

Washington University in St. Louis

Washington University Open Scholarship

McKelvey School of Engineering Theses & Dissertations

McKelvey School of Engineering

Spring 5-2020

Predicting Disease Progression Using Deep Recurrent Neural Networks and Longitudinal Electronic Health Record Data

Seunghwan Kim

Washington University in St. Louis

Follow this and additional works at: https://openscholarship.wustl.edu/eng_etds



Part of the [Artificial Intelligence and Robotics Commons](#), [Databases and Information Systems Commons](#), [Engineering Commons](#), [Numerical Analysis and Scientific Computing Commons](#), and the [Other Medicine and Health Sciences Commons](#)

Recommended Citation

Kim, Seunghwan, "Predicting Disease Progression Using Deep Recurrent Neural Networks and Longitudinal Electronic Health Record Data" (2020). *McKelvey School of Engineering Theses & Dissertations*. 690.

https://openscholarship.wustl.edu/eng_etds/690

This Thesis is brought to you for free and open access by the McKelvey School of Engineering at Washington University Open Scholarship. It has been accepted for inclusion in McKelvey School of Engineering Theses & Dissertations by an authorized administrator of Washington University Open Scholarship. For more information, please contact digital@wumail.wustl.edu.

Abstract

Predicting Disease Progression Using Deep Recurrent Neural Networks and Longitudinal
Electronic Health Record Data

By

Seunghwan Kim

Master of Science in Computer Science

Washington University in St. Louis, 2020

Research Advisor: Professor Philip R.O. Payne

Electronic Health Records (EHR) are widely adopted and used throughout healthcare systems and are able to collect and store longitudinal information data that can be used to describe patient phenotypes. From the underlying data structures used in the EHR, discrete data can be extracted and analyzed to improve patient care and outcomes via tasks such as risk stratification and prospective disease management. Temporality in EHR is innately present given the nature of these data, however, and traditional classification models are limited in this context by the cross-sectional nature of training and prediction processes. Finding temporal patterns in EHR is especially important as it encodes temporal concepts such as event trends, episodes, cycles, and abnormalities. Previously, there have been attempts to utilize temporal neural network models to predict clinical intervention time and mortality in the intensive care unit (ICU) and recurrent neural network (RNN) models to predict multiple types of medical conditions as well as medication use. However, such work has been limited in scope and generalizability beyond the immediate use cases that have been focused upon. In order to extend the relevant knowledge-base, this study demonstrates a predictive modeling pipeline that can extract and integrate

clinical information from the EHR, construct a feature set, and apply a deep recurrent neural network (DRNN) to model complex time stamped longitudinal data for monitoring and managing the progression of a disease condition. It utilizes longitudinal data of pediatric patient cohort diagnosed with Neurofibromatosis Type 1 (NF1), which is one of the most common neurogenetic disorders and occurs in 1 of every 3,000 births, without predilection for race, sex, or ethnicity. The prediction pipeline is differentiable from other efforts to-date that have sought to model NF1 progression in that it involves the analysis of multi-dimensional phenotypes wherein the DRNN is able to model complex non-linear relationships between event points in the longitudinal data both temporally and also within the cross-sectional observation. Such an approach is critical when seeking to transition from traditional evidence-based care models to precision medicine paradigms. Furthermore, our predictive modeling pipeline can be generalized and applied to manage the progression and stratify the risks in other similar complex diseases, as it can predict multiple set of sub-phenotypical features from training on longitudinal event sequences.

Commented [SK1]: Updated per request from Aditi