Contribution ID: **51**   Type: **Oral Presentation**

Analysis of Type 1 diabetes verbal autopsy data by machine learning techniques

Tuesday, 9 July 2019 10:20 (20 minutes)

Big data is a term used for data sets with large, diverse and complex structures that are often quite difficult to analyze or visualize using traditional computing methods and approaches. Machine learning (ML) techniques are effective in analyzing these types of data and extracting information from these types of data. Large sets of data are generated by health care systems from record keeping of patients and this data supports a wide range of medical decisions like population health surveillance and disease management for the overall improvement of the quality of health care delivery. In areas where there are no health registration systems like the rural areas of most underdeveloped and developing African countries, a method of verbal autopsy is relied on to give information of a likely cause of death. In this study, type 1 diabetes (T1DM) verbal autopsy data from MRC/Wits Agincourt Unit was used as a test case for applying modern machine learning classification methods to ascertain the cause of death by type 1 diabetes. Machine learning techniques of artificial neural networks (ANNs) and random forests (RF) which are realized with a keras and tensorflow front end were used for the classification task. Machine learning algorithms automatically learn to make accurate predictions based on past observations by learning patterns in the data and for this study, they learn the features present in diabetic patients and are able to identify patients who actually could have died from the disease. This is the first study on type 1 diabetes verbal autopsy data by the two machine learning techniques in South Africa. The dataset was negatively skewed and performance metrics of precision, recall, confusion matrix and the roc-score were used on these classifiers. Results obtained show that the random forest classifier did the classification task of deaths by diabetes better than the artificial neural network. In particular the roc-score compares favourably with the study that was done by two clinician specialists in the disease whose study was similar, ascertaining the number of deaths by type 1 diabetes from the data.

Apply to be
 considered for a student
   award (Yes / No)?

yes

Level for award
 (Hons, MSc,
   PhD, N/A)?

PhD

Primary author: Ms MANAKA, THOKOZILE (LESOTHO)

Co-authors: Dr WADE, Alisha (University of the Witwatersrand);   KAR, Deepak (University of Witwatersrand)

Presenter: Ms MANAKA, THOKOZILE (LESOTHO)

Session Classification: Applied Physics

Track Classification: Track F - Applied Physics