How to Improve Diagnostic AI in Medicine?

The University of Johannesburg recently publicised a study published in *Informatics in Medicine Unlocked* examining how to improve machine learning (ML) algorithms for healthcare use. Algorithms can better identify healthy individuals since they are often trained on 'unbalanced' datasets containing more healthy than sick individuals. Since fewer data points exist describing ill patients, algorithms learning from these datasets can be inaccurate at diagnosing sick individuals.

Therefore, the study's investigators, Drs Ibomoiye Domor Mienye and Yanxia Sun, determined how building cost sensitivity into algorithmic models affected their diagnostic performance (i.e., algorithms receive larger penalties for false negatives than for false positives). The rationale is that telling a sick person that they are healthy carries more danger than the other way round.

Logistic regression, decision tree, XGBoost, and random forest models were trained on public learning datasets for diabetes, breast cancer, cervical cancer and chronic kidney disease obtained from the University of California, Irvine machine learning repository. These are supervised binary classification algorithms that learn from 'yes/no' datasets. Each dataset contains a diagnosis and the relevant diagnostic data for each patient.

In almost every case, adding a penalty boosted performance so that the algorithms identified fewer healthy people as sick (precision) and sick people as healthy (recall). For example, in chronic kidney disease, cost-sensitivity increased random forest precision to 0.990 and recall to a perfect 1.000 from 0.972 and 0.946, respectively. For cervical cancer, cost-sensitivity in random forest and XGBoost improved precision and recall to 1.000 from high score. Thus, adding penalties effectively compensated for the imbalance in the datasets.

Source: Informatics in Medicine Unlocked, Press Release

Published on : Mon, 8 Nov 2021