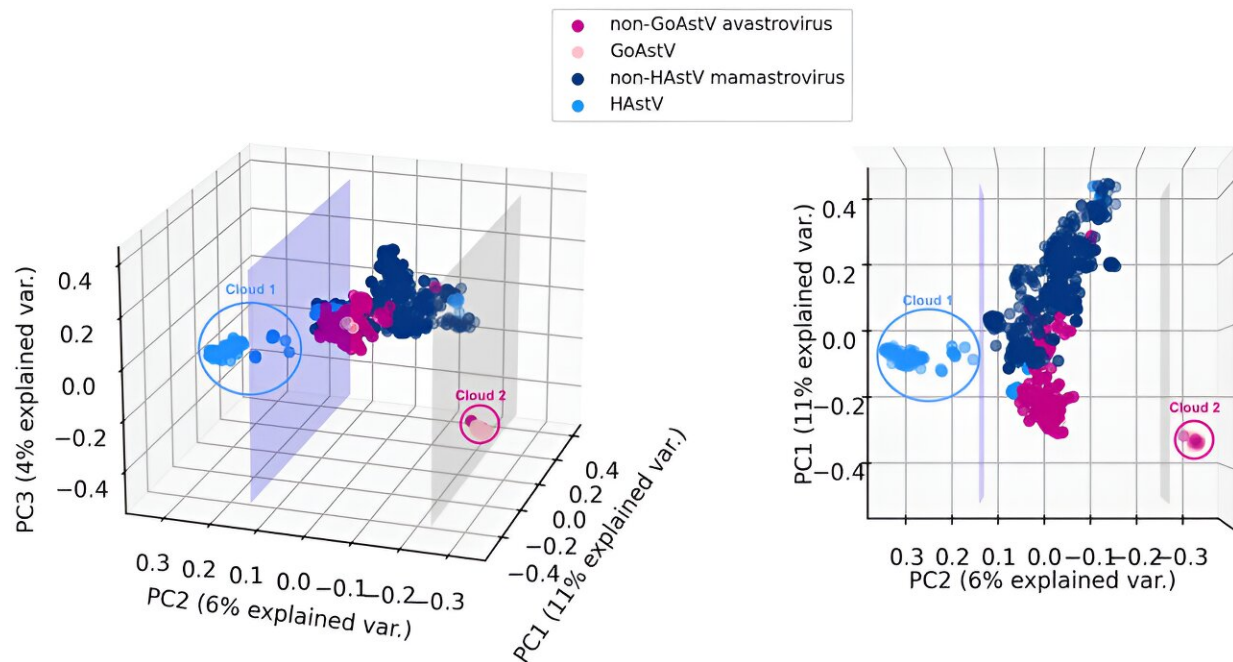


Machine learning classifies 191 of the world's most damaging viruses

April 29 2024



Two views of 3D PCA data visualizations of Mamastrovirus and Avastrovirus sequences k-mer frequencies: astrovirus sequences in Dataset 2 (known genus labels), together with the 191 astrovirus genomes with genus labels predicted by 3PCM. For comparison purposes, HAstV and GoAstV are highlighted with different colors compared to the rest of Mamastroviruses (non-HAstV Mamastroviruses) respectively the rest of the Avastroviruses (non-GoAstV Avastroviruses). The lavender plane illustrates the separation between two possible subgenera of Mamastrovirus. The gray plane illustrates the separation between two possible subgenera of Avastrovirus. Credit: *Frontiers in Molecular Biosciences* (2024). DOI: 10.3389/fmolb.2023.1305506

Researchers from the University of Waterloo have successfully classified 191 previously unidentified astroviruses using a new machine learning-enabled classification process.

The study, "Leveraging [machine learning](#) for taxonomic classification of emerging astroviruses," was recently [published](#) in *Frontiers in Molecular Biosciences*.

Astroviruses are some of the most damaging and widespread viruses in the world. These viruses cause severe diarrhea, which kills more than 440,000 children under the age of 5 annually. In the [poultry industry](#), astroviruses like [avian flu](#) have an 80% infection rate and a 50% mortality rate among livestock, leading to economic devastation, supply chain disruption, and food shortages.

Astroviruses mutate quickly and can spread easily across their more than 160 [host species](#), putting researchers and [public health officials](#) in a constant race to classify and understand new astroviruses as they emerge. In 2023, there were 322 unidentified astroviruses with distinct genomes. This year, that number has risen to 479.

"At any given point, between 2% and 9% of humans carry one of these viruses. That number can be as high as 30% in some countries," said Fatemeh Alipour, Ph.D. candidate in [computer science](#) at Waterloo and the lead computer science author of the research study. "Understanding and classifying these viruses effectively is essential for developing vaccines."

The astrovirus research team included computer science researchers at Waterloo and biology researchers at the University of Western Ontario.

The new three-part classification method includes supervised machine learning, unsupervised machine learning, and manual labeling of each

astrovirus's host.

"The main idea behind the classification method is to leverage machine learning to classify species by learning from their 'genomic signatures,'" said Lila Kari, professor in the David R. Cheriton School of Computer Science. "The classification method is exciting both in its speed and general applicability."

"This method can help us understand how viruses are transmitted between different animals. It can also be used to classify viruses in other virus families like HIV and Dengue."

More information: Fatemeh Alipour et al, Leveraging machine learning for taxonomic classification of emerging astroviruses, *Frontiers in Molecular Biosciences* (2024). [DOI: 10.3389/fmolb.2023.1305506](https://doi.org/10.3389/fmolb.2023.1305506)

Provided by University of Waterloo

Citation: Machine learning classifies 191 of the world's most damaging viruses (2024, April 29) retrieved 18 May 2024 from <https://phys.org/news/2024-04-machine-world-viruses.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.