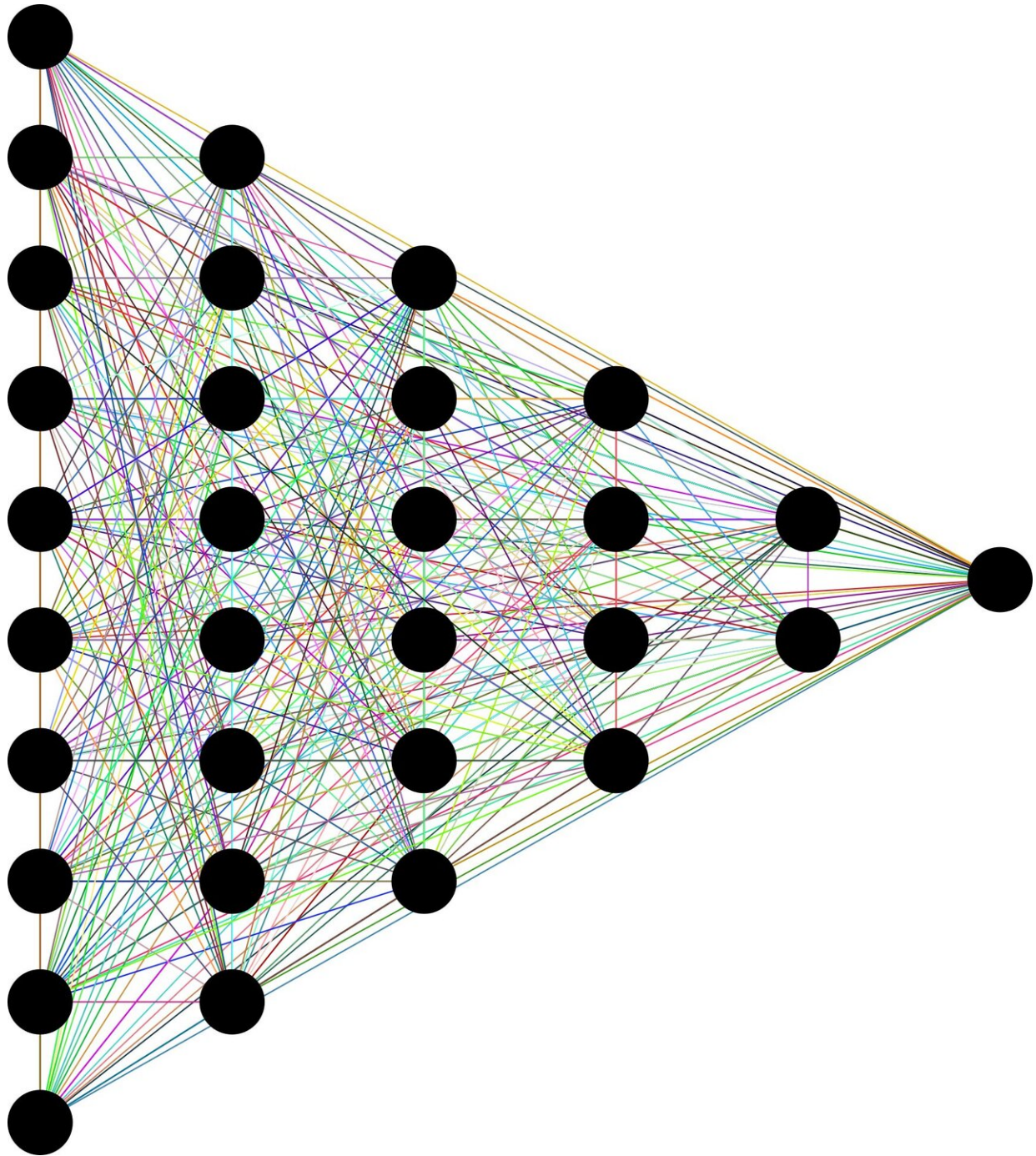# Machine learning research may aid many industries

December 4 2019, by Jacob T. Kerr

Credit: CC0 Public Domain

Spam emails, bank fraud, diabetes, workers quitting their jobs. What do

these topics have in common? The answer can be found in machine learning research at Binghamton University.

Dana Bani-Hani, a doctoral student studying industrial and systems engineering, has spent the past few years teaching machines how to read data sets in any industry. The system she coded, called a Recursive General Regression Neural Network Oracle (R-GRNN Oracle), takes data inputs and creates prediction outputs.

Regression models are not new in data science and analytics, but what Bani-Hani created goes beyond the basics. A typical system uses algorithms, called classifiers, that run through a data set of many different variables to create a prediction. Oracles are created to run multiple sets of these classifiers to see which algorithm creates the most accurate prediction.

For example, a classifier can look at myriad emails and factor in certain word usage, word count and several other variables to determine if the email is spam. An oracle looks at the different classifier outputs and determines which most accurately predicted the spam emails.

What sets the R-GRNN Oracle apart from other oracles is its capability to take classifier outputs and rank them based on their accuracy. Based on the ranking, classifiers are given weights and are combined to produce a prediction superior to any one classifier on its own.

Think of this process like an orchestra. Each instrument has its own strengths, just like different classifiers, so it is useful to include them all. The conductor, like the R-GRNN Oracle, directs the different instruments to play loudly or more softly based on how the instrument makes the final symphony sound.

At this point, the system would be called a General Regression Neural

Network (GRNN), which has been created before at Binghamton University. The real crux of Bani-Hani's work lies in the first letter, R, standing for Recursion.

The R-GRNN Oracle takes the original GRNN output, and uses that entire system as an input for another GRNN prediction. This is combined with the most successful of the original classifiers.

So, back to the orchestra: The original symphony is recorded, and then played back again later. This time, along with the recording, a few instruments play again to further fine-tune the important sounds of the orchestra.

"Because of the way [the GRNN] works, I was able to create the recursive model," Bani-Hani says. "The concept of recursion is not widely used in machine learning, so I decided to put an oracle inside of an oracle."

Mohammad Khasawneh, professor and department chair in systems science and industrial engineering, supervised Bani-Hani's research. He says systems like the GRNN and R-GRNN are underutilized and are vital in serious life events.

"The traditional GRNN Oracle has received limited attention in the literature as only very few researchers have published work on the algorithm," Khasawneh says. "But many real-life problems that apply machine learning models to automate classifying unknown observations require accurate predictions. Tasks such as diagnosing diseases entail precision to avoid serious issues that could potentially lead to problems such as lawsuits or even deaths."

Bani-Hani says the R-GRNN Oracle produces more accurate predictions than any single classifier alone, as well as one GRNN on its own. The R-

GRNN Oracle took in thousands of email samples, programmed to factor 57 variables, and then produced a spam prediction superior to all other classifiers tested.

Bani-Hani also used the R-GRNN to predict credit card application fraud, diabetes diagnosis and whether a worker will quit based on past workplace experiences. In each case, the R-GRNN came out as the most accurate predictor.

She plans to focus her model on specific fields, such as business or finance, as well as package both the GRNN Oracle and the R-GRNN Oracle so companies do not have to create the entire code from scratch.

Bani-Hani's journey to machine learning research started nearly 6,000 miles away from Binghamton in Jordan. After completing her bachelor's degree in architectural engineering, she heard about Binghamton University through Watson School faculty and academic leaders, and from her father's supportive suggestions. She initially pursued a master's degree in industrial engineering, but she soon found a new passion: data mining and machine learning.

"Getting a Ph.D. has been a dream of mine for the last 15 years," Bani-Hani says. "I mainly attribute this to having a family with advanced degrees. I am thankful to my professors here at Binghamton University for introducing me to the topics that make up my research."

Provided by Binghamton University

study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.