and F1 measure score. Consequently, deep learning allows the disclosure of some unknown or unexpressed knowledge during prediction procedure, which is beneficial for decision-making in medical practice and provide useful suggestions and warnings to patient about unpredictable stroke.