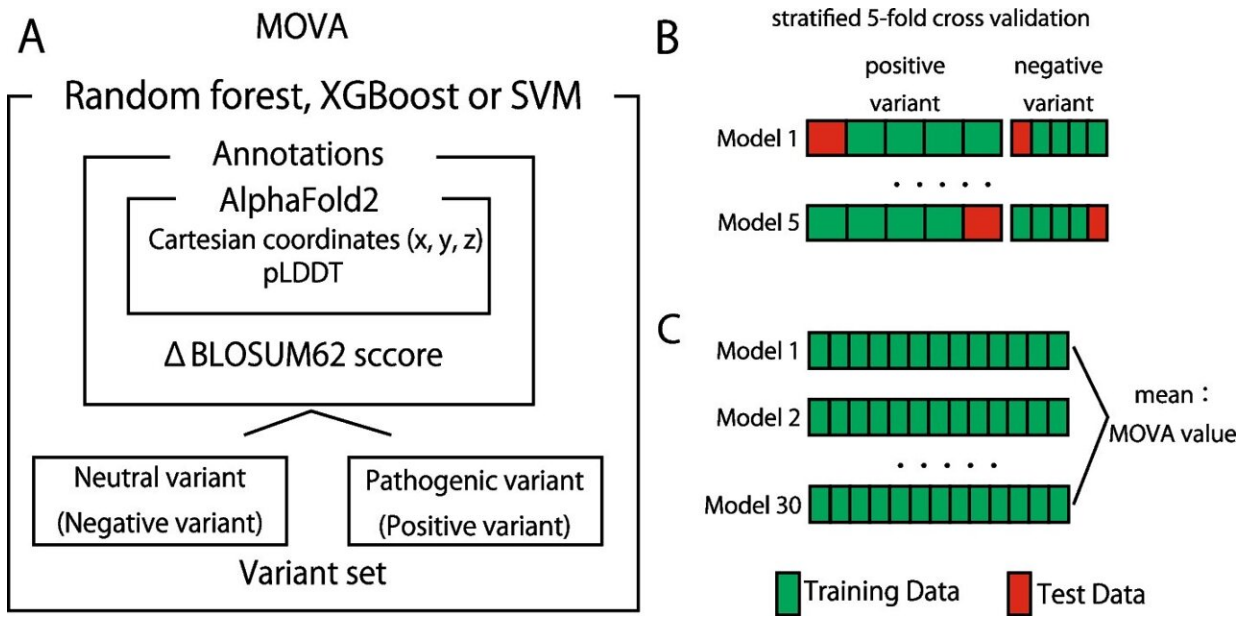
A new method for evaluating the pathogenicity of missense variants using AlphaFold2

June 14 2023



Work flowchart for MOVA. The x, y, z coordinates, and the plddt score for the amino acid residues at the substitution sites in the protein in the pdb file of the Alphafold2 database, and the ΔBLOSUM62 of the substituted amino acid residue, were used as parameters for random forest, XGBoost, or support vector machine (SVM) training (A). The sample group was randomly divided into five subsets as avoiding bias in objective variables. With one subset as the test cases and the rest as the training cases, we built the model. The predictions were calculated and validated using the test data. The models were iteratively built so that all five subsets were test cases. (B). The model was generated 30 times with all variants in the dataset as training data. The probability of each possible variant of the gene being pathogenic was predicted, and the average of the

predictions was used as the MOVA value (C). Credit: *BMC Bioinformatics* (2023). DOI: 10.1186/s12859-023-05338-5

The Department of Neurology at Niigata University has developed a new in silico method for evaluating the pathogenicity of missense variants using AlphaFold2 (MOVA).

Rare variants in the causative gene of ALS are present in 10 to 30% of sporadic ALS cases, which highlights the need for accurate and efficient pathogenicity prediction methods. To predict the pathogenicity of the variants, in silico analysis methods are commonly used. In some ALS causal genes, the mutations are concentrated in specific regions, and the accuracy of pathogenicity prediction can be improved by considering the positional information of the variants.

However, existing methods have not considered information on the position of variants in the protein's structure. MOVA was developed to address this issue and focuses on using positional information in the 3D structure to evaluate the pathogenicity of missense variants. The use of a machine learning method, random forest, in the development of MOVA has also shown promising results.

"The comparison of MOVA with existing in silico analysis methods, such as PolyPhen-2, CADD, REVEL, EVE, and AlphScore, demonstrates its potential in pathogenicity prediction. Combining MOVA with existing methods, such as REVEL and CADD, further improves performance beyond existing pathogenicity prediction methods alone. Moreover, MOVA also showed superior pathogenicity discrimination of hotspot mutations in the TARDBP and FUS genes. This highlights the importance of considering the positional information of variants in protein structures for improved pathogenicity prediction,"

says Dr. Hatano and Dr. Ishihara.

The results of the study were published in the online edition of the journal *BMC Bioinformatics*.

More information: Yuya Hatano et al, Accuracy of a machine learning method based on structural and locational information from AlphaFold2 for predicting the pathogenicity of TARDBP and FUS gene variants in ALS, *BMC Bioinformatics* (2023). DOI: 10.1186/s12859-023-05338-5