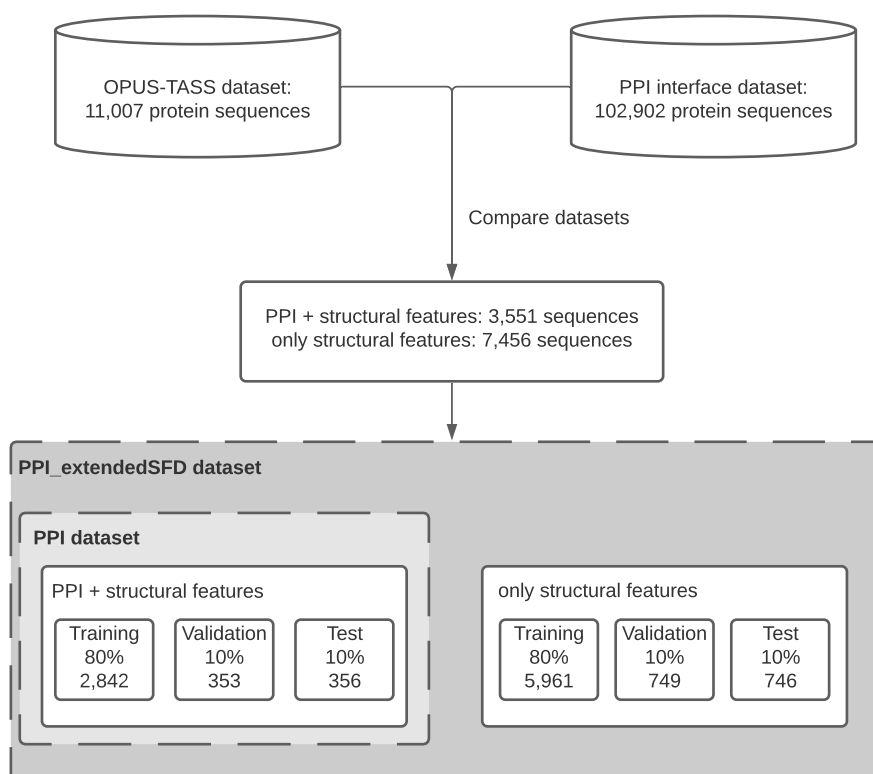


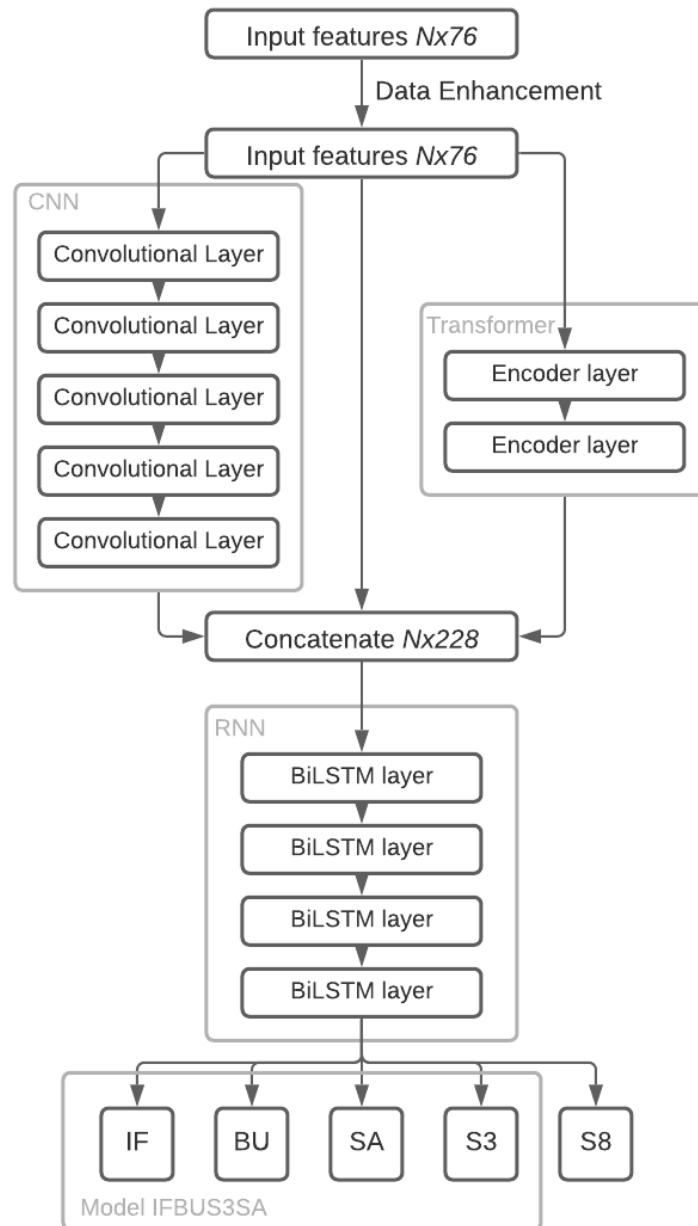
# Supplementary information for Multi-task learning to leverage partially annotated data for PPI interface prediction

Henriette Capel<sup>1</sup>, K. Anton Feenstra<sup>1</sup>, Sanne Abeln<sup>1,\*</sup>

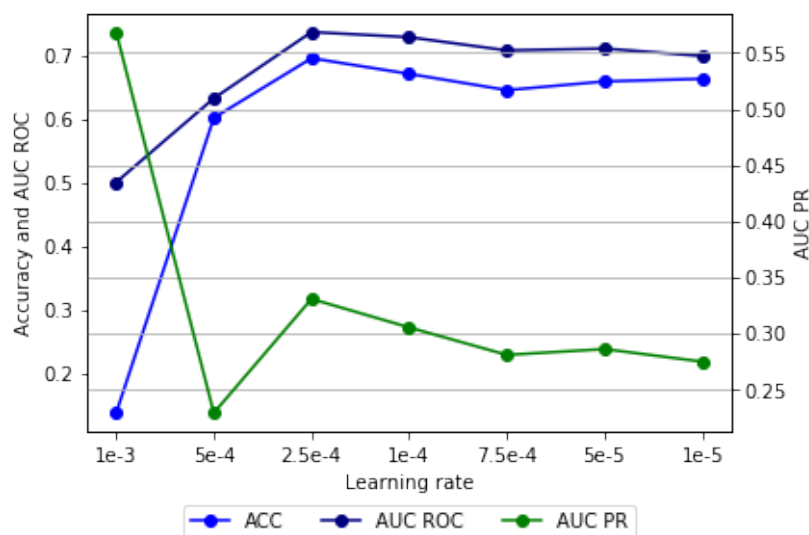
<sup>1</sup> IBIVU – Center for Integrative Bioinformatics, Vrije Universiteit Amsterdam, Amsterdam 1081HV, The Netherlands



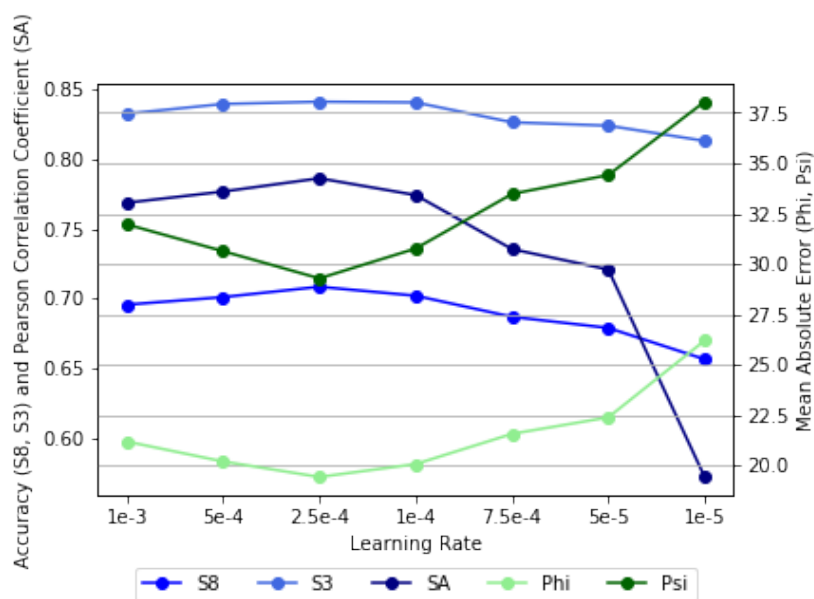
**Supplementary Fig. 1.** Construction of the PPI and PPI\_extendedSFD dataset based on the combined training and validation dataset of OPUS-TASS. For one third of this dataset PPI interface annotations were available and stored in the PPI interface dataset. The PPI dataset consist of all proteins of the OPUS-TASS dataset for which PPI interface annotations are available. This dataset is a subset of the PPI\_extendedSFD dataset which is augmented by the residual proteins of the OPUS-TASS dataset.



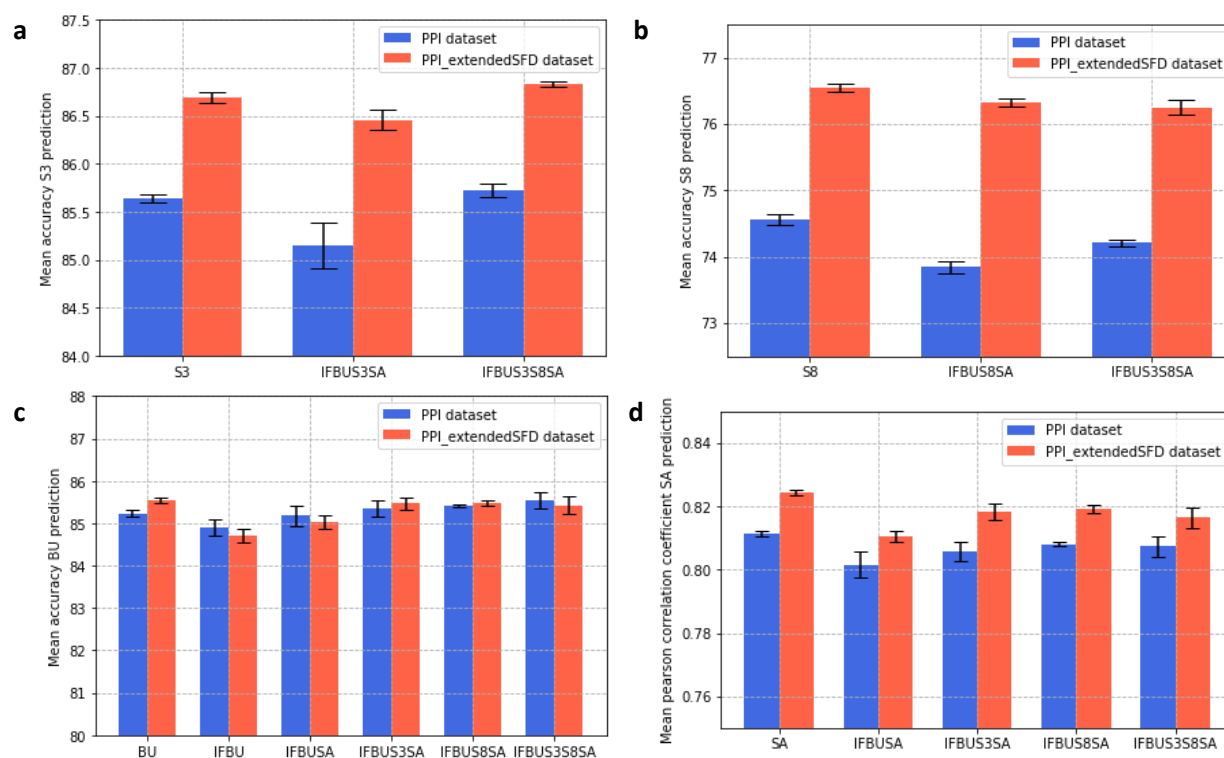
**Supplementary Fig. 2.** The architecture of the deep learning model based on the OPUS-TASS model of Xu et al. (2020). The input of the model is a matrix of size length of the protein (N) by the 76 input features. The model consists of a convolutional neural network (CNN), a transformer, and a recurrent neural network (RNN), specifically a bidirectional long short-term memory (LSTM). The output of the model are the possible prediction tasks. The predictions for the different tasks are coupled via the network architecture, and their inclusion in the cost function.



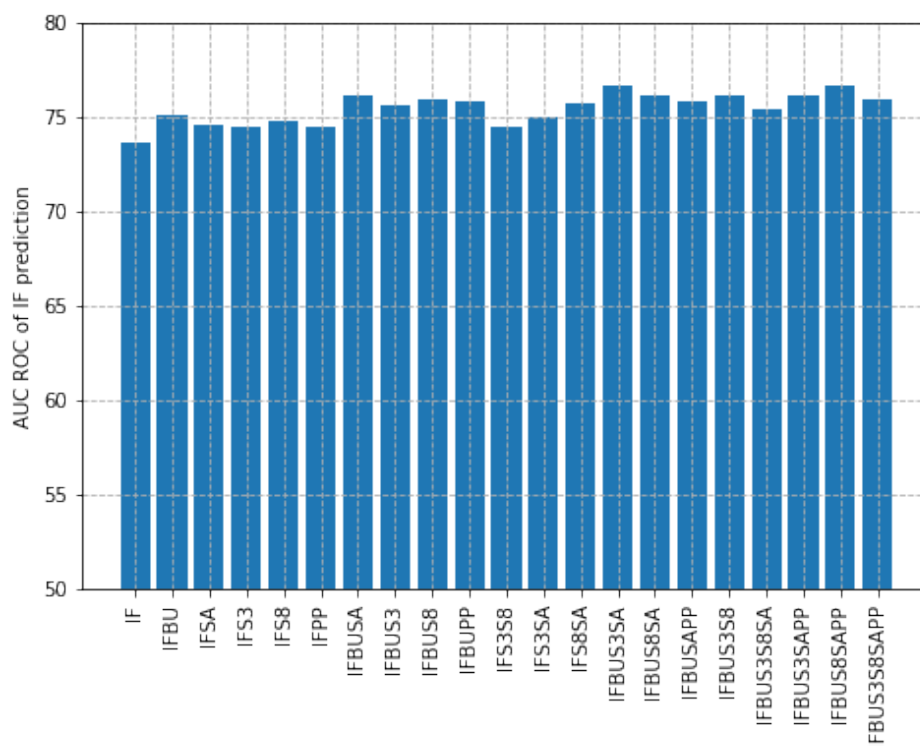
**Supplementary Fig. 3.** Model performance of the single task model on the PPI interface prediction task for different initial learning rates. The models are trained on the PPI dataset and performance is measured by the accuracy (ACC, in blue), area under the receiver operator characteristics curve (AUC ROC, in dark blue) and area under the precision-recall curve (AUC PR, in green) on the validation set. Highest prediction performances are shown for the initial learning rate  $2.5e-4$ . Note that the highest AUC PR score at  $1e-3$  was only obtained when all residues were predicted to be non-interface residues.



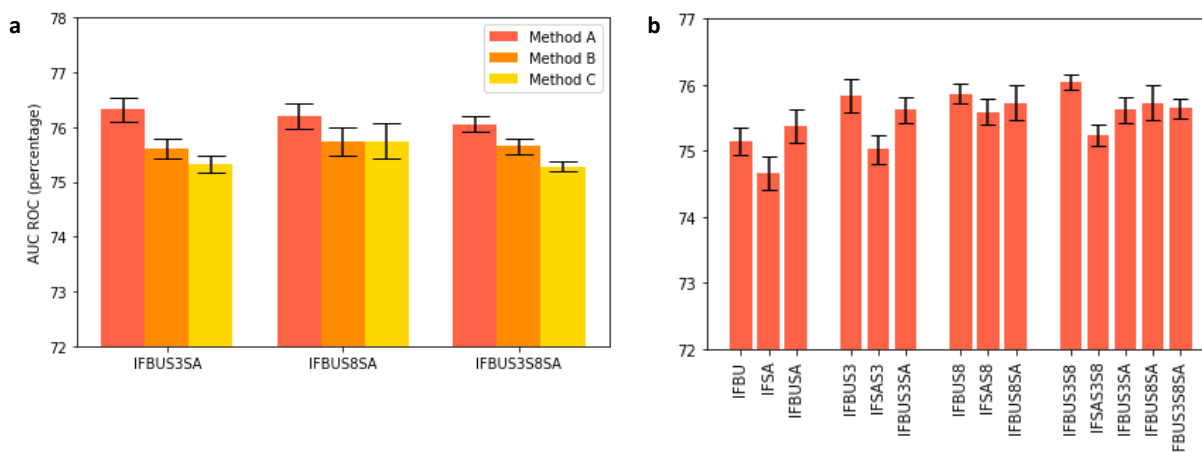
**Supplementary Fig. 4.** The model performance on the different prediction tasks (see legend) for different initial learning rates. The model used, is the best performing model (C4) in the work by Xu et al. (2020). The presented performances are accuracy of the secondary structure prediction in three (S3, light blue) and eight (S8, blue) classes, pearson correlation coefficient of the absolute solvent accessibility prediction (SA, dark blue), and the mean absolute error of the phi and psi angle (Phi and Psi, in light and dark green). The highest accuracy and pearson correlation coefficient, and the lowest mean absolute error for all tasks is reached by using the initial learning rate of  $2.5e-4$ . This result is in line with the hyperparameter tuning result based on the PPI interface prediction.



**Supplementary Fig. 5.** Model performances for the related prediction task of the single-task models and the multi-task model. Note that the single-task model is optimised by using its single-task performance measure in the stopping criteria and the multi-task models are optimised for the PPI interface prediction. All models are trained four times on both the PPI and PPI\_extendedSFD dataset. Note that the extended dataset contains training labels for all tasks, except for the IF task which is only partially annotated. Mean performance and standard error are shown for the validation set. The multi-task models attain in most cases similar performance as the single-task models. **(a)** Model performances measured in accuracy of the secondary structure prediction in three classes. **(b)** Model performances measured in accuracy of the secondary structure prediction in eighth classes. **(c)** Model performances measured in accuracy of the buried residue prediction. **(d)** Model performances measured in pearson correlation coefficient of the absolute solvent accessibility prediction.



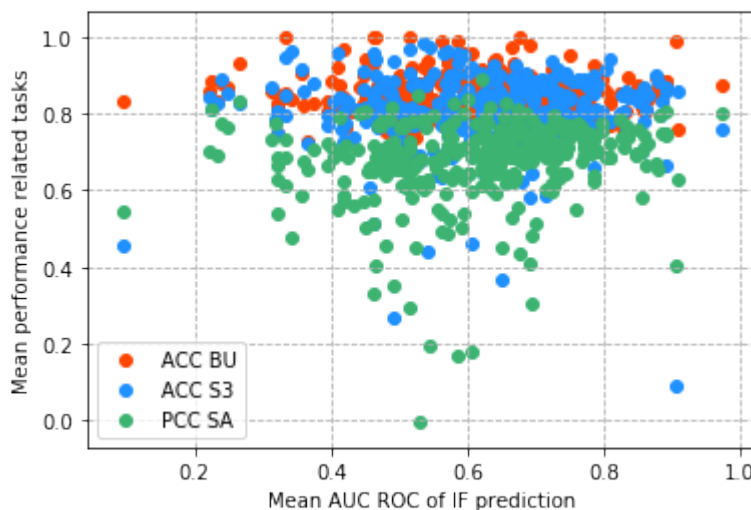
**Supplementary Fig. 6.** Comparison of the single-task model (IF) against multiple multi-task models. Model performance is measured by the area under the receiver operator characteristics curve (AUC ROC). All models are trained once and performance is shown for the validation set.



**Supplementary Fig. 7.** Increasing the weight of the fraction of the PPI interface prediction loss in the cost function. **(a)** Comparing the standard method in which the weights are set to 1.0 (Method A, red), against two different weighting methods. For Method B (orange) we halved the weights of pairs of similar prediction tasks, such both sum up to a weight equal to the other tasks. For Method C (yellow) we assigned a weight to the interface loss such that its fraction is 50% of the total cost function. The AUC ROC scores on the interface prediction tasks is shown for models IFBUS3SA, IFBUS8SA and IFBUS3S8SA. **(b)** AUC ROC scores on the interface prediction task are shown for models in which Method B is applied. Note that a weight is kept the same as in Method A if only one of the two correlated prediction task is present. The models are grouped based on the prediction tasks they included and are considered as similar

| Model        | Dataset         | SS8<br>(ACC)             | SS3<br>(ACC)             | BUR<br>(ACC)             | ASA<br>(PCC)             | PHI<br>(MAE) | PSI<br>(MAE) | PPI<br>(AUC ROC)         |
|--------------|-----------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------|--------------|--------------------------|
| IF           | PPI             |                          |                          |                          |                          |              |              | 0.7364                   |
| IFPP         | PPI             |                          |                          |                          |                          | 18.53        | 27.39        | 0.7486                   |
| IFPP         | PPI_extendedSFD |                          |                          |                          |                          | 17.94        | 26.17        | 0.7441                   |
| IFBUPP       | PPI             |                          |                          | 0.8527<br><b>+0.0037</b> |                          | 18.47        | 27.25        | 0.7550<br><b>+0.0042</b> |
| IFBUPP       | PPI_extendedSFD |                          |                          | 0.8538<br><b>+0.0066</b> |                          | 17.75        | 25.61        | 0.7582<br><b>+0.0078</b> |
| IFBUSAPP     | PPI             |                          |                          | 0.8548<br><b>+0.0030</b> | 0.8087<br><b>+0.0070</b> | 18.26        | 26.92        | 0.7515<br>+0.0006        |
| IFBUSAPP     | PPI_extendedSFD |                          |                          | 0.8550<br><b>+0.0046</b> | 0.8206<br><b>+0.0100</b> | 17.65        | 25.38        | 0.7585<br><b>-0.0030</b> |
| IFBUS3SAPP   | PPI             |                          | 0.8563<br><b>+0.0050</b> | 0.8532<br>-0.0003        | 0.8069<br>+0.0010        | 18.59        | 26.72        | 0.7601<br>+0.0028        |
| IFBUS3SAPP   | PPI_extendedSFD |                          | 0.8676<br><b>+0.0030</b> | 0.8549<br>+0.0002        | 0.8197<br>+0.0015        | 17.56        | 24.96        | 0.7617<br><b>-0.0046</b> |
| IFBUS8SAPP   | PPI             | 0.7449<br>+0.0064        |                          | 0.8549<br>+0.0007        | 0.8092<br><b>+0.0010</b> | 18.09        | 26.41        | 0.7586<br>-0.0006        |
| IFBUS8SAPP   | PPI_extendedSFD | 0.7667<br>+0.0034        |                          | 0.8556<br>+0.0008        | 0.8208<br><b>+0.0015</b> | 17.28        | 24.31        | 0.7664<br><b>+0.0049</b> |
| IFBUS3S8SAPP | PPI             | 0.7470<br>+0.0049        | 0.8587<br>+0.0015        | 0.8587<br><b>+0.0034</b> | 0.8091<br>+0.0017        | 18.19        | 26.23        | 0.7515<br><b>-0.0060</b> |
| IFBUS3S8SAPP | PPI_extendedSFD | 0.7658<br><b>+0.0033</b> | 0.8690<br>+0.0007        | 0.8560<br>+0.0017        | 0.8189<br>+0.0023        | 17.46        | 24.51        | 0.7596<br><b>+0.0052</b> |

**Supplementary Table 1.** Comparison of the multi-task models including the torsion angle predictions (PP) as related prediction tasks against models without this prediction. Performances for the phi- and psi- angles are measured by the mean absolute error and are shown for the validation set after training the models once. The plus and minus signs indicate the increase or decrease of the models including the torsion angles compared to the model without this prediction task. Values in bold indicate that they lie outside the standard error region.



**Supplementary Fig. 8.** Error analyse on the individual proteins of the test set after training the IFBUS3SA model on the PPI\_extendedSFD dataset including only one tenth of the PPI interface information. The mean accuracy of the prediction of the secondary structural component and buried residues, and the mean pearson correlation coefficient of the absolute solvent accessibility are plotted against the mean AUC ROC score of the PPI interface prediction per protein.