

Postgraduate Seminar Series

Speaker Information

Joydeb Kumar Sana (Student No. 1018054003) is a part-time Ph.D. student in the Department of Computer Science and Engineering (CSE) at the Bangladesh University of Engineering and Technology (BUET). He earned his undergraduate degree from Rajshahi University of Engineering and Technology (RUET) in 2006 and M.Sc. degree from BUET in 2015. His research interests encompass Applied Machine Learning, Big Data, Financial Informatics, Health Informatics and Data Privacy. He is currently doing his Ph.D. under the guidance of Prof. M Saifur Rahman. In this talk, he will discuss his ongoing research.

Patient Similarity Computation for Clinical Decision Support: An efficient use of Data Transformation and Spark Computation

Patient similarity computation (PSC) is a fundamental problem in healthcare informatics. The aim of the patient similarity computation is to measure the similarity among patients according to their historical clinical records, which helps to improve clinical decision support. A novel data transformation (DT) based distributed patient similarity computation (DPSC) technique has been presented in this paper utilizing an effective combination of time series and static data. Time series data from patients admitted to the intensive care unit (ICU), including metrics like heart rate, blood pressure, Oxygen saturation, respiration etc. are measured at regular intervals. The static data are mainly patient background and demographic data, including age, weight, height and gender. Static data has been used for clustering the patients. Before feeding the demographic data to the machine learning model adaptive Weight-of-Evidence (aWOE) and Z-score data transformation (DT) methods have been performed which improve the prediction performances. In aWOE-based patient similarity models, sensitive patient information has been processed using aWOE which preserves the data privacy of the trained models. We used the dynamic time warping (DTW) for time series similarity which is robust and a very popular method. However, DTW is not suitable for big data due to the huge computational run-time. To overcome this problem distributed DTW computation is used in this study. For Diagnosis, our data transformation based approach boosts prediction performance by as much as 11.4%, 10.20% and 12.6% in terms of AUC, accuracy, and F-measure, respectively. In case of Congestive heart failure (CHF), our proposed method achieves performance enhancement up to 15.9%, 10.5%, 21.9% for the same measures, respectively. The proposed method reduces the computation time by as high as 40%. We are confident in the sturdiness and effectiveness of the proposed technique for finding similar patients, making it well-suited for integration into clinical decision support systems.