Researchers develop new technique to keep drinking water safe using machine learning

September 6 2022

Credit: CC0 Public Domain

Waterborne illness is one of the leading causes of infectious disease outbreaks in refugee and internally displaced persons (IDP) settlements, but a team led by York University has developed a new technique to keep drinking water safe using machine learning, and it could be a game

changer. The research is published in the journal *PLOS Water*.

As drinking water is not piped into homes in most settlements, residents instead collect it from public tap stands using storage containers.

"When water is stored in a container in a dwelling it is at high risk of being exposed to contaminants, so it's imperative there is enough free residual chlorine to kill any pathogens," says Lassonde School of Engineering Ph.D. student Michael De Santi, who is part of York's Dahdaleh Institute for Global Health Research, and who led the research.

Recontamination of previously [safe drinking water](#) during its collection, transport and storage has been a major factor in outbreaks of cholera, hepatitis E, and shigellosis in refugee and IDP settlements in Kenya, Malawi, Sudan, South Sudan, and Uganda.

"A variety of factors can affect chlorine decay in stored water. You can have [safe water](#) at that collection point, but once you bring it home and store it, sometimes up to 24 hours, you can lose that residual chlorine, pathogens can thrive and illness can spread," says Lassonde Adjunct Professor Syed Imran Ali, a Research Fellow at York's Dahdaleh Institute for Global Health Research, who has firsthand experience working in a settlement in South Sudan.

Using machine learning, the research team—including Associate Professor Usman Khan, also of Lassonde—has developed a new way to predict the probability that enough chlorine will remain until the last glass is consumed. They used an artificial neural network (ANN) along with ensemble forecasting systems (EFS), something that is not typically done. EFS is a probabilistic model commonly used to predict the probability of precipitation in weather forecasts.

"ANN-EFS can generate forecasts at the time of consumption that take a

variety of factors into consideration that affect the level of residual chlorine, unlike the typically used models. This new probabilistic modeling is replacing the currently used universal guideline for chlorine use, which has been shown to be ineffective," says Ali.

Factors such as local temperature, how the water is stored and handled from home to home, the type and quality of the water pipes, water quality and whether a child dipped their hand in the water container can all play a role in how safe the water is to drink.

"However, it's really important that these probabilistic models be trained on data at a specific settlement as each one is as unique as a snowflake," says De Santi. "Two people could collect the same water on the same day, both store it for six hours, and one could still have all the chlorine remaining in the water and the other could have almost none of it left. Another 10 people could have varying ranges of chlorine."

The researchers used routine water quality monitoring data from two refugee settlements in Bangladesh and Tanzania collected through the Safe Water Optimization Tool Project. In Bangladesh, the data was collected from 2,130 samples by Médecins Sans Frontières from Camp 1 of the Kutupalong-Balukhali Extension Site, Cox's Bazaar between June and December 2019 when it hosted 83,000 Rohingya refugees from neighboring Myanmar.

Determining how to teach the ANN-EFS to come up with realistic probability forecasts with the smallest possible error required out-of-the-box thinking.

"How that error is measured is key as it determines how the model behaves in the context of probabilistic modeling," says De Santi. "Using cost-sensitive learning, a tool that morphs the cost function towards a targeted behavior when using machine learning, we found it could

improve probabilistic forecasts and reliability. We are not aware of this being done before in this context."

For example, this model can say that under certain conditions at the tap with a particular amount of free residual chlorine in the water, there is a 90 percent chance that the remaining chlorine in the stored water after 15 hours will be below the safety level for drinking.

"That's the kind of probabilistic determination this modeling can give us," says De Santi. "Like with weather forecasts, if there is a 90 percent chance of rain, you should bring an umbrella. Instead of an umbrella, we can ask water operators to increase the chlorine concentration so there will be a greater percentage of people with safe drinking water."

"Our Safe Water Optimization Tool takes this machine learning work and makes it available to aid workers in the field. The only difference for the water operators is we ask them to sample water in the container at the tap and in that same container at the home after several hours," says Ali.

"This work Michael is doing is advancing the state of practice of machine learning models. Not only can this be used to ensure safe drinking water in refugee and IDP settlements, it can also be used in other applications."

Provided by York University