International Conference on

Big Data Analytics and Data Mining

November 25-26, 2019 | Frankfurt, Germany

Predicting preterm birth from electrical impedance spectroscopy: A comparative study using different machine learning algorithms

David Tian
University of Sheffield, United Kingdom

Preterm birth (PTB) refers to labour before 37 weeks of pregnancy. PTB can cause new-born deaths and long term diseases for the babies. The cost of care for surviving premature babies during their first year of life could be much higher than for term babies. Therefore, predicting and preventing PTB accurately early during pregnancy can effectively improve birth outcomes. However, current approaches to PTB prediction give limited prediction of PTB risk. Electrical impedance spectroscopy (EIS) is a technology concerned with applying a sinusoidal current to a sample under test to measure its impedance over a specified frequency range. EIS has been a powerful technique for measuring the electrical properties of materials. In machine learning, the class distribution of a training set critically affects the performance of the model trained in that a model trained on an imbalanced training set tends to bias towards predicting the majority class for unseen patterns. An EIS dataset collected for PTB prediction is often imbalanced in that the majority of the patients in the dataset are on-term and only contains a small number of preterm patients. This work proposes a machine learning methodology for predicting PTB from an imbalanced cervical EIS dataset collected during 20 to 22 weeks of pregnancy. The methodology integrates 3 pre-processing algorithms (polynomial feature construction, oversampling and information gain feature selection) with 4 classification algorithms: logistic regression, neural networks, naive Bayes and random forest, respectively. The EIS dataset was randomly split into training (66%) and test (34%) sets. The methodology was run 100 times on different training and test data and each time the 4 classifiers were trained on the training set and evaluated on the corresponding test set, respectively. The results show that neural networks achieved the highest AUC 78.2% on a test set compared with the other classification algorithms.

Biography

David Tian received a Ph.D. in Computer Science from The University of Manchester in 2009. His research interests are AI, machine learning, data science and their applications to healthcare, engineering and biology. He is a research associate at The University of Sheffield. He has 16 publications at journals and international conferences in AI, machine learning and data science. He has been a reviewer of reputed journals and international conferences including IEEE Transactions on Fuzzy Systems and IEEE International Conference on Data Mining.

e: dtian09@gmail.com

Notes: