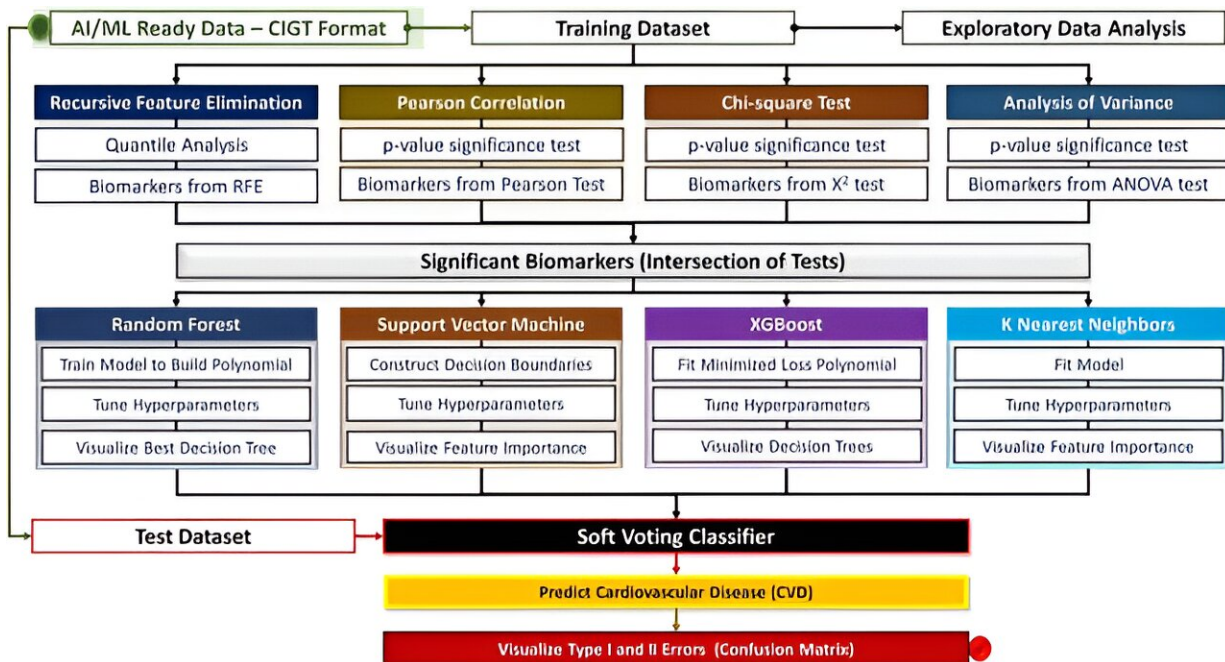# Researchers develop software to predict diseases

January 23 2024

Methodology and study design, workflow, and bioinformatics. This figure presents implemented statistical tests (Recursive Feature Elimination, Pearson correlation, Chi-square test, and Analysis of Variance) for the exploratory data analysis to assess the differences in genomics and phenotypic features between healthy individuals and patients with CVD and observe significant biomarkers. Next, applied a nexus of Machine Learning (ML) algorithms (Random Forest, Support Vector Machine, Xtreme Gradient Boosting Decision Trees, and k-Nearest Neighbors) to predict CVD. In addition, it includes Training Dataset, Test Dataset, Soft Voting Classifier, and Visualization of Type I and II errors. Credit: *Scientific Reports* (2024). DOI: 10.1038/s41598-023-50600-8

IntelliGenes, a first-of-its-kind software created at Rutgers Health, combines artificial intelligence (AI) and machine-learning approaches to measure the significance of specific genomic biomarkers to help predict diseases in individuals, according to its developers.

A study published in *Bioinformatics* explains how IntelliGenes can be utilized by a wide range of users to analyze multigenomic and clinical data.

Zeeshan Ahmed, lead author of the study and a faculty member at Rutgers Institute for Health, Health Care Policy and Aging Research (IFH), said there currently are no AI or machine-learning tools available to investigate and interpret the complete human genome, especially for nonexperts. Ahmed and members of his Rutgers lab designed IntelliGenes so anyone can use the platform, including students or those without strong knowledge of bioinformatics techniques or access to high-performing computers.

The software combines conventional statistical methods with cutting-edge machine learning algorithms to produce personalized patient predictions and a visual representation of the biomarkers significant to disease prediction.

In another study, published in *Scientific Reports*, the researchers applied IntelliGenes to discover novel biomarkers and predict cardiovascular disease with high accuracy.

"There is huge potential in the convergence of datasets and the staggering developments in artificial intelligence and machine learning," said Ahmed, who also is an assistant professor of medicine at Robert Wood Johnson Medical School.

"IntelliGenes can support personalized early detection of common and

rare diseases in individuals, as well as open avenues for broader research ultimately leading to new interventions and treatments."

Researchers tested the software using Amarel, the high-performance computing cluster managed by the Rutgers Office of Advanced Research Computing. The office provides a research computing and data environment for Rutgers researchers engaged in complex computational and data-intensive projects.

Co-authors of the study include William DeGroat, Dinesh Mendhe, Atharva Bhusari and Habiba Abdelhalim of IFH and Saman Zeeshan of Rutgers Cancer Institute of New Jersey.

 More information: William DeGroat et al, IntelliGenes: a novel machine learning pipeline for biomarker discovery and predictive analysis using multi-genomic profiles, *Bioinformatics* (2023). DOI: 10.1093/bioinformatics/btad755

William DeGroat et al, Discovering biomarkers associated and predicting cardiovascular disease with high accuracy using a novel nexus of machine learning techniques for precision medicine, *Scientific Reports* (2024). DOI: 10.1038/s41598-023-50600-8

Provided by Rutgers University