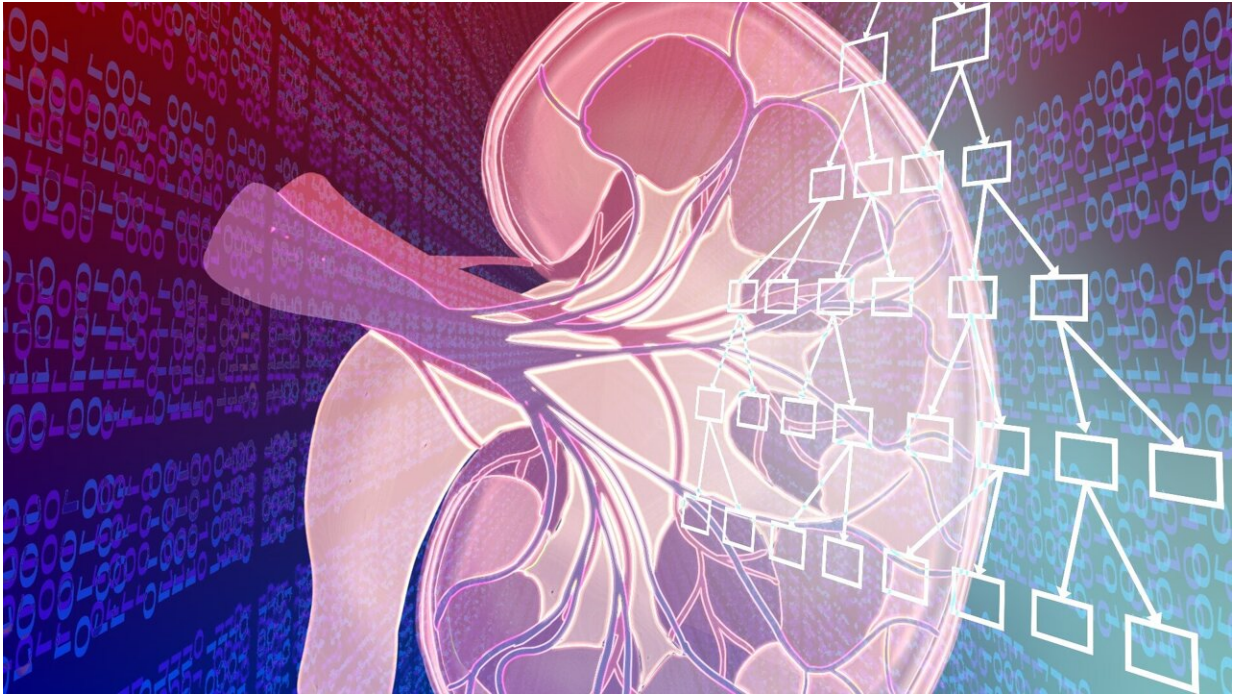# Giving AI penalties to get better diagnoses

November 1 2021

Telling sick people that they're healthy, can happen when a human doctor sees a patient. It also happens when Artificial Intelligence (AI) learns to diagnose disease. But giving a big penalty to an algorithm for false negatives results in much better precision, UJ researchers find.The research appears in Informatics in Medicine Unlocked, at https://doi.org/10.1016/j.imu.2021.100690. Credit: Graphic design by Therese van Wyk, University of Johannesburg. Based on Pixabay images.

Anyone waiting for the results of a medical test knows the anxious question: "Will my life change completely when I know?" And the relief

if you test negative.

Today, artificial intelligence (AI) is increasingly deployed to predict life-threatening diseases. But there remains a big challenge in getting the machine learning (ML) algorithms to be precise enough—specifically, in getting the algorithms to correctly diagnose if someone is sick.

Machine learning (ML) is the branch of AI where algorithms learn from datasets and get smarter in the process. "Let's say there is a dataset about a serious disease. The dataset has 90 people who do not have the disease. But 10 of the people do have the disease," says Dr. Ibomoiye Domor Mienye. Mienye is a post-doctoral AI researcher at the University of Johannesburg (UJ).

"As an example, an ML algorithm says that the 90 do not have the disease. That is correct so far. But it fails to diagnose the 10 that do have the disease. The algorithm is still regarded as 90% accurate," he says.

This is because accuracy has been defined in this way. But for health outcomes, it may be urgent to diagnose the 10 people with the disease and get them into treatment. That may be more important than complete accuracy about the 90 who do not have the condition, he adds.

Penalties against AI

In a research study published in *Informatics in Medicine Unlocked*, Mienye and Prof Yanxia Sun show how ML algorithms can be improved significantly for medical purposes. They used logistic regression, decision tree, XGBoost, and random forest algorithms.

These are supervised binary classification algorithms. That means they only learn from the 'yes/no' datasets provided to them.

Dr. Mienye and Prof Sun are both from the Department of Electrical and Engineering Science at UJ. The researchers built cost sensitivity into each of the algorithms. This means the algorithm gets a much bigger penalty for telling a sick person in the dataset that they are healthy, than the other way round. In medical terms, the algorithms get bigger penalties for false negatives than for false positives.

Dr. Mienye and Prof Sun used public learning datasets for diabetes, breast cancer, cervical cancer (858 records) and chronic kidney disease (400 records).

The datasets come from large hospitals or healthcare programs. In these binary datasets, people are classified as either having a disease, or not having it at all.

The algorithms they used are binary also. These can say "yes the person has the disease" or "no they don't have it." They tested all the algorithms on each dataset, both without and with the cost-sensitivity.

Significantly improved precision and recall

The results make it clear that the penalties work as intended in these datasets. For chronic kidney disease for example, the Random Forest algorithm had precision at 0.972 and recall at 0.946, out of a perfect 1.000. After the cost-sensitivity was added, the algorithm improved significantly to precision at 0.990 and recall at a perfect 1.000.

For CKD, the three other algorithms' recall improved from high scores to a perfect 1.000. Precision at 1.000 means the algorithm did not predict one or more false positives across the entire dataset. Recall at 1.000 means the algorithm did not predict one or more false negatives across the entire dataset.

With the other datasets, the results were different for different algorithms. For [cervical cancer](#), the cost-sensitive random forest and XGBoost algorithms improved from high scores to perfect precision and recall. However, the [logistic regression](#) and decision tree algorithms improved to much higher scores but did not reach 1.000.

The precision problem

In general, algorithms have been more accurate at saying people do not have a disease than identifying the ones who are sick, says Mienye. This is an ongoing challenge in healthcare AI.

The reason is the way the algorithms learn. The algorithms learn from datasets that come from large hospitals or state healthcare programs. But most of the people in those datasets do not have the conditions they are being tested for, says Mienye. "At a large hospital, a person comes in to get tested for chronic kidney disease (CKD). Their doctor sent them there because some of their symptoms are CKD symptoms. The doctor would like to rule out CKD. Turns out, the person does not have CKD.

"This happens with lots of people. The dataset ends up with more people who do not have CKD, than people who do. We call this an imbalanced dataset."

When an algorithm starts learning from the dataset, it learns far less about CKD than it should, and isn't accurate enough in diagnosing ill patients—unless the algorithm is adjusted for the imbalance.

AI on the other side of a boat ride

Mienye grew up in a village near the Atlantic Ocean, that is not accessible by road. "You have to use a speedboat from the nearest town

to get there. The boat ride takes two to three hours," he says. The nearest clinic is in the bigger town, on the other side of the boat ride. The deep rural setting of his home village inspired him to see how AI can help people with little or no access to healthcare.

An old lady from his village is a good example of how more advanced AI algorithms may assist in future, he says. A cost-sensitive multiclass ML algorithm could assess the measured data for her blood pressure, sodium levels, blood sugar and more.

If her data is recorded correctly on a computer, and the algorithm learns from a multiclass dataset, that future AI could tell clinic staff which stage of chronic kidney disease she is at. This village scenario is in the future, however.

Meanwhile the study's four algorithms with cost sensitivity, are far more precise at diagnosing disease in their numerical datasets. And they learn quickly, using the ordinary computer that one could expect to find in a remote town.

More information: Ibomoiye Domor Mienye et al, Performance analysis of cost-sensitive learning methods with application to imbalanced medical data, *Informatics in Medicine Unlocked* (2021). DOI: 10.1016/j.imu.2021.100690

Provided by University of Johannesburg