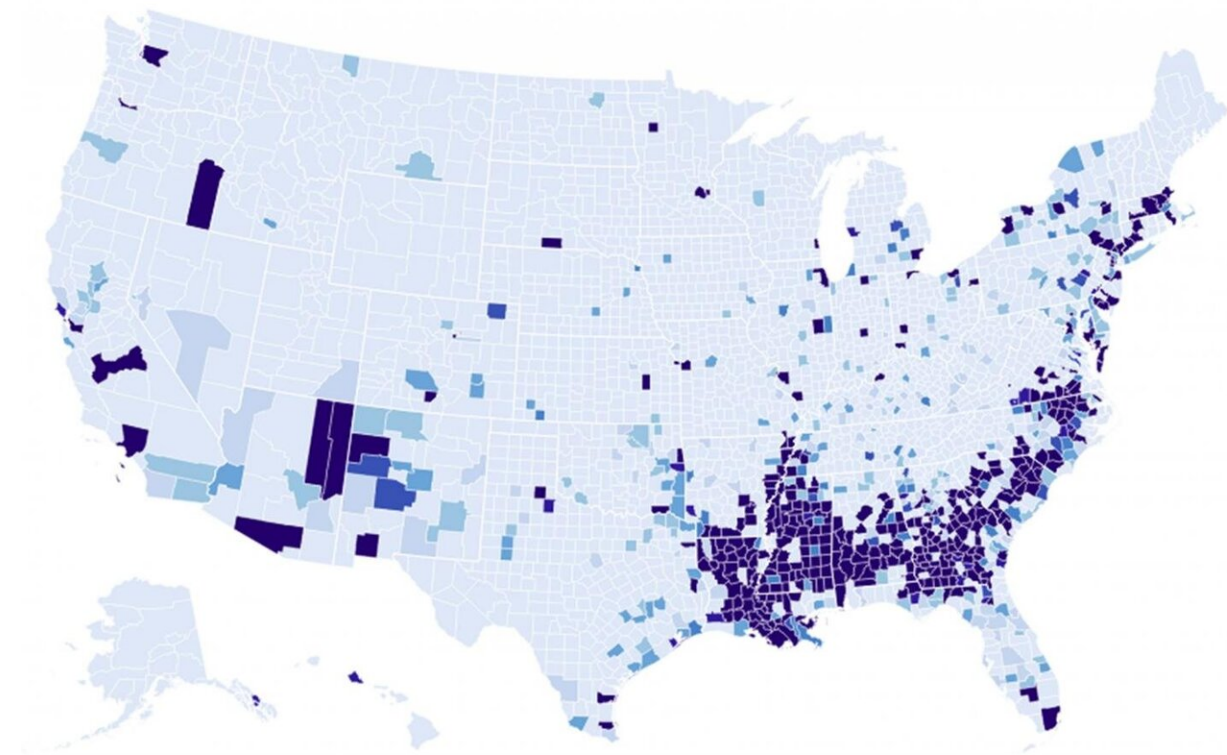# Data mining system unearths US counties most at risk for COVID deaths

July 8 2020

This "alert map" indicates the number of times a US county is part of a 'high risk' set. Darker colors indicate a higher level of risk a county has for high COVID-19 death rates. Credit: Akai Kaeru, LLC

The task of controlling the COVID-19 pandemic nationwide and predicting where cases will spike next and which areas may have high

mortality rates remains daunting for scientists and public officials. A new machine learning tool developed by researchers at a startup company (Akai Kaeru LLC) affiliated with Stony Brook University's Department of Computer Science and the Institute for Advanced Computational Science (IACS)may help gauge areas most at risk for the virus and high death rates. The software they use analyzes a massive data set from all 3,007 U.S. counties. They found that combinations of factors such as poverty, rural settings, low education, low poverty but housing debt, and sleep deprivation are associated with higher death rates in counties.

The researchers use an automatic pattern mining engine and software to analyze a data set with approximately 500 attributes, which cover details related to demographics, economics, race and ethnicity, and infrastructure in all U.S. counties. After analyzing and assessing the data within counties they created nearly 300 sets of counties at a "high risk" for COVID-19 and related death rates.

Many of these counties within the sets—but not all—are in Southern U.S. states and include close to 1,000 counties. Some of the counties include Hancock, Ga.; Attala, Miss.; Lee, S.C.; Swisher Texas; Adams, Ohio; Torrance, N.M.; and Madison, Fla. Mississippi, Louisiana and Georgia are the most at risk, with 80-90 percent of their counties covered by these sets.

"Our software algorithm identifies counties with specific conditions that appear to lead to higher than average U.S. death rates due to COVID-19," said Klaus Mueller, Ph.D., Professor of Computer Science, IACS faculty member, CEO of startup Akai Kaeru, LLC, and Principal Investigator of the company study. "We cannot say that a specific county will have a higher than usual death rate, but we can predict this for the sets of counties that fit certain conditions."

According to Mueller, the software and method used to analyze the data and identify high-risk counties can help inform officials based on important correlations related to COVID-19 death rates and help direct allocation of resources, such as testing kits and stations. The method and findings may also help to target community-based information campaigns about COVID-19 and measures to contain the pandemic and potentially reduce cases.

The researchers found that several conditions must be present at the same time to expose a county to elevated risk. Some of these condition sets are:

- Poor rural counties with aging residents.
- Sleep-deprived, under-educated counties with low participation in health insurance.
- Counties with low Asian but high minority populations where black children live in poverty.
- Counties with high home ownership and low poverty. For this set of counties there also exists a significant correlation between death rate and the amount of housing debt the county residents have.

"Each of these sets of conditions tells a unique story and makes the Artificial Intelligence behind our algorithm explainable." Mueller says. "For instance, what we might conclude from the 'high home ownership and low poverty' pattern is that there are homeowners in these wealthy counties with high home ownership who cannot afford their homes and as a result run high housing debt. Then, as the percentage of these types of homeowners in a county grows, so does the risk of COVID-19 infection and potentially death."

"We also observe in a different county set that poor and aging counties with low population density are on average especially hard hit by

COVID-19," explains Mueller. "While it is well known now that older residents are more vulnerable to COVID-19, the pattern tells us that this high risk seems to be amplified by two factors related to accessibility:

(1) The residents live in sparsely populated areas which offer fewer urgent care facilities and (2) the residents are mostly poor which hampers their ability to use and pay for these services."

Mueller emphasizes that any conclusions about conditions related to high death rates from COVID-19 in county sets or specific counties will continue to need further investigation because a pandemic is not static and factors contributing to disease and death are often complicated.

Akai Kaeru is a start-up company developed and located in the New York State Center of Excellence in Wireless and Information Technology (CEWIT). Created in 2003, CEWIT is the anchoring building to Stony Brook University's Research and Development Park to conduct research and commercialize it.

More information: The entire high-risk county sets analysis can be viewed in more detail on this website: akaikaeru.com/covid-19-1

Provided by Stony Brook University