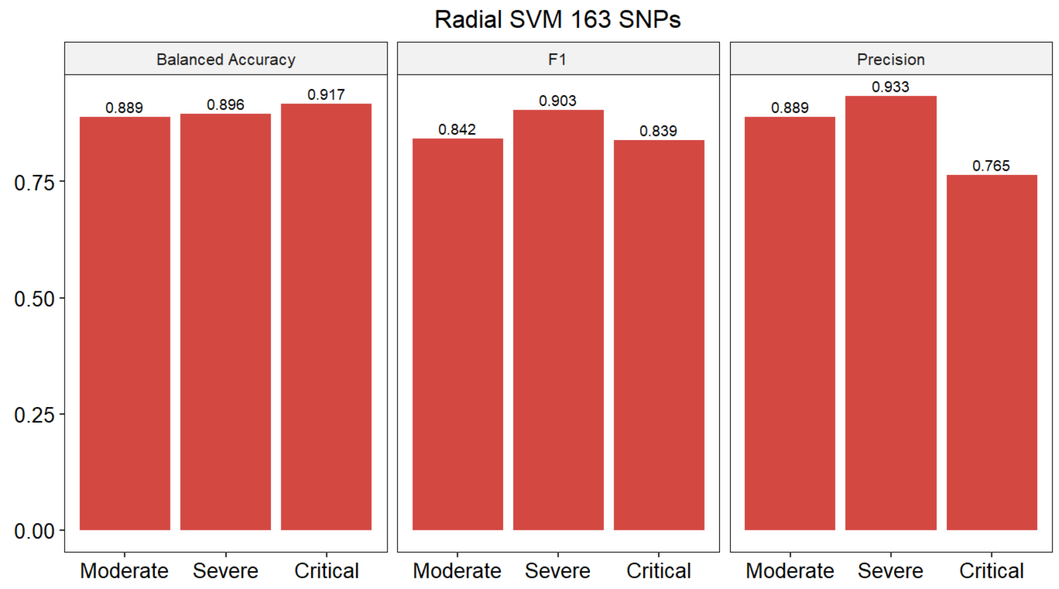
Predicting COVID-19 symptom severity based on host genetic predisposition

Doris Repušić[1*], Marina Korolija[2], Ana Livun[3], Tomo Svaguša[3], Robert Belužić[4], Vjekoslav Tomaić[4], Rosa Karlić[1], Kristian Vlahoviček[1]
[1] Bioinformatics Group, Division of Molecular Biology, Department of Biology, Faculty of Science, University of Zagreb, Horvatovac 102a, Zagreb, Croatia
[2] Forensic Science Centre "Ivan Vučetić", Ilica 335, Zagreb, Croatia
[3] Dubrava University Hospital, Avenija Gojka Šuška 6, Zagreb, Croatia
[4] Department of Molecular Medicine, Ruđer Bošković Institute, Bijenička 54, Zagreb, Croatia
dbrkic@bioinfo.hr

ML Model	Radial Support Vector Machine
Predictors	163 SNPs
Mean Train Accuracy (10-fold crossvalidation)	87.33%
Test Accuracy	87.5%
AUC	0.969

A support vector machine (SVM) transforms data into a higher-dimensional space to simplify finding a linear separation or better classify the data.



Balanced accuracy - the average between the sensitivity and the specificity,
F1 score - model's performance on each class individually,
Precision - the number of true positives divided by the total number of positive predictions

WE FOUND 163 SNPs THAT PREDICT WITH 87.5% ACCURACY HOW SEVERE YOUR COVID-19 SYMPTOMS COULD BE!

DATASET

Genomic data in this study was obtained by whole exome sequencing of 184 patients hospitalized with SARS-CoV-2 infection in Dubrava University Hospital. Based on their health record data, patients were classified into three severity groups Moderate, Severe, and Critical following the National Institute of Health's classification guidelines.

METHODS

Single nucleotide polymorphisms (SNPs) within genes previously related to the COVID-19 disease pathway were identified, and their allele frequencies within severity groups were calculated. SNPs with statistically significant differences in allele frequencies between symptom severity categories were used as predictors in machine-learning models, while symptom severity was used as the response variable we tried to predict.

CONCLUSIONS

- 163 SNPs in the Croatian population with significant allele frequency differences between symptom severity categories.
- High test accuracy (87,5%) of radial support vector machine model in predicting symptom severity based on 163 SNPs.
- Host genetic factors can be used to identify high-risk individuals for getting severe and critical COVID-19 symptoms.