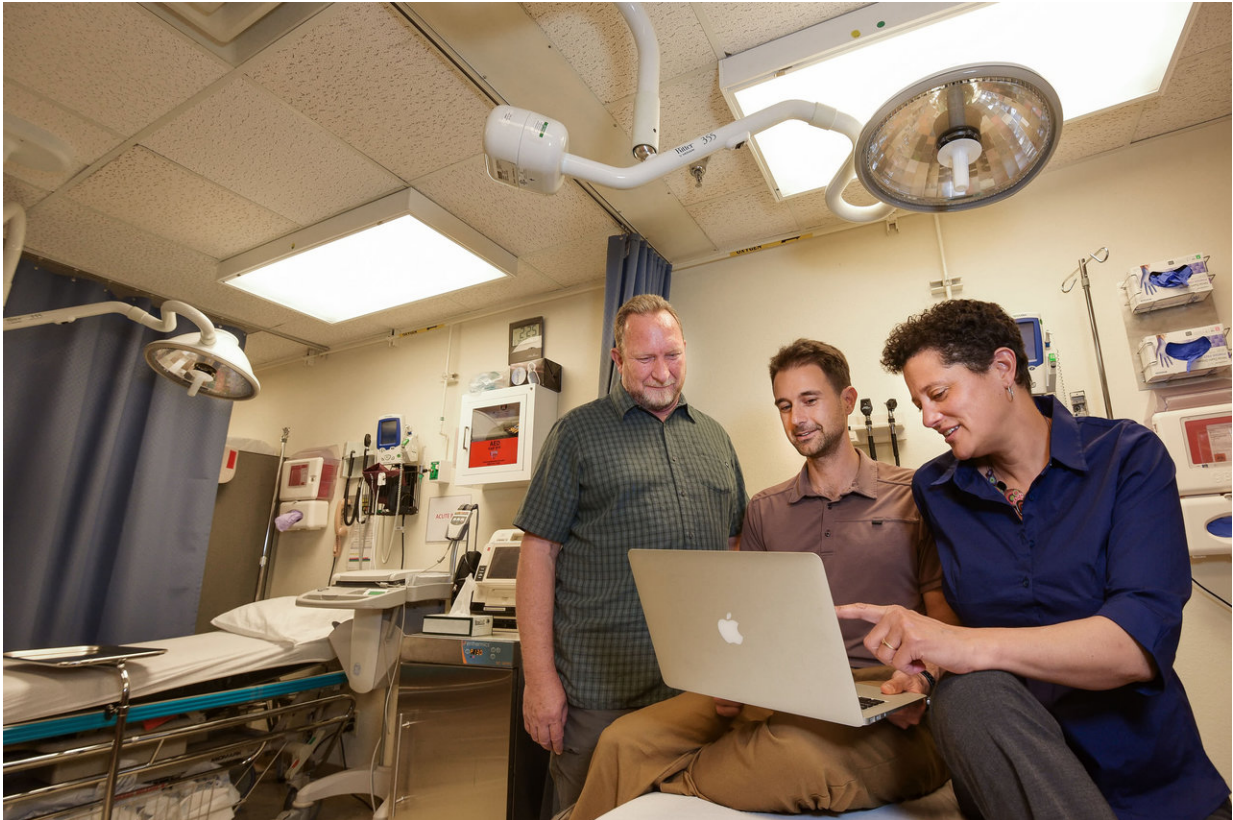# Using biomimicry to detect outbreaks faster

April 9 2018

Sandia National Laboratories computer scientists Pat Finley, left; and Drew Levin, center; and University of New Mexico Professor Melanie Moses have taken inspiration from the immune system to improve how the U.S. detects emerging outbreaks. Levin started on the project as a UNM graduate student and after he graduated he was hired by Sandia to continue working on the project. Credit: Randy Montoya

Our immune systems are made up of billions of white blood cells searching for signs of infections and foreign invaders, ready to raise the alarm.

Sandia National Laboratories computer scientists Pat Finley and Drew Levin have been working to improve the U.S. biosurveillance system that alerts authorities to disease outbreaks by mimicking the human immune system. They are working with researchers at the University of New Mexico and the Centers for Disease Control and Prevention.

The CDC coordinates the National Syndromic Surveillance Program. It collects anonymized data from most emergency departments around the nation and analyzes public health indicators to speed up the response to hazardous events and disease outbreaks.

"The national biosurveillance system serves essentially the same purpose as the human immune system, just on a larger scale," said Levin, who started working on the project as a UNM graduate student and was hired by Sandia to continue his work after he graduated. "The immune system is made up of numerous T-cells that all operate independently. There's no centralized controller and yet we do pretty well not dying."

The CDC uses traditional statistical analyses to look for anomalies, such as a large or sudden increase in ER visits, and determine the likelihood of an outbreak. These algorithms are based on reliable, decades-old math but usually only look at one variable at a time, said Levin.

The faster an emerging outbreak is detected, the more lives are saved; however, flagging non-outbreaks can waste resources.

Finley said the biosurveillance system has the dual challenge of detecting new outbreaks of old diseases, such as seasonal influenza, as well as outbreaks of new diseases, such as the next Zika virus, a very tough

problem. However, it is a problem the immune system has been working on for millions of years.

Synthetic T-cells monitor multiple variables for nuanced alerts

T-cells are a type of white blood cell that recognize and kill virus-infected cells and other foreign pathogens. They recognize the foreign invaders after undergoing a negative-selection "training" process where every T-cell that attacks normal body cells is destroyed. Other than this initial "training," there's no central "brain" controlling the T-cells.

Finley thought that mimicking how T-cells work might speed up outbreak detection. In 2015, he began collaborating with immune system modeling experts at UNM as part of Sandia's Academic Alliance program. The Academic Alliance is a partnership Sandia has built with five universities to promote collaborative research on tough problems and attract top talent to work on these challenges.

"The adaptive immune system in vertebrates is one of the most complex systems in biology with trillions of cells, dozens of cell types and signaling molecules," said Melanie Moses, professor of computer science and biology at UNM involved in the project. "Through computer modeling and simulation, we understand how the immune system works which, in the long term, can lead to improved immuno-therapies, allergy treatments and vaccines. It also provides inspiration for the design of other decentralized systems for surveillance and protection."

Working together, the team created synthetic, mathematical "T-cells" that look at multiple different variables at the same time, such as number of clinic visits, day of the year and intake temperature. Then, mimicking the T-cell negative selection process, Levin ran the synthetic T-cell

algorithms against past data collected by the CDC and New Mexico Department of Health. He compared the algorithms and selected the most accurate.

In 2016, initial tests on a pilot-scale biosurveillance system showed that Levin's synthetic T-cells performed better than the traditional statistical methods, said Finley. Also, because the synthetic T-cells track multiple variables intrinsically, they could provide more nuanced alerts, such as separating an outbreak of a new disease from seasonal influenza, he said.

Brain-inspired machine learning improves chief complaint deciphering

The first piece of data the CDC receives from each emergency room visit is called the chief complaint. This is a concise statement describing why a patient has gone to the emergency room or clinic, before they've seen a doctor and have been diagnosed. Chief complaints range from "chest pain" and "fever three days" to specialized abbreviations.

These terse statements are full of medical jargon and even misspelled words, making them difficult to decipher by simple keyword searches or by the inexperienced. Also, many words describe the same symptoms, such as fever, hot, temperature and chills.

Technology companies have been using deep learning for similar natural-language processing problems. Deep learning is brain-inspired machine learning that excels at finding patterns without being explicitly programmed on what to look for. One such algorithm, called Word2vec, converts the context of words into mathematical vectors.

When Levin ran the Word2vec algorithm on anonymized chief complaint data collected by the New Mexico Department of Health, it

out-performed a standard keyword search, as well as other state-of-the-art machine learning algorithms. However, it still had troubles with misspelled words and abbreviations.

To work around this, Levin tried two related neural network algorithms: one that converts letters into vectors and another that converts words into random vectors. The algorithm that converted words into random, or untrained, vectors was most accurate, possibly because the trained Word2vec algorithm places antonyms too close together in vector space, said Levin.

Though more optimization is needed, the team's deep-learning algorithm for deciphering chief complaints could be particularly useful for the opioid epidemic, said Finley. He added, "New terms for street drugs tend to appear much more quickly than the public health community realizes. If we find that a weird word is popping up a lot in an area, it could be a new variety of fentanyl."

Future of distributed biosurveillance centers

Lymph nodes are distributed throughout the body and act as immune system hubs, chock full of T-cells and the B-cells that produce antibodies to fight off infections.

Finley and his team are just beginning to explore how mimicking lymph nodes might improve the biosurveillance system. Finley believes it would be particularly helpful in detecting outbreaks of regional diseases like Lyme disease, plague and Hantavirus. Also, distributed detection algorithms could be more efficient by bypassing the physical and power consumption limits that Moore's Law computers are now running up against, added Levin.

"We are working closely with the CDC to test a number of our deep

learning approaches on a subset of the national data flow," said Finley.

His goal is to have his biologically inspired system set up by October to allow side-by-side comparisons with the traditional statistical methods at the national scale. He believes the different approaches will have different strengths, and combining them will improve the speed and accuracy of outbreak detection.

This research was funded by Sandia's Laboratory Directed Research and Development program. Sandia computer scientists Walt Beyeler and Michael Mitchell and UNM postdoctoral fellow Tatiana Flanagan also worked on the project, focusing on the lymph system-mimicking distributed detection algorithm.

"This project with Sandia has provided us with an opportunity to test the practical application of the concepts we've learned from our models," said Moses. "Ultimately, this project will lead to a more complete understanding of the immune system, as well as a practical way to quickly identify and respond to disease outbreaks and other biological threats."

Provided by Sandia National Laboratories