Longitudinal Prostate Cancer Outcome Predictive Model for CRPC Treatment Using Phased-LSTM

Jihwan Park\textsuperscript{1,2}, Mi Jung Rho\textsuperscript{2}, Hyong Woo Moon\textsuperscript{3}, and Ji Youl Lee\textsuperscript{3}

\textsuperscript{1}Department of Biomedicine & Health Sciences, College of Medicine, The Catholic University of Korea, Seoul, 06591 Republic of Korea
\textsuperscript{2}Catholic Cancer Research Institute, College of Medicine, The Catholic University of Korea, Seoul, 06591 Republic of Korea
\textsuperscript{3}Department of Urology, Seoul St. Mary’s Hospital, College of Medicine, the Catholic University of Korea, Seoul, 06591 Republic of Korea

Corresponding author: Ji Youl Lee (e-mail: uroljy@catholic.ac.kr).

This work was supported by Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government (MSIT) (2018-2-00861, Intelligent SW Technology Development for Medical Data Analysis). This research was supported by Smart Healthcare Research Grant through the Daewoong Foundation (DF-201907-0000001).

ABSTRACT

It is difficult to make accurate predictions based on medical data because patients’ treatment periods are sampled irregularly. It is particularly desirable to predict castration-resistant prostate cancer (CRPC) in prostate cancer (PCa) patients and this study aims to predict patients’ likely outcomes to support physicians’ decision making. Serial data is collected from 1592 PCa patients and phased long-term short-term memory (phased-LSTM) model with a special module called a “time-gate” is used to process the irregularly sampled data sets. A synthetic minority oversampling technique is used to overcome the data imbalance between two patient groups, those with and without CRPC treatment. The phased-LSTM model is able to predict the CRPC outcome with an accuracy of 90.1\% using 120 days of data or 95.9\% using 360 days of data. The validation loss converged slowly with 120 days of data and quickly with 360 days of data. In both cases, the prediction model takes 4 epochs to build. The overall CPRC Likely outcome prediction model using irregularly sampled serial medical data is accurate and can be used to support physicians’ decision making, which saves time compared to cumbersome serial data reviews. This study can be extended to make clinically meaningful prediction models.

Keywords

Prostate Cancer, Castration-resistant Prostate Cancer, Deep Learning, Phased Long-term Short-Term Memory