

# A better approach to disease prediction through big data analytics

July 17 2017

---

Big data holds great promise to change health care for the better. However, much of the technology that will someday transform health care and its delivery is not yet mature enough for hospitals and other systems to use.

The [Second IEEE/ACM Conference on Connected Health: Applications, Systems and Engineering Technologies](#) will bring experts from academics, business and government together to share information and help accelerate [health care](#)'s transformation. This leading international conference will take place in Philadelphia this week from July 17—19.

Mooi Choo Chuah, professor of computer science and engineering at Lehigh University and co-director of Lehigh's undergraduate computer engineering program, is serving as technical co-chair, along with Professor Insup Lee of the University of Pennsylvania. Chuah is a top expert in next generation wireless network architecture design, network and Smart Grid security, and mobile/cloud computing related research. Recently, she also started to do some healthcare [data](#) mining research.

In addition to co-leading the technical program committee charged with planning and implementing the conference's content, Chuah will present a paper on Tuesday, July 18th called "Incentivizing High Quality Crowdsourcing Clinical Data for Disease Prediction"

According to Chuah, her group's latest research offers two contributions. The first is an approach she developed with her graduate student

collaborator Qinghan Xue that uses a large dataset to demonstrate an improved disease prediction model that combines data cleaning and careful feature selection with effective machine learning techniques.

Chuah utilized a dataset made public by the non-profit [Prize4Life](#), which partnered to develop the Pooled Resource Open-Access ALS Clinical Trials (PRO-ACT) data base, the largest database of clinical data from Amyotrophic Lateral Sclerosis (ALS) patients ever created. In 2012, Prize4Life held a crowdsourced competition to create a method to accurately predict ALS disease outcomes based on PRO-ACT dataset.

Among the outcomes the participating teams sought to predict were which patients with ALS—a progressive degenerative nerve disease—would experience a slowly-progressing disease, which an average-progressing disease and which a fast-progressing disease. The challenge also asked researchers to predict how long ALS patients would survive from the date of diagnosis. Two teams won the top awards for these two different prediction tasks.

Similar to the crowdsourced competition, Chuah used the PRO-ACT database (which contains more than 10,700 records with 6,318 features) to predict which patients would fall into the three clusters of progression: slow, average or fast.

The challenge, says Chuah, was that the dataset was "very noisy."

"For example, some data were missing," says Chuah. "Some data were non-numeric—and, as you know, computers like numeric values."

Their model cleaned up the data and demonstrated an improved accuracy rate in predicting a patient's disease progression. In fact, Chuah's method performed better than the winning team's did—at 58.3% accuracy compared to 40.5%—and with fewer required features

and higher quality data.

"We were able to predict where a patient would fall on the disease progression spectrum with more accuracy and faster," says Chuah. "This has both cost-saving implications—as a physician might see a patient with a faster-progressing disease more frequently, but less frequently for slow-progressing patients—as well as for improved [health](#) outcomes."

The paper's second contribution presents a solution to one of the major challenges of healthcare: the fact that no single hospital or health care system has enough of their own data for useful predictive disease analysis.

"Hospitals and other health care systems collect troves of data," explains Chuah. "However, each has a limited number of patients experiencing a particular disease—such as ALS or diabetes, for example. We have designed an incentive method to encourage hospitals to share data so that better prediction models can be created."

The algorithm that she and her team developed is designed to provide a "reward function" for each health care provider, identifying the cost per patient to participate in a crowdsourced database. An individual hospital would be able to use the incentive model to evaluate whether to participate. The model provides a "reward" for offering truthful, high-quality data.

Chuah believes that both elements of her latest research could positively impact the accuracy and usefulness of predictive [disease](#) models and, most importantly, improve health outcomes for [patients](#).

She adds: "In my work, I'm always looking to solve problems that I know will have some kind of positive social impact."

Provided by Lehigh University

Citation: A better approach to disease prediction through big data analytics (2017, July 17)  
retrieved 26 April 2024 from <https://phys.org/news/2017-07-approach-disease-big-analytics.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.