CLAIMS

1. A computer-implemented method comprising:

accessing an initial pathogenicity score previously generated by a variant pathogenicity machine-learning model for a target amino acid at a target protein position within a protein based on an amino-acid sequence for the protein;

determining, for the protein and by utilizing a trained temperature prediction machine-learning model, a temperature weight based on the initial pathogenicity score and the amino-acid sequence or a nucleotide sequence corresponding to the protein, wherein the temperature weight estimates a temperature indicating a degree of certainty for pathogenicity scores generated by the variant pathogenicity machine-learning model; and

generating, for the target amino acid at the target protein position, a calibrated pathogenicity score based on the initial pathogenicity score and the temperature weight.

2. The computer-implemented method of claim 1, wherein determining the temperature weight comprises determining the temperature weight for the target protein position of the protein.

3. The computer-implemented method of claim 1 or 2, wherein determining the temperature weight comprises processing the initial pathogenicity score and the amino-acid sequence or the nucleotide sequence through the trained temperature prediction machine-learning model to generate the temperature weight for the protein.

4. The computer-implemented method of claim 1, wherein the trained temperature prediction machine-learning model used for determining the temperature weight comprises a multilayer perceptron (MLP), a convolutional neural network (CNN), a triangle attention neural network, a recurrent neural network (RNN), a long short-term memory (LSTM), a transformer machine-learning model, or a decision tree model.

5. The computer-implemented method of any one of claims 1-4, wherein determining the temperature weight comprises applying a non-linear activation function to an initial weight to determine a positive temperature weight.

6. The computer-implemented method of any one of claims 1-5, wherein the temperature weight indicates the degree of certainty caused by the variant pathogenicity machine-learning or by data input into the variant pathogenicity machine-learning model.

7. The computer-implemented method of any one of claims 1-6, further comprising generating, for display, a graphical visualization of the temperature weight indicating the degree of certainty for the pathogenicity scores generated by the variant pathogenicity machine-learning model for the protein or the target protein position.

8. The computer-implemented method of any one of claims 1-7, wherein determining the temperature weight comprises determining an average temperature weight from initial temperature weights at the target protein position.

9. The computer-implemented method of claim 8, wherein determining the temperature weight comprises utilizing a Gaussian blur model, a median filter, or a bilateral filter to determine the average temperature weight from the initial temperature weights for various amino acids at the target protein position.

10. The computer-implemented method of any one of claims 1-9, wherein the variant pathogenicity machine-learning model that generates the initial pathogenicity score comprises a transformer machine-learning model, a convolutional neural network (CNN), a sequence-to-sequence model, a variational autoencoder (VAE), a multilayer perceptron (MLP), a recurrent neural network (RNN), a long short-term memory (LSTM), or a decision tree model.

11. The computer-implemented method of claim 1, wherein determining the temperature weight comprises determining, utilizing a triangle attention neural network, the temperature weight for the protein by:

determining one or more of an amino-acid pairwise-index-differences embedding representing differences between amino acids in an amino-acid sequence for the protein, an amino-acid pairwise-atom-distances matrix representing pairwise distances between atoms within the protein, a reference-residues embedding representing reference residues for the protein, a conservation multiple-sequence-alignment matrix representing a multiple sequence alignment for the protein from multiple species, and a pathogenicity-scores matrix representing pathogenicity scores generated by the variant pathogenicity machine-learning model for amino acids in the protein;

determining a residue-pair representation based on one or more of the amino-acid pairwise-index-differences embedding, the amino-acid pairwise-atom-distances matrix, the reference-residues embedding, the conservation multiple-sequence-alignment matrix, and the pathogenicity-scores matrix;

projecting temperature weights for protein positions based on the residue-pair representation; and

identifying, from among the temperature weights, the temperature weight for the target protein position within the protein.

12. The computer-implemented method of claim 1, further comprising:

accessing, from an additional variant pathogenicity machine-learning model, an additional pathogenicity score for the target amino acid at the target protein position;

normalizing the additional pathogenicity score and the calibrated pathogenicity score for the target amino acid; and

combining the normalized additional pathogenicity score and the normalized calibrated pathogenicity score to generate a combined pathogenicity score for the target amino acid at the target protein position.

13. The computer-implemented method of claim 1, further comprising:

generating, utilizing an additional variant pathogenicity machine-learning model, an additional pathogenicity score for the target amino acid at the target protein position; and

generating, utilizing a meta variant pathogenicity machine-learning model, a refined pathogenicity score for the target amino acid at the target protein position based on the calibrated pathogenicity score and the additional pathogenicity score.

14. A system comprising at least one processor and a non-transitory computer readable medium comprising instructions that, when executed by the at least one processor, cause the system to perform a method according to any of claims 1-13.

15. A non-transitory computer-readable medium storing instructions that, when executed by at least one processor, cause a system to perform a method according to any of claims 1-13.