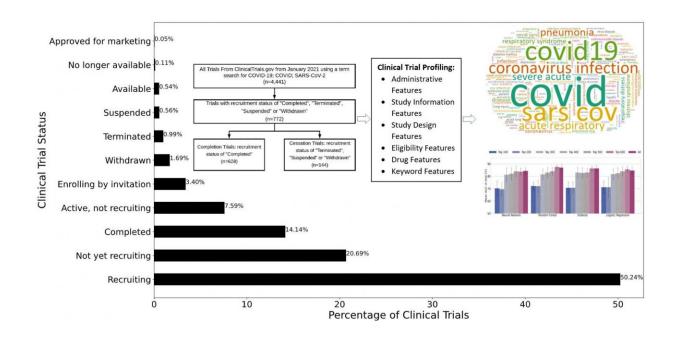# Predicting if COVID-19 clinical trials will fail or succeed

July 21 2021



As of July 15, more than 6,180 COVID-19 clinical trials have been registered through ClinicalTrials.gov. Credit: Florida Atlantic University/College of Engineering and Computer Science

In order to win the battle against COVID-19, studies to develop vaccines, drugs, devices and re-purposed drugs are urgently needed. Randomized clinical trials are used to provide evidence of safety and efficacy as well as to better understand this novel and evolving virus. As of July 15, more than 6,180 COVID-19 clinical trials have been registered through ClinicalTrials.gov, the national registry and database

for privately and publicly funded clinical studies conducted around the world. Knowing which ones are likely to succeed is imperative.

Researchers from Florida Atlantic University's College of Engineering and Computer Science are the first to model COVID-19 completion versus cessation in clinical trials using machine learning algorithms and ensemble learning. The study, published in *PLOS ONE*, provides the most extensive set of features for clinical trial reports, including features to model trial administration, study information and design, eligibility, keywords, drugs and other features.

This research shows that computational methods can deliver effective models to understand the difference between completed vs. ceased COVID-19 trials. In addition, these models also can predict COVID-19 trial status with satisfactory accuracy.

Because COVID-19 is a relatively novel disease, very few trials have been formally terminated. Therefore, for the study, researchers considered three types of trials as cessation trials: terminated, withdrawn, and suspended. These trials represent research efforts that have been stopped/halted for particular reasons and represent research efforts and resources that were not successful.

"The main purpose of our research was to predict whether a COVID-19 clinical trial will be completed or terminated, withdrawn or suspended. Clinical trials involve a great deal of resources and time including planning and recruiting human subjects," said Xingquan 'Hill' Zhu, Ph.D., senior author and a professor in the Department of Computer and Electrical Engineering and Computer Science, who conducted the research with first author Magdalyn 'Maggie' Elkin, a second-year Ph.D. student in computer science who also works full-time. "If we can predict the likelihood of whether a trial might be terminated or not down the road, it will help stakeholders better plan their resources and procedures.

Eventually, such computational approaches may help our society save time and sources to combat the global COVID-19 pandemic."

For the study, Zhu and Elkin collected 4,441 COVID-19 trials from ClinicalTrials.gov to build a testbed. They designed four types of features (statistics features, keyword features, drug features and embedding features) to characterize clinical trial administration, eligibility, study information, criteria, drug types, study keywords, as well as embedding features commonly used in state-of-the-art machine learning. In total, 693 dimensional features were created to represent each clinical trial. For comparison purposes, researchers used four models: Neural Network; Random Forest; XGBoost; and Logistic Regression.

Feature selection and ranking showed that keyword features derived from the MeSH (medical subject headings) terms of the clinical trial reports, were the most informative for COVID-19 trial prediction, followed by drug features, statistics features and embedding features. Although drug features and study keywords were the most informative features, all four types of features are essential for accurate trial prediction.

By using ensemble learning and sampling, the model used in this study achieved more than 0.87 areas under the curve (AUC) scores and more than 0.81 balanced accuracy for prediction, indicating high efficacy of using computational methods for COVID-19 clinical trial prediction. Results also showed single models with balanced accuracy as high as 70 percent and an F1-score of 50.49 percent, suggesting that modeling clinical trials is best when segregating research areas or diseases.

"Clinical trials that have stopped for various reasons are costly and often represent a tremendous loss of resources. As future outbreaks of COVID-19 are likely even after the current pandemic has declined, it is

critical to optimize efficient research efforts," said Stella Batalama, Ph.D., dean, College of Engineering and Computer Science. "Machine learning and AI driven computational approaches have been developed for COVID-19 health care applications, and deep learning techniques have been applied to medical imaging processing in order to predict outbreak, track virus spread and for COVID-19 diagnosis and treatment. The new approach developed by professor Zhu and Elkin will be helpful to design computational approaches to predict whether or not a COVID-19 clinical trial will be completed so that stakeholders can leverage the predictions to plan resources, reduce costs, and minimize the time of the clinical study."

**More information:** Magdalyn E. Elkin et al, Understanding and predicting COVID-19 clinical trial completion vs. cessation, *PLOS ONE* (2021). DOI: 10.1371/journal.pone.0253789

Provided by Florida Atlantic University