Convolutional ProteinUnetLM competitive with LSTM-based protein secondary structure predictors

Katarzyna Stąpor[1], Krzysztof Kotowski[1], Piotr Fabian[1], and Irena Roterman[2]

[1]Politechnika Slaska
[2]Uniwersytet Jagiellonski w Krakowie Collegium Medicum

October 21, 2022

Abstract

The protein secondary structure (SS) prediction plays an important role in the characterization of general protein structure and function. In recent years, a new generation of algorithms for SS prediction based on embeddings from protein language models (pLMs) is emerging. These algorithms reach state-of-the-art accuracy without the need for time-consuming multiple sequence alignment (MSA) calculations. LSTM-based SPOT-1D-LM and NetSurfP-3.0 are the latest examples of such predictors. We present the ProteinUnetLM model using a convolutional Attention U-Net architecture that provides prediction quality and inference times at least as good as the best LSTM-based models for 8-class SS prediction (SS8). Additionally, we address the issue of the heavily imbalanced nature of the SS8 problem by extending the loss function with the Matthews correlation coefficient (MCC), and by proper assessment using previously introduced adjusted geometric mean metric (AGM). ProteinUnetLM achieved better AGM and sequence overlap score (SOV) than LSTM-based predictors, especially for the rare structures 310-helix (G), beta-bridge (B), and high curvature loop (S). It is also competitive on challenging datasets without homologs, free-modeling targets, and chameleon sequences. Moreover, ProteinUnetLM outperformed its previous MSA-based version ProteinUnet2, and provided better AGM than AlphaFold2 for 1/3 of proteins from the CASP14 dataset, proving its potential for making a significant step forward in the domain. To facilitate the usage of our solution by protein scientists, we provide an easy-to-use web interface under https://biolib.com/SUT/ProteinUnetLM/.

Introduction

Proteins are macromolecules built from amino acids (AAs) and the kind and order of the AAs (known as primary structure) are determined by DNA sequence. Most proteins are built out of 20 different AAs, which only differ in the organic side chain (R groups). The backbone is the same for all AAs, it consists of an amino group, a central carbon atom, and a carboxyl group. From a linear sequence of amino acids, a protein sequence folds rapidly into secondary or local arrangements, and then into a tertiary or three-dimensional structure. The regular structural elements created by hydrogen bonds between hydrogen donors in the nitrogen part and hydrogen acceptors in the carboxyl group of the backbone define the secondary structure (SS) of a protein. These structures stabilize the protein. Different types of secondary structures are known: the alpha-helix, beta-sheet, and loop are the three most important ones. There are different SS assignment methods among which the most commonly used is the algorithm of the Dictionary of Protein Secondary Structure (DSSP) [1]. DSSP proposed 8 classes of secondary protein structures which provide more

information about the actual 3D formation of the protein. There are three helix states: 310-helix (G), alpha-helix (H), and pi-helix (I); two beta-sheet states: beta-bridge (B) and beta-strand (E); and three loop (or irregular) classes: high curvature loop (S), beta-turn (T), and random coil (C).

SS prediction plays an important role in protein tertiary structure prediction as well as in the characterization of general protein structure and function. Because the SS provides the first step toward native or tertiary structure prediction, it is utilized in many protein folding algorithms [2–5] and in a variety of bioinformatics areas, including proteome and gene annotation[6–9], the determination of protein flexibility[10], the subcloning of protein fragments for expression and the assessment of evolutionary trends among organisms[2]. Therefore, protein SS prediction remains an active area of research and an integral part of protein analysis.

Three generations of methods and algorithms are described in the literature for secondary structure prediction [11]. The first generation, represented by the Chou-Fasman method[12], exploited statistical propensities of residues to a particular secondary structure class. These methods usually achieved a prediction accuracy of less than 60%. The second generation of methods was developed in the 1980s. They used advanced statistical methods, machine learning techniques, and information about neighbor residues, usually using a sliding window approach. These methods include, among others, GOR [13] and Lim [14]. The accuracy of predicting secondary structure as assessed by the Q3 parameter was less than 65% [15]. The third generation of methods appeared in the 1990s. They used neural networks and additional features based on multiple sequence alignment (MSA) profiles, e.g., PSSM - position-specific scoring matrices [16] or HHblits (iterative protein sequence search according to the hidden profile) Markov models [17]. The Q3 accuracy of these methods exceeded 80% for models such as PSIPRED [18]. Given the increasing number of known protein sequences and more efficient neural network architectures, the latest methods can predict the secondary structure with over 70% accuracy for an 8-class problem such as NetSurfP-2.0 (Q8 = 71.43% for CASP12) [19] or SPOT-1D (Q8 = 73.67% for CASP12) [20] based on long-term memory (LSTM) bidirectional recursive neural networks (BRNN).

The fourth, recently emerging, generation of methods uses protein language models (pLMs) inspired by advancements in the natural language processing (NLP) field[21]. Secondary structure predictors of the latest generation use embeddings from models like SeqVec [22]or transformers-based networks like ProtTrans [23], ESM[24], or BERT [25] that learn the*grammar* of the *language of life* . The concept of embedding in machine learning is an idea of encoding categorical parameters (i.e., sequences of amino acids) as highly informative numerical vectors that can be used as inputs of neural networks. LM-based classifiers are able to achieve SS prediction performance close to or better than the previous generation of methods, e.g., NetSurfP-3.0 [26]and ProtT5Sec [23] which are comparable to NetSurfP-2.0; or SPOT-1D-LM [27] which improves over SPOT-1D. Most importantly, the sequence embeddings can be generated in a fraction of the time with respect to MSA-based features [23]. Additionally, the recent success of AlphaFold2 [28]proved that NLP-inspired mechanisms like attention and transformers may be extremely useful in protein tertiary structure prediction. It predicted protein structures near the X-ray resolution in the latest Critical Assessment of protein Structure Prediction (CASP14)[29]. However, there is still room for improving AlphaFold2 predictions in terms of SS. The newest results demonstrate that its accuracy decreases for longer loop regions, and it has a tendency to slightly overpredict helices and beta-sheets[30].

In our study, we present how our proposed ProteinUnetLM model based on Attention U-Net architecture and LM-based features improves over its previous MSA-based version [31], and over its closest LM-based competitors (SPOT-1D-LM, NetSurfP-3.0, and ProtT5Sec): (i) its prediction performance measured by sequence-level adjusted geometric mean (AGM) is better than all other LM-based networks while being comparable in segment overlap metric (SOV8) and Q8 accuracy, (ii) it provides the best results for rare structures G, B, and S, (iii) its prediction time is comparable to the fastest NetSurfP-3.0. These results support our hypothesis that LSTMs are not needed to achieve state-of-the-art performance as our fully-convolutional Attention U-Net architecture is at least as accurate and at least as fast as any LSTM-based competitor. We especially focus on the issue of imbalance in the SS8 prediction problem, so besides proposing proper metrics and statistical methodology, we extended the loss function of the network with the Matthews correlation coeffi-

cient (MCC) which improved the performance for rare structures. In comparison with secondary structures parsed from tertiary structure predictions of famous AlphaFold2[28] on the CASP14 dataset, ProteinUnetLM achieved better AGM for 10 out of 30 sequences and better precision for helix (H) structure proving its potential for making a significant step forward in the domain.

Materials and Methods

Datasets

For a fair comparison with previous MSA-based models ProteinUnet2 and SPOT-1D, we use the same training, validation, and test sets. The training set TR10029 contains 10'029 sequences, the validation set VAL983 has 983 sequences, and there are two test sets: TEST2016 with 1213 and TEST2018 with 250. See [20] for the details about these datasets.

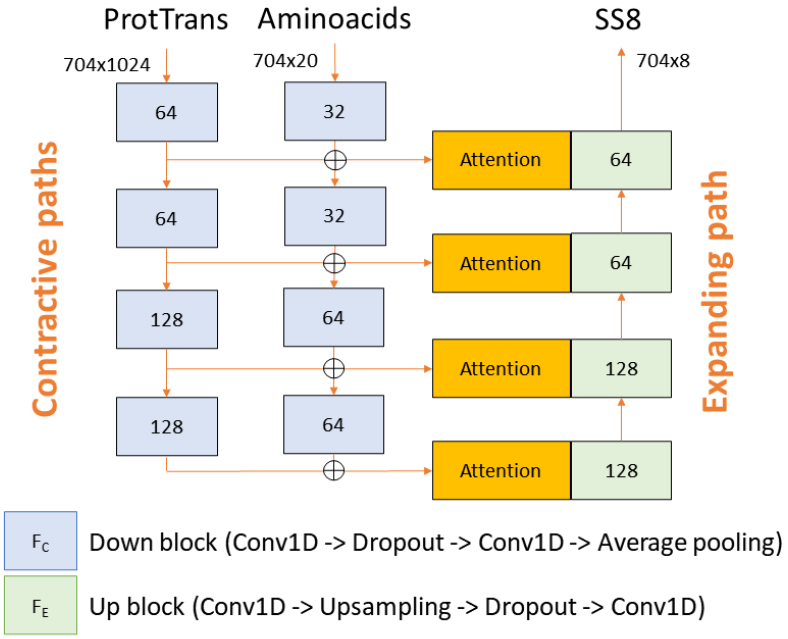
For comparison with LM-based models, we use four additional test sets introduced in SPOT-1D-LM [27]. The largest TEST2020 contains 547 sequences deposited between the years 2018 and 2020 where remote homology to proteins released before 2018 was removed. Two separate test subsets were extracted from TEST2020 to assess the performance of algorithms in specific cases, TEST2020-HQ with 121 sequences with high-resolution structures ($<$ 2.5 A), and Neff1-2020 with 46 sequences with no homologs (Neff=1). The original list of sequences in the TEST2020 set contained 671 sequences but we could not find some sequence codes in Protein Data Bank (PDB) or ProteinNet [32]. Thus, we attached our lists of sequences and their DSSP-generated SS8 for TEST2020 as Supplementary File 2. The smallest CASP12-FM dataset contains 22 sequences without any known structural templates (free-modeling) from CASP12.

Finally, we compare all the networks on the newest CASP14 dataset of 30 proteins (the same as used in [31]) for which the PDB targets were available on the official CASP14 challenge page (predictioncenter.org/download_area/CASP14/targets).

The ratios of each SS8 in all mentioned datasets are given in Supplementary Figure S1. It gives an overview of how imbalanced is the problem of SS8 prediction.

Attention U-Net for secondary structure prediction

U-Net is a state-of-the-art architecture in image segmentation tasks[33–35] and we previously successfully introduced it into the domain of protein secondary structure prediction by creating the ProteinUnet model [36]. The Attention U-Net architecture of ProteinUnetLM presented in this paper adapts our previous ProteinUnet2 model [31] to the features from pLMs.



The network learns higher-level features in convolutional contractive paths, concatenates them, and passes them to the attention gates that learn to filter out irrelevant features [37,38]. Finally, the filtered features are passed to the convolutional expanding path that learns to predict the sequence of 8-class secondary structures as the output layer with softmax activation connected to the last up-block (Figure 1). As in ProteinUnet2, taking into account that the receptive field of our network includes 710 residues [36], we limited the input sequence length to 704. We also resigned from predicting 3-class secondary structures (SS3) as they can be easily derived from 8-class predictions (SS8), and we did not notice any advantages of including SS3 output in the network training in our previous works. Other hyperparameters were the same as in the ProteinUnet2 paper, to enable direct comparisons between the architectures. Specifically, we have 2 convolutions with 1D kernels of length 7 and ReLU activations, and dropouts with a 0.1 rate between convolutional layers in all blocks. In overall, the model has 2'501'260 trainable parameters. It is worth noticing that ProteinUnetLM is a single model, not an ensemble of 10 models like in previous versions of ProteinUnet. The ensembling provides a bit better performance in some metrics, as presented in Supplementary Table S1, but we decided to sacrifice it to improve the inference time.

ProteinUnetLM takes a sequence of feature vectors $X = (x_1, x_2, x_3, \ldots, x_N)$as input, where x_i is the feature vector corresponding to the$i$ th residue, and it returns a vector$Y = (y_1, y_2, y_3, \ldots, y_N)$ as output, where y_i is the vector 8 probabilities of i th residue being in one of the SS8 states. Our model is fed with 1024 features from ProtTransT5-XL-U50 [23]. Each feature is standardized to ensure a 0 mean and SD of 1 in the training data. Using features from the ESM-1b model [39] instead of ProtTrans resulted in suboptimal performance as presented in Supplementary Table S1. Additionally, we use a one-hot encoded sequence of amino acids as the second input to keep a close comparison with ProteinUnet2. However, Supplementary Table 1 suggests that it has a minor impact on the results.

Training procedures and improved loss function

We trained a single ProteinUnetLM model using TR10029 as a training set and VAL983 as a validation set. The model was trained to simultaneously minimize the categorical cross-entropy (CCE, Equation 1) and

maximize the Matthews correlation coefficient (MCC, Equation 2) by defining a loss function as a difference between average CCE and average MCC across the training batch (Equation 3).

$CCE = -\sum_{i=1}^{8} y_i \log \hat{y}_i$, where $\mathbf{y}$ is a target vector and $\hat{\mathbf{y}}$ is a model output, (1)

(2)

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)} + e}$$

where $TP = \mathbf{y} \cdot \hat{\mathbf{y}}$, $TN = (\mathbf{1} - \mathbf{y}) \cdot (\mathbf{1} - \hat{\mathbf{y}})$, $FP = (\mathbf{1} - \mathbf{y}) \cdot \hat{\mathbf{y}}$, $FN = \mathbf{y} \cdot (\mathbf{1} - \hat{\mathbf{y}})$, and e is a very small number preventing division by zero,

$Loss = CCE - MCC$ (3)

MCC was already evaluated as one of the most reliable, universal, and informative metrics in machine learning and bioinformatics problems[40–42]. We involved MCC in the training loss to address the imbalance problem of the protein SS prediction and improve the results on rare structures. The ablation study in Supplementary Table S1 suggests that it was successfully achieved by improving metrics for TEST2018.

We used an Adam optimizer [43] with batch size 8 and an initial learning rate of 0.001. The learning rate was reduced by a factor of 0.1 when there was no improvement in the validation loss for 4 epochs. The training was stopped when the validation loss was not improving for 6 epochs and the checkpoint with the lowest validation loss among all epochs was selected as the final ProteinUnetLM model.

ProteinUnetLM was implemented in the environment containing Python 3.8 with TensorFlow 2.9 accelerated by CUDA 11.7 and cuDNN 8. The inference code and trained models are available on the CodeOcean platform (https://codeocean.com/capsule/7112101) ensuring high reproducibility of the results. An easy-to-use web interface is accessible on Biolib (https://biolib.com/SUT/ProteinUnetLM/). The code for training can be run in the Google Colab notebook (https://colab.research.google.com/drive/1Onh6xlg-a-_QDy2EL_-t9XmKa8T3VLVEv).

Metrics and statistical testing

Following the reasoning from the ProteinUnet2 paper, we utilize the Adjusted Geometric Mean (AGM) metric as a primary metric for assessing the prediction performance. It is well-suited for bioinformatics imbalance problems, it performs better than F-score in these problems, and it has no parameters (like a beta in F-score) [44]. It is given by Equation 4 where GM is the geometric mean (Equation 5) and N_n is the proportion of negative samples. It takes value from range <0, 1> where 1 is a perfect prediction. The metric can be calculated both at the residue and at the sequence level. By *the residue level* , we mean calculating the metric once for all residues in all sequences in the dataset, and by *the sequence level* , we mean calculating the metric separately for each sequence in the dataset and taking an average out of scores. To aggregate the metric across 8 classes, we use macro averaging – we calculate the AGM score separately for each class and average the results to create the macro-AGM score.

$$AGM = \begin{cases} \frac{GM \ + \ Specificity * N_n}{1+N_n}, & \& Sensitivity > 0 \\ 0, & \& Sensitivity = 0 \end{cases} \quad (4)$$

$$GM = \sqrt{Precision * Sensitivity} \quad (5)$$

Additionally, we use a segment overlap score for 8 classes (SOV8) as defined by the SOV_refine algorithm [45]. The SOV score was designed specifically to compare two sequences of protein secondary structures in which the continuity of segments has important meanings. It promotes classifiers that are able to more consistently predict segments of the same structure without breaking it with incorrect prediction. It takes value from range <0, 1> where 1 is a perfect prediction. It can be calculated only at the sequence level.

We report Q8 accuracy at the residue level just for compatibility with previous literature, we consider it highly inappropriate for such an imbalanced problem [31] and we do not perform any statistical tests on Q8. To avoid any potential bias towards the MCC metric that was optimized during training, we decided to avoid its assessment during testing. The implementations of all the mentioned metrics are available in our computational capsule on CodeOcean.

For fair experimental classifier evaluation [46], we apply paired statistical comparisons between ProteinUnetLM and other networks using two-sided paired sample permutation tests for difference in mean classifier performances (*perm.paired.loc* function from*wPerm* R package with 10'000 replications). We decided to use the permutation approach as it does not make any assumptions about a sampling distribution or a sample size. The tests were performed at the sequence level, that is, we first calculated metric values for each sequence separately, and then we run statistical tests on them. We selected a significance level of 0.05 but we use Bonferroni correction for multiple hypothesis testing (MHT) when ProteinUnetLM is compared with more than 1 other classifier on the same dataset. It means that the significance level is divided by the number of comparisons, i.e., the significance level for TEST2016 is 0.025 (2 comparisons), for TEST2018 is 0.01 (5 comparisons), and so on. To quantify the effect size and its direction, as proposed previously in [31], we use Cohen's d effect for paired samples calculated as the mean difference divided by the standard deviation of the differences[47].

Comparison with LSTM-based networks

We compare ProteinUnetLM with the three latest networks for protein secondary structure prediction based on features from protein LMs: NetSurfP-3.0 [26], ProtT5Sec [23], and SPOT-1D-LM [27]. SPOT-1D-LM and NetSurfP-3.0 are hybrids of convolutional feature extractors and bidirectional recurrent neural networks (BRNN) with long short-term memory (LSTM) units. SPOT-1D-LM uses ResNet convolutional encoder [48], and NetSurfP-3.0 uses two convolutional layers with very large kernels (129 and 257) and paddings (64 and 128) followed by 0.5 dropouts and ReLU activations. In fact, SPOT-1D-LM, unlike ProteinUnetLM and NetSurfP-3.0, is an ensemble of three separate networks, BRNN, ResNet, and ResNet-BRNN hybrid, which increases the complexity and time of the training and prediction. Interestingly, the authors of NetSurfP-3.0 showed that replacing their downstream architecture with a transformer-based encoder resulted in suboptimal performance [26].

The main purpose of LSTM networks is to learn both short and distant dependencies within sequences [49]. Distant dependencies are not possible to capture by convolutional layers because of their limited receptive field, but this can be overcome by using an attention mechanism [50]. The positive impact of the attention mechanism on the results of ProteinUnetLM can be observed in Supplementary Table S1. Additionally, long skip connections in U-Net architecture, besides stabilizing gradient updates in deep architectures, prevent from losing fine-grained details of the input sequence [51]. Moreover, the training time of LSTM networks is several times longer than for fully-convolutional networks[36,52]. Mainly because RNNs are harder to parallelize and they take less advantage of GPU processing [53].

ProtT5Sec was introduced as a simple classification backbone based on ProtTrans features [23]. The authors tested four different classifiers: logistic regression, fully-connected network, fully-convolutional network (CNN), and BRNN-LSTM. They concluded that two-layer CNN (32 filters of size 7) provided the best performance while being computationally less expensive than LSTM which reached similar results. In our paper, we build on this conclusion, and we hypothesize that LSTM networks are not necessary to achieve state-of-the-art results in protein secondary structure prediction and can be effectively replaced by the proposed

Attention U-Net architecture when using features from pLMs.

Results

Comparison with MSA-based classifiers

First, we directly compared ProteinUnetLM with its previous version ProteinUnet2 based on multiple sequence alignment (MSA) features (https://codeocean.com/capsule/0425426), and its competitor SPOT-1D network [20](https://sparks-lab.org/server/spot-1d/). ProteinUnetLM achieved the best results in 7 out of 8 combinations of the test set (TEST2016 and TEST2018) and metric (macro-AGM at residue and sequence levels, SOV8, and Q8 at the residue level) presented in Table 1. In terms of statistical significance, ProteinUnetLM had statistically significantly better macro-AGM in comparison to ProteinUnet2 (p=0.0054 on TEST2016) and SPOT-1D (p=2e-5 on TEST2016 and p=0.0077 on TEST2018) with small effect sizes (d $<$ 0.2). There were no large or significant differences between networks for the SOV8 metric. ProteinUnetLM achieved the highest Q8 among all networks.

The separate AGM scores for each SS8 in Supplementary Table S2 show that ProteinUnetLM is better than ProteinUnet2 for all structures on the biggest TEST2016 and TEST2018 datasets, especially on the rare structures B, G, S, and I. Importantly, ProteinUnetLM achieved correct predictions for the rarest structure "I" which was not possible using MSA features in ProteinUnet2. It confirms that LMs provide better features for protein SS than MSA-based methods like PSSM or HHblits. Especially, taking into account the fact that ProteinUnetLM is a single model, not an ensemble of 10 models like ProteinUnet2.

Table 1 . The comparison of **macro-AGM** at the**sequence** and **residue level** , **SOV8** at the**sequence level,** and **Q8** at the **residue level**on two test sets for ProteinUnetLM vs ProteinUnet2 and SPOT-1D. The best results for each dataset are boldfaced. The green shading of sequence level scores denotes the statistical significance that ProteinUnetLM has a better mean with standard deviations (SD), p-values, and Cohen's effect size (d) given below the score.

Test set	**Model**	**Macro-AGM**	**Macro-AGM**	**SOV8**	**Q8**
		Residue level	**Sequence level (±SD/p-value/d)**	**Sequence level (±SD/p-value/d)**	**Residue level**
TEST2016	ProteinUnetLM	**0.829**	**0.767** ± 0.135	**0.786** ± 0.113	**0.771**
	ProteinUnet2	0.729	0.759 ±0.122/p=0.005/d=0.163	0.784 ±0.102/p=0.526/d=0.014	0.766
	SPOT-1D	0.809	0.747 ±0.127/p=2e-5/d=0.154	0.784 ±0.103/p=0.385/d=0.019	**0.771**
TEST2018	ProteinUnetLM	**0.728**	**0.728** ±0.182	0.760 ±0.144	**0.756**
	ProteinUnet2	0.721	0.719 ±0.175/p=0.273/d=0.012	**0.763** ±0.129/p=0.643/d=-0.020	0.746
	SPOT-1D	0.718	0.707 ±0.176/p=0.008/d=0.113	0.762 ±0.132/p=0.767/d=-0.012	0.754

Comparison with LM-based classifiers

We compared our network with the three latest networks of similar utility based on features from pLMs: NetSurfP-3.0 [26], ProtT5Sec [23], and SPOT-1D-LM [27]. SPOT-1D-LM uses features from both ProtTransT5-XL-U50[23] and ESM-1b [39] LMs, NetSurfP-3.0 uses only ESM-1b with 1280 features, and ProtT5Sec only ProtTransT5-XL-U50 with 1024 features. We run SPOT-1D-LM from its source code (https://github.com/jas-preet/SPOT-1D-LM), and we used web interfaces to run NetSurfP-3.0 (https://dtu.biolib.com/NetSurfP-3/) and ProtT5Sec (https://api.bioembeddings.com/). It needs to be noted that these networks were trained on different, but partially overlapping datasets. ProteinUnetLM was trained on 10029 (TR10029 dataset) and validated on 983 sequences (VAL983 dataset), NetSurfP-3.0 and ProtT5Sec were trained on 10337 and validated on 500 sequences, and SPOT-1D-LM was trained on 38913 (including most of the sequences from TR10029 and TEST2016) and validated on 100 sequences. To ensure no overlap between the train and test sets, we used only test sets from SPOT-1D-LM for comparisons in this section. We attempted to train the ProteinUnetLM model using the larger datasets from SPOT-1D-LM but surprisingly the results were suboptimal (as presented in Supplementary Table S1), so we decided to keep the model based on the TR10029 dataset.

The comparison of ProteinUnet2 with these three networks on 5 different test sets is presented in Table 2. First of all, ProteinUnetLM was statistically significantly better than NetSurfP-3.0 for all test sets in macro-AGM and SOV8 metrics, with relatively large effect sizes (d $>$ 0.3). ProteinUnet2 had also much better residue level metrics, excluding macro-AGM for TEST2018 for which NetSurfP-3.0 correctly predicted the rarest structure "I" (Supplementary Table S3). The main advantage of ProteinUnetLM over the SPOT-1D-LM network was better macro-AGM for all test sets, statistically significant (with a small effect size d [?] 0.1) for the three largest sets TEST2018, TEST2020, and TEST2020-HQ. It comes from the fact that ProteinUnetLM achieves much better results for rare structures B, G, and S without losing much accuracy for the frequent ones. For the same reason, SPOT-1D-LM had better Q8 on most of the test sets (excluding CASP12-FM), but as mentioned in Section 2.4, this metric is not appropriate for assessing SS8 prediction.

Table 2 . The comparison of **macro-AGM** and **Q8**at the **residue level** , and **SOV8** at the **sequence level** , on 5 test sets for ProteinUnetLM vs NetSurfP-3.0, ProtT5Sec, and SPOT-1D-LM. The best results for each dataset are boldfaced. The green shading of sequence level scores denotes the statistical significance that ProteinUnetLM has a better mean with standard deviations (SD), p-values, and Cohen's effect size (d) given below the score.

Test set	Model	Macro-AGM	Macro-AGM	SOV8	Q8
		Residue level	**Sequence level (±SD/p-value/d)**	**Sequence level (±SD/p-value/d)**	**Residue level**
TEST2018	ProteinUnetLM	0.728	**0.728** ±0.182	**0.760** ±0.144	0.756
	NetSurfP-3.0	0.693	0.644 ±0.196/p=2e-5/d=0.443	0.695 ±0.171/p=2e-5/d=0.411	0.716
	ProtT5Sec	**0.800**	0.688 ±0.193/p=2e-5/d=0.212	0.748 ±0.153/p=3e-4/d=0.080	0.749
	SPOT-1D-LM	0.725	0.710 ±0.197/p=0.004/d=0.096	0.757 ±0.161/p=0.414/d=0.019	**0.764**
TEST2020	ProteinUnetLM	0.667	**0.599** ±0.200	**0.648** ±0.173	0.683
	NetSurfP-3.0	0.718	0.509 ±0.204/p=2e-5/d=0.447	0.588 ±0.192/p=2e-5/d=0.325	0.662

Test set	Model	Macro-AGM	Macro-AGM	SOV8	Q8
	ProtT5Sec	**0.761**	0.568 ±0.200/p=2e-5/d=0.156	0.638 ±0.179/p=0.002/d=0.055	0.681
	SPOT-1D-LM	0.661	0.569 ±0.208/p=2e-5/d=0.149	0.643 ±0.181/p=0.140/d=0.028	**0.693**
TEST2020-HQ	ProteinUnetLM	**0.678**	**0.623** ±0.192	0.685 ±0.167	0.689
	NetSurfP-3.0	0.640	0.529 ±0.179/p=2e-5/d=0.501	0.625 ±0.178/p=2e-5/d=0.347	0.657
	ProtT5Sec	0.668	0.595 ±0.186/p=5e-4/d=0.144	0.681 ±0.165/p=0.585/d=0.021	0.688
	SPOT-1D-LM	0.674	0.601 ±0.189/p=0.008/d=0.116	**0.687** ±0.165/p=0.761/d=-0.012	**0.701**
Neff1-2020	ProteinUnetLM	**0.653**	**0.585** ±0.216	**0.653** ±0.202	0.687
	NetSurfP-3.0	0.635	0.506 ±0.229/p=2e-4/d=0.348	0.573 ±0.202/p=2e-4/d=0.391	0.678
	ProtT5Sec	0.652	0.584 ±0.208/p=0.991/d=0.021	0.645 ±0.213/p=0.469/d=0.040	0.689
	SPOT-1D-LM	0.647	0.574 ±0.227/p=0.368/d=0.029	0.652 ±0.193/p=0.900/d=0.007	**0.701**
CASP12-FM	ProteinUnetLM	**0.663**	**0.609** ±0.161	**0.650** ±0.114	**0.651**
	NetSurfP-3.0	0.563	0.541 ±0.149/p=0.001/d=0.127	0.574 ±0.121/p=5e-4/d=0.631	0.610
	ProtT5Sec	0.584	0.603 ± 0.155/p=0.736/d=0.085	0.633 ±0.109/p=0.059/d=0.150	0.640
	SPOT-1D-LM	0.656	0.589 ±0.168/p=0.165/d=0.118	0.627 ±0.120/p=0.039/d=0.196	0.643

There were no statistically significant differences between ProteinUnetLM and SPOT-1D-LM in terms of SOV8, but in most cases (excluding TEST2020-HQ) ProteinUnetLM had a better mean and smaller standard deviation (SD). The only advantage of ProtT5Sec over ProteinUnetLM was a correct prediction of the rarest structure "I" that highly improved the macro-AGM at the residue level for TEST2018 and TEST2020. ProteinUnetLM was better than ProtT5Sec in all other aspects. ProteinUnetLM was not statistically significantly worse than competitors in any metric or dataset. The competitive results of ProteinUnetLM on Neff1-2020 (sequences without homologs) and CASP12-FM (free modeling targets) prove the abilities of the network to generalize well beyond the protein folds included in the training/validation sets

Comparison on CASP14

In the context of the recent success of AlphaFold2 in the CASP14 contest[28], it is necessary to compare our network with SS8 predictions derived from AlphaFold2 tertiary structures (using DSSP) submitted to that contest, in order to support the desirability of our work. This comparison is far from being fair as AlphaFold2 is a much bigger model trained on a much bigger dataset. Despite this, ProteinUnetLM was able to achieve better macro-AGM for 10 out of 30 sequences from the CASP14 dataset (Supplementary Table S4) with

better residue level AGM for the rare class G (Supplementary Table S5) and is not statistically significantly different than AlphaFold2 in that metric (Table 3). It supports the claim that ProteinUnetLM provides state-of-the-art results in terms of the AGM metric. AlphaFold2 dominated other metrics and structures. It has a much better SOV8 which confirms the abilities of this metric to evaluate the quality of tertiary structure prediction at the secondary structure level[45], and a much higher Q8.

Setting aside AlphaFold2, ProteinUnetLM dominated all other networks in terms of macro-AGM at the sequence level with relatively large effect sizes (d > 0.3) and achieved the highest SOV8 (statistically significantly better than ProteinUnet2 and NetSurfP-3.0). As for TEST2018 and TEST2020, ProtT5Sec was able to predict the rarest structure "I" (Supplementary Table S5), so it surpassed ProteinUnetLM in macro-AGM at the residue level; it turns out to be one of the most prominent features of the ProtT5Sec network. In terms of Q8, ProteinUnetLM only gave way to SPOT-1D-LM

Table 3 . The comparison of **macro-AGM** at the**sequence and residue level** , **SOV8** at the**sequence level,** and **Q8** at the **residue level**on CASP14 for ProteinUnetLM vs all other networks. The best results for each metric are **boldfaced** and the second best are*underlined* . The green/red shading of sequence level scores denotes the statistical significance that ProteinUnetLM has a better/worse mean with standard deviations (SD), p-values, and Cohen's effect size (d) given below the score.

Model	**Macro-AGM**	**Macro-AGM**	**SOV8**	**Q8**
	Residue level	**Sequence level (±SD/p-value/d)**	**Sequence level (±SD/p-value/d)**	**Residue level**
ProteinUnetLM	0.697	*0.658* ±0.152	*0.668* ±0.129	0.694
ProteinUnet2	0.643	0.584 ±0.139/p=6e-4/d=0.499	0.601 ±0.111/p=0.001/d=0.545	0.618
NetSurfP-3.0	0.655	0.538 ±0.162/p=2e-5/d=0.747	0.617 ±0.134/p=4e-5/d=0.379	0.661
SPOT-1D-LM	0.683	0.606 ±0.139/p=5e-4/d=0.347	0.662 ±0.138/p=0.344/d=0.044	*0.701*
SPOT-1D	0.680	0.621 ±0.170/p=0.068/d=0.307	0.639 ±0.168/p=0.325/d=0.189	0.683
ProtT5Sec	*0.766*	0.610 ±0.152/p=0.004/d=0.311	0.654 ±0.127/p=0.099/d=0.107	0.686
AlphaFold2	**0.830**	**0.717** ±0.106/p=0.042/d=-0.442	**0.748** ±0.110/p=0.001/d=-0.659	**0.726**

3.3.1 Analysis of chameleon sequences

To understand the differences between the networks on the CASP14 dataset, we applied the analysis of predictions for chameleon sequences (ChSeqs) – specific sequences of amino acids that are known to adopt different 3-class SS (H, E, C) in different unrelated proteins. This analysis is considered one of the most rigorous tests for SS predictors because the conformations of ChSeqs depend on non-local protein-specific interactions [54,55]. We searched CASP14 for all 4-element ChSeqs according to the database by [56] and created a CASP14-ChSeqs set containing 3202 such 4-element sequences and their associated SS (for the first element in the sequence) according to DSSP. In Supplementary Figure S2, we compared the numbers and types of mistakes made for CASP14-ChSeqs by all the networks. The largest number of mistakes and largest differences between networks were observed for the loop class. ProteinUnetLM mistook helix for coil (H - C) over 2x less often than ProteinUnet2, reaching a level similar to AlphaFold2. The biggest disadvantage of AlphaFold2 was the overprediction of helices instead of coils (with the highest number of C - H errors out of all networks) which is in line with the conclusions from[30]. MSA-based networks (ProteinUnet2 and SPOT-1D) made over 80 mistakes more than their LM-based counterparts which confirms the higher predictive

power of LM features for challenging chameleon sequences. ProteinUnetLM achieved 3rd best result after AlphaFold2 and SPOT-1D-LM, beating NetSurfP-3.0.

Running time comparison

For a comparison of running times of LM-based models, we used a laptop with Nvidia RTX 2080 Max-Q GPU and Intel i7-10750H CPU. In the prediction time, we take into account the time needed for feature generation and the time of inference of the networks for SS prediction only (i.e., excluding regression-based networks for generating other outputs of SPOT-1D-LM) using batch size 1. We do not take into account the time needed for program initialization, data loading, or saving the results on disk. We were unable to measure the inference time of NetSurfP-3.0 on the same computer, as the model is accessible only for online end-to-end prediction. However, we assumed that the inference time of NetSurfP-3.0 is 5.3x shorter than for SPOT-1D-LM, based on the information from article [26], this assumption was marked with an asterisk in Table 4 which presents the times.

The features calculation time for ProtTransT5-XL-U50 (ProteinUnetLM) and ESM-1b (NetSurfP-3.0) on GPU is similar, with ESM-1b being 1.5x faster on CPU. The features calculation time for SPOT-1D-LM is a sum of both which makes it around 2x longer. ProteinUnetLM has nearly 3x shorter inference time on CPU (3 s) than on GPU (8 s). This is because ProteinUnetLM is so lightweight that loading features from pLMs (1024 x 704 values) into GPU and retrieving the result takes longer than simply running the model on the CPU. It leads to the situation where the optimal approach is to generate features using GPU and to run inference on CPU. It makes the inference time around 7x shorter than for SPOT-1D-LM on GPU and around 66x shorter on CPU. It results in 38 s (152 ms per sequence) of prediction time which is on par with the estimated prediction time of NetSurfP-3.0 on GPU and 2.4 times shorter than SPOT-1D-LM on GPU. Additionally, ProteinUnetLM can be effectively used without GPU with a prediction time shorter than 3 s per sequence. It is worth adding that ProteinUnetLM can be additionally sped up without losing much accuracy by training without AA on input (Supplementary Table S1) if necessary

Table 4 . The comparison of running times for ProteinUnetLM, SPOT-1D-LM, and NetSurfP-3.0 on the TEST2018 set with 250 sequences.

Model	Features calculation time (s)	Features calculation time (s)	Inference time (s)	Inference time (s)	Pr
	GPU	CPU	GPU	CPU	G
ProteinUnetLM	35	693	8	3	38
SPOT-1D-LM	69	1155	22	199	91
NetSurfP-3.0	34	462	4*	38*	38

*Estimated running times based on the assumption that NetSurfP-3.0 is 5.3x faster than SPOT-1D-LM

GPU+CPUFeature calculation on GPU and inference on CPU

Conclusion

The Attention U-Net convolutional architecture of ProteinUnet was shown to predict SS8 much better when using features from protein language models than features from MSA (as in the previous ProteinUnet2 network). Our experiments suggest that it has at least as good prediction quality as SPOT-1D-LM while being much faster; as fast as NetSurfP-3.0 which achieves much worse results. It supports our hypothesis that state-of-the-art in SS8 prediction can be achieved without using LSTM networks. Additionally, ProteinUnetLM has better results than the ProtT5Sec classifier which suggests that our architecture provided a significant improvement over this simple fully-convolutional network. Our focus on the issue of imbalance in

SS8 prediction, i.e., by adjusting the loss function of the network, allowed ProteinUnetLM to achieve state-of-the-art results in AGM metric by providing results competitive with AlphaFold2, and by dominating all other networks on CASP14 dataset. ProteinUnetLM can be considered one of the most efficient (prediction time shorter than 200 ms per sequence) and effective (macro-AGM 0.653–0.829, SOV8 0.648–0.786, Q8 0.651–0.771, depending on the test set) networks for predicting rare secondary structures, such as 310-helix (G), beta-bridge (B) and high curvature loop (S) while maintaining high performance for other structures. It can be run even on computers without GPU, so it is an ideal solution for embedded chips, mobile devices, and low-end computers. To support the reproducibility of the research and to encourage the community to adopt our network, we shared models and a complete code (for both training and inference), and an easy-to-use web interface.

The only limitation of Attention U-Net in comparison to LSTMs is the limited size of the input sequence (i.e., 704 residues). However, such long sequences rarely occur in nature and they can be still predicted by Attention U-Net in fragments if necessary. Moreover, an additional performance boost can be achieved by training an ensemble of 10 ProteinUnetLM models in the way described in the ProteinUnet2 publication [31]. The architecture can be easily extended to predict torsion angles and protein features like half sphere exposure, accessible solvent area, or contact number, as presented in the first ProteinUnet publication [36]. We plan to extend ProteinUnetLM with those outputs to enhance the utility of our network. Also, we plan to train a larger version of ProteinUnetLM on much larger datasets to reach AlphaFold2 accuracy. We hope that the thesis stated in the title will inspire researchers to apply Attention U-Net architecture to eliminate less energy-efficient LSTMs in other domains beyond bioinformatics, such as the classification of electroencephalography signals [57].

Acknowledgments

We would like to thank Jędrzej Kozal and Michał Woźniak from Wrocław University of Science and Technology for sharing their computational resources.

Funding

Conflict of Interest: none declared.

References

1. Kabsch W, Sander C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features.*Biopolymers* . 1983;22:2577-2637.

2. Liu J, Rost B. Comparing function and structure between entire proteomes. *Protein Science* . 2001;10(10):1970-1979. doi:10.1110/ps.10101

3. McGuffin LJ, Bryson K, Jones DT. What are the baselines for protein fold recognition? *Bioinformatics* . 2001;17(1):63-72. doi:10.1093/bioinformatics/17.1.63

4. Meiler J, Baker D. Coupled prediction of protein secondary and tertiary structure. *Proc Natl Acad Sci U S A* . 2003;100(21):12105-12110. doi:10.1073/pnas.1831973100

5. Hung LH, Samudrala R. PROTINFO: secondary and tertiary protein structure prediction. *Nucleic Acids Research* . 2003;31(13):3296-3299. doi:10.1093/nar/gkg541

6. Myers JK, Oas TG. Preorganized secondary structure as an important determinant of fast protein folding. *Nat Struct Mol Biol* . 2001;8(6):552-558. doi:10.1038/88626

7. Gardy JL, Spencer C, Wang K, et al. PSORT-B: improving protein subcellular localization prediction for Gram-negative bacteria.*Nucleic Acids Research* . 2003;31(13):3613-3617. doi:10.1093/nar/gkg602

8. Van Domselaar GH, Stothard P, Shrivastava S, et al. BASys: a web server for automated bacterial genome annotation. *Nucleic Acids Research* . 2005;33(suppl_2):W455-W459. doi:10.1093/nar/gki593

9. Mewes HW, Frishman D, Mayer KFX, et al. MIPS: analysis and annotation of proteins from whole genomes in 2005. *Nucleic Acids Research* . 2006;34(suppl_1):D169-D172. doi:10.1093/nar/gkj148

10. Wishart DS, Case DA. Use of Chemical Shifts in Macromolecular Structure Determination. In: James TL, Dötsch V, Schmitz U, eds.*Methods in Enzymology* . Vol 338. Nuclear Magnetic Resonance of Biological Macromolecules Part A. Academic Press; 2002:3-34. doi:10.1016/S0076-6879(02)38214-4

11. Smolarczyk T, Roterman-Konieczna I, Stapor K. Protein Secondary Structure Prediction: A Review of Progress and Directions. *Current Bioinformatics* . 2020;15:90-107.

12. Chou PY, Fasman GD. Prediction of protein conformation.*Biochemistry* . 1974;13:222-245.

13. Garnier J, Osguthorpe DJ, Robson B. Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. *Journal of Molecular Biology* . 1978;120:97-120.

14. Lim VI. Algorithms for prediction of α-helical and β-structural regions in globular proteins. *Journal of Molecular Biology* . 1974;88:873-894.

15. Jiang Q, Jin X, Lee SJ, Yao S. Protein secondary structure prediction: A survey of the state of the art. *Journal of Molecular Graphics and Modelling* . 2017;76:379-402. 10.1016/j.jmgm.2017.07.015.

16. Rost B, Sander C. Improved prediction of protein secondary structure by use of sequence profiles and neural networks. *Proceedings of the National Academy of Sciences of the United States of America* . 1993;90:7558-7562.

17. Remmert M, Biegert A, Hauser A, Söding J. HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nature Methods* . 2012;9:173-175.

18. Jones DT. Protein secondary structure prediction based on position-specific scoring matrices. *Journal of Molecular Biology* . 1999;292:195-202.

19. Klausen MS, Jespersen MC, Nielsen H, Jensen KK, Jurtz VI. NetSurfP-2.0: Improved prediction of protein structural features by integrated deep learning. *Proteins: Structure, Function, and Bioinformatics* . 2019;87:520-527.

20. Hanson J, Paliwal K, Litfin T, Yang Y, Zhou Y. Improving prediction of protein secondary structure, backbone angles, solvent accessibility and contact numbers by using predicted contact maps and an ensemble of recurrent and residual convolutional neural networks. Valencia A, ed.*Bioinformatics* . 2019;35(14):2403-2410. doi:10.1093/bioinformatics/bty1006

21. Ofer D, Brandes N, Linial M. The language of proteins: NLP, machine learning & protein sequences. *Computational and Structural Biotechnology Journal* . 2021;19:1750-1758. doi:10.1016/j.csbj.2021.03.022

22. Heinzinger M, Elnaggar A, Wang Y, et al. Modeling aspects of the language of life through transfer-learning protein sequences. *BMC Bioinformatics* . 2019;20(1):723. doi:10.1186/s12859-019-3220-8

23. Elnaggar A, Heinzinger M, Dallago C, et al. ProtTrans: Towards Cracking the Language of Lifes Code Through Self-Supervised Deep Learning and High Performance Computing. *IEEE Trans Pattern Anal Mach Intell* . Published online 2021:1-1. doi:10.1109/TPAMI.2021.3095381

24. Rives A, Meier J, Sercu T, et al. Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences. *Proc Natl Acad Sci USA* . 2021;118(15):e2016239118. doi:10.1073/pnas.2016239118

25. Vig J, Madani A, Varshney LR, Xiong C, socher richard, Rajani N. {BERT}ology Meets Biology: Interpreting Attention in Protein Language Models. In: *International Conference on Learning Representations* . ; 2021. https://openreview.net/forum?id=YWtLZvLmud7

26. Høie MH, Kiehl EN, Petersen B, et al. NetSurfP-3.0: accurate and fast prediction of protein structural features by protein language models and deep learning. *Nucleic Acids Research* . 2022;50(W1):W510-W515. doi:10.1093/nar/gkac439

27. Singh J, Paliwal K, Litfin T, Singh J, Zhou Y. Reaching alignment-profile-based accuracy in predicting protein secondary and tertiary structural properties without alignment. *Sci Rep* . 2022;12(1):7607. doi:10.1038/s41598-022-11684-w

28. Jumper J, Evans R, Pritzel A, et al. Highly accurate protein structure prediction with AlphaFold. *Nature* . 2021;596(7873):583-589. doi:10.1038/s41586-021-03819-2

29. Kryshtafovych A, Schwede T, Topf M, Fidelis K, Moult J. Critical assessment of methods of protein structure prediction (CASP)—Round XIV. *Proteins: Structure, Function, and Bioinformatics* . 2021;89(12):1607-1617. doi:https://doi.org/10.1002/prot.26237

30. Stevens AO, He Y. Benchmarking the Accuracy of AlphaFold 2 in Loop Structure Prediction. *Biomolecules* . 2022;12(7):985. doi:10.3390/biom12070985

31. Stapor K, Kotowski K, Smolarczyk T, Roterman I. Lightweight ProteinUnet2 network for protein secondary structure prediction: a step towards proper evaluation. *BMC Bioinformatics* . 2022;23(1):100. doi:10.1186/s12859-022-04623-z

32. AlQuraishi M. ProteinNet: a standardized data set for machine learning of protein structure. *BMC Bioinformatics* . 2019;20(1):311. doi:10.1186/s12859-019-2932-0

33. Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat Methods* . 2021;18(2):203-211. doi:10.1038/s41592-020-01008-z

34. Kotowski K, Adamski S, Machura B, Zarudzki L, Nalepa J. Coupling nnU-Nets with Expert Knowledge for Accurate Brain Tumor Segmentation from MRI. In: Crimi A, Bakas S, eds. *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries* . Lecture Notes in Computer Science. Springer International Publishing; 2022:197-209. doi:10.1007/978-3-031-09002-8_18

35. Isensee F, Ulrich C, Wald T, Maier-Hein KH. Extending nnU-Net is all you need. Published online August 23, 2022. doi:10.48550/arXiv.2208.10791

36. Kotowski K, Smolarczyk T, Roterman-Konieczna I, Stapor K. ProteinUnet—An efficient alternative to SPIDER3-single for sequence-based prediction of protein secondary structures. *Journal of Computational Chemistry* . 2021;42(1):50-59. doi:https://doi.org/10.1002/jcc.26432

37. Vaswani A, Shazeer N, Parmar N, et al. Attention is All you Need. In: *Advances in Neural Information Processing Systems* . Vol 30. Curran Associates, Inc.; 2017. Accessed September 14, 2022. https://papers.nips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html

38. Oktay O, Schlemper J, Folgoc LL, et al. Attention U-Net: Learning Where to Look for the Pancreas. In: *Medical Imaging with Deep Learning* . ; 2018. https://openreview.net/forum?id=Skft7cijM

39. Rao R, Meier J, Sercu T, Ovchinnikov S, Rives A. Transformer protein language models are unsupervised structure learners. Published online December 15, 2020:2020.12.15.422761. doi:10.1101/2020.12.15.422761

40. Chicco D, Jurman G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics* . 2020;21(1):6. doi:10.1186/s12864-019-6413-7

41. Chicco D, Totsch N, Jurman G. The Matthews correlation coefficient (MCC) is more reliable than balanced accuracy, bookmaker informedness, and markedness in two-class confusion matrix evaluation. *BioData Mining* . 2021;14(1):13. doi:10.1186/s13040-021-00244-z

42. Abhishek K, Hamarneh G. Matthews Correlation Coefficient Loss for Deep Convolutional Networks: Application to Skin Lesion Segmentation. Published online February 20, 2021. Accessed September 18, 2022. http://arxiv.org/abs/2010.13454

43. Kingma DP, Ba J. Adam: A Method for Stochastic Optimization. In: Bengio Y, LeCun Y, eds. *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings* . ; 2015. http://arxiv.org/abs/1412.6980

44. Batuwita R, Palade V. Adjusted geometric-mean: a novel performance measure for imbalanced bioinformatics datasets learning. *J Bioinform Comput Biol* . 2012;10(04):1250003. doi:10.1142/S0219720012500035

45. Liu T, Wang Z. SOV_refine: A further refined definition of segment overlap score and its significance for protein structure similarity.*Source Code Biol Med* . 2018;13:1. doi:10.1186/s13029-018-0068-7

46. Stapor K, Ksieniewicz P, Garcia S, Woźniak M. How to design the fair experimental classifier evaluation. *Applied Soft Computing* . 2021;104:107219. doi:10.1016/j.asoc.2021.107219

47. Cohen J. *Statistical Power Analysis for the Behavioral Sciences* . 2nd Edition. Routledge; 1988.

48. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* . ; 2016:770-778. doi:10.1109/CVPR.2016.90

49. Hochreiter S, Schmidhuber J. Long short-term memory. *Neural computation* . 1997;9(8):1735-1780.

50. Wang X, Girshick R, Gupta A, He K. Non-local Neural Networks. In:*2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* . ; 2018:7794-7803. doi:10.1109/CVPR.2018.00813

51. Drozdzal M, Vorontsov E, Chartrand G, Kadoury S, Pal C. The Importance of Skip Connections in Biomedical Image Segmentation. In: Carneiro G, Mateus D, Peter L, et al., eds. *Deep Learning and Data Labeling for Medical Applications* . Lecture Notes in Computer Science. Springer International Publishing; 2016:179-187. doi:10.1007/978-3-319-46976-8_19

52. Weytjens H, De Weerdt J. Process Outcome Prediction: CNN vs. LSTM (with Attention). In: Del Río Ortega A, Leopold H, Santoro FM, eds.*Business Process Management Workshops* . Lecture Notes in Business Information Processing. Springer International Publishing; 2020:321-333. doi:10.1007/978-3-030-66498-5_24

53. Li B, Zhou E, Huang B, et al. Large scale recurrent neural network on GPU. In: *2014 International Joint Conference on Neural Networks (IJCNN)* . ; 2014:4062-4069. doi:10.1109/IJCNN.2014.6889433

54. Jacoboni I, Martelli PL, Fariselli P, Compiani M, Casadio R. Predictions of protein segments with the same aminoacid sequence and different secondary structure: A benchmark for predictive methods.*Proteins: Structure, Function, and Bioinformatics* . 2000;41(4):535-544. doi:10.1002/1097-0134(20001201)41:4<535::AID-PROT100>3.0.CO;2-C

55. Saravanan KM, Selvaraj S. Performance of secondary structure prediction methods on proteins containing structurally ambivalent sequence fragments. *Peptide Science* . 2013;100(2):148-153. doi:10.1002/bip.22178

56. Ghozlane A, Joseph AP, Bornot A, de Brevern AG. Analysis of protein chameleon sequence characteristics. *Bioinformation* . 2009;3(9):367-369.

57. Kotowski K, Stapor K, Ochab J. Deep Learning Methods in Electroencephalography. In: Tsihrintzis GA, Jain LC, eds. *Machine Learning Paradigms: Advances in Deep Learning-Based Technological Applications* . Learning and Analytics in Intelligent Systems. Springer International Publishing; 2020:191-212. doi:10.1007/978-3-030-49724-8_8