

# The Whole is Greater than its Parts – Ensembling Improves Protein Contact Prediction

Wendy Billings<sup>1</sup>, Connor Morris<sup>1</sup>, and Dennis Della Corte<sup>1</sup>

<sup>1</sup>Brigham Young University

October 20, 2020

## Abstract

The prediction of amino acid contacts from protein sequence is an important problem, as protein contacts are a vital step towards the prediction of folded protein structures. We propose that a powerful concept from deep learning, called ensembling, can increase the accuracy of protein contact predictions by combining the outputs of different neural network models. We show that ensembling the predictions made by different groups at the recent Critical Assessment of Protein Structure Prediction (CASP13) outperforms all individual groups. Further, we show that contacts derived from the distance predictions of three additional deep neural networks – AlphaFold, trRosetta, and ProSPr – can be substantially improved by ensembling all three networks. In a final assessment, we show that ensembling these recent deep neural networks with the best CASP13 group creates a superior contact prediction tool. In order to build novel methods on these findings, we propose the creation of a better protein contact benchmark set and additional open-source contact prediction methods.

## Hosted file

Ensembling\_Improves\_Contact\_Predictions\_20201017.pdf available at <https://authorea.com/users/368631/articles/487707-the-whole-is-greater-than-its-parts-ensembling-improves-protein-contact-prediction>