Shedding Light on the Black Box: Explaining Deep Neural Network Prediction of Clinical Outcome

Authors : Yijun Shao, Yan Cheng, Rashmee U. Shah, Charlene R. Weir, Bruce E. Bray, Qing Zeng-Treitler

Abstract : Deep neural network (DNN) models are being explored in the clinical domain, following the recent success in other domains such as image recognition. For clinical adoption, outcome prediction models require explanation, but due to the multiple non-linear inner transformations, DNN models are viewed by many as a black box. In this study, we developed a deep neural network model for predicting 1-year mortality of patients who underwent major cardio vascular procedures (MCVPs), using temporal image representation of past medical history as input. The dataset was obtained from the electronic medical data warehouse administered by Veteran Affairs Information and Computing Infrastructure (VINCI). We identified 21,355 veterans who had their first MCVP in 2014. Features for prediction included demographics, diagnoses, procedures, medication orders, hospitalizations, and frailty measures extracted from clinical notes. Temporal variables were created based on the patient history data in the 2-year window prior to the index MCVP. A temporal image was created based on these variables for each individual patient. To generate the explanation for the DNN model, we defined a new concept called impact score, based on the presence/value of clinical conditions' impact on the predicted outcome. Like (log) odds ratio reported by the logistic regression (LR) model, impact scores are continuous variables intended to shed light on the black box model. For comparison, a logistic regression model was fitted on the same dataset. In our cohort, about 6.8% of patients died within one year. The prediction of the DNN model achieved an area under the curve (AUC) of 78.5% while the LR model achieved an AUC of 74.6%. A strong but not perfect correlation was found between the aggregated impact scores and the log odds ratios (Spearman's rho = 0.74), which helped validate our explanation.

Keywords : deep neural network, temporal data, prediction, frailty, logistic regression model
Conference Title : ICHI 2019 : International Conference on Health Informatics
Conference Location : Rome, Italy
Conference Dates : January 17-18, 2019

Open Science Index, Health and Medical Engineering Vol:13, No:01, 2019 publications.waset.org/abstracts/99910.pdf