

A hybrid deep learning and handcrafted feature approach for the prediction of protein structural class

Rached Yagoubi, Abdelouahab Moussaoui,
Mohamed Bachir Yagoubi

Abstract

The knowledge of the protein structural class is one of the most important sources of information related to protein structure or that about function analysis and drug design. But researchers still face difficulties to predict the protein structural class when it is a question about low-similarity sequences. In this paper, we propose to make the prediction using a hybrid deep learning method and handcrafted features instead of shallow classifier. We input only nine features, mostly from predicted secondary structure information, to a feed-forward deep neural network. The latter will automatically explore and extend those features through many layers and discover the representations needed for classification. The obtained results, when applying the jackknife test on two low-similarity benchmark datasets (25PDB and FC699), proved to be very significant. After comparing our method to others, it has turned out that using deep learning methods affords satisfactory performance in the field of protein structural class prediction.

Keywords: Protein structural class, Deep Learning, feed-forward deep neural network, predicted secondary structure information.

MSC 2010: 97R40, 92B20, 68T05, 92D20.

1 Introduction

The importance of the protein structural class is displayed in multiple domains. For instance, in the field of protein structure, function analy-

sis, drug design, and a lot of other biomedical applications [1], [2]. Protein structural class was introduced firstly by Levitt and Chothia [3] in 1976. The study they led was related to the polypeptide chain topologies. It consisted in studying a dataset composed of 31 globular proteins. The results yielded to the categorization of the known structure protein domains into four structural classes: *all* - α , *all* - β , α/β and $\alpha + \beta$ classes. The *all* - α classes represent structures that consist of mainly α -helices, and *all* - β classes represent structures that consist of mainly β -strands. The α/β and $\alpha + \beta$ classes contain both α -helices and β -strands, where the α/β class includes mainly parallel β -sheets, and the $\alpha + \beta$ class includes anti-parallel β -sheets. Since then, many other studies in the field emerged based on the four main classes. However, due to the rapid advance of sequencing technology, genomics, and proteomics, the number of newly discovered protein sequences has grown exponentially. This growth has led to a substantial gap between the number of sequence-known and the number of structure-known proteins. As a result of such inquiry, efforts have been made to reduce this gap, by setting newly developed methods that allow fast and accurate determination of the protein structure classes. In the same line of thought and to perform the structural class prediction, we generally need to follow two steps. The first one is linked to the transformation of the Amino Acid sequences into fixed-length feature vectors, whereas the second step consists of the feeding of the feature vectors to a classification algorithm.

In a parallel way during the last three decades, many efforts have followed trying to improve the accuracy of the prediction. In that, multiple sequence features representing protein sequences have been applied. We may cite the amino acid composition (AAC) [4], pseudo amino acid composition (PseAA) [5], polypeptide composition [6], functional domain composition [7], predicted secondary structure information [8], [9], and PSI-BLAST profile [10]. In these terms, a series of classification algorithms have been used to make the prediction. As example, we can name Bayesian classification [11], information discrepancy [12], fuzzy clustering [13], rough sets [14], logistic regression [15], k-nearest neighbors [16], Fisher's linear discriminate algorithm [17],

support vector machine (SVM) [18], [19], [2], classifier ensembles [20], and Convolutional Neural Network [21] .

The year 2006 has witnessed the emergence of deep learning, also called hierarchical learning, on behalf of machine learning research [22]. The principal aim of deep learning methods is to learn feature hierarchies. This learning will be achieved through features from the higher levels of the hierarchy already composed of lower-level features [23]. We may add that deep learning methods are seen as a clue in finding solutions to problems that had displayed resistance to the artificial intelligence community for a long period [24]. For instance, deep learning has proved its high efficiency in image recognition [25] and speech recognition [26], also in predicting activities in the fields of protein structure prediction [27], prediction of protein subcellular localization [28].

In this paper, we will exploit the power of deep neural networks for the prediction of protein structural class problem. Firstly, we have used a previous custom-designed feature vector from [9] that includes only nine hand-crafted features (features that are designed beforehand by human experts). After that, the vector is fed to a feed-forward deep neural network to automatically explore those features and discover the representations needed for classification. For more objectivity, we have chosen the jackknife cross-validation test as a means of evaluation to assess the proposed prediction method. The test is used on a couple of largely adopted datasets (25PDB and FC699) both related to be low-similarity benchmark datasets.

The paper is organized as follows. Section 2 gives a small description of the two datasets, the feature vector, the proposed feed-forward deep neural network architecture, and the different performance measure indexes. The obtained results are detailed in Section 3. Section 4 concludes the paper.

2 Materials and methods

2.1 Datasets

One has to know that the effect of sequence similarity on prediction accuracy weighs heavily, since less similarity makes prediction too dif-

difficult to perform [29], [30]. This is obvious regarding the definition of similarity by the amount of amino acids existing in the protein sequences which present resemblance after aligning that sequence with or to other sequences from a given dataset [30]. To test the performance of our approach, two widely used benchmark datasets were selected. The first is the 25PDB dataset [29]. It contains 1673 low similarity proteins (Lower than 25%). The other dataset is FC699 [9] including 858 sequences with similarity lower than 40%. Table 1 presents detailed compositions of each dataset. 1 presents detailed compositions of each dataset.

Table 1. The composition of 25PDB and FC699 datasets

Dataset	$all - \alpha$	$all - \beta$	α/β	$\alpha + \beta$	Total
25PDB	443	443	346	441	1673
FC699	130	269	377	82	858

2.2 Feature vector

In this work, we have used a custom-designed feature vector that includes 9 features previously defined on [9]. 8 out of them are relying on information obtained from secondary structure predicted with PSIPRED [31], [32]. The last feature is taken directly from the amino acid sequence. Table 2 gives a small description of each feature.

For example, if we have the following amino acid sequence:
MDPFLVLLHSVSSSLSSSELTELKYLCLGRVGRKRLERVQSGLD
LFSMLLEQNDLEPGHTELLRELLASLRRHDLLRRVDDFELEHH
HHHH

The predicted secondary structure sequence is:
CCHHHHHHHHHHCCCCCHHHHHHHHHHHHHHHHHCCCCCCCC
CCHHHHHHHHHHHHCCCCCHHHHHHHHHHHHCHHHHHHH
HHHHHHHHHHCCC

The obtained features are:
PSIPRED- $NC_{count_H}^6 = 1$
PSIPRED- $NC_{count_H}^8 = 1$

Table 2. The description of the used features

Feature	Description
PSIPRED- $NCOUNT_H^6$	Normalized count of α -helix segments (including at least 6 residues)
PSIPRED- $NCOUNT_H^8$	Normalized count of α -helix segments (including at least 8 residues)
PSIPRED- CMV_H^1	Composition moment vector of α -helix segments (first order)
PSIPRED- $NAvgSeg_H$	Normalized average length of α -helix segments
PSIPRED- CV_E	Composition vector of β -strand segments
PSIPRED- $NCOUNT_E^5$	Normalized count of β -strand segments (including at least 5 residues)
PSIPRED- $MaxSeg_E$	Length of the longest β -strand segment
PSIPRED- $NAvgSeg_E$	Normalized average length of β -strand segments
CV_{L---G}	Count of collocated amino acid pair (L, G) separated by 3 gaps

$PSIPRED-CMV_H^1 = 0.3543$
 $PSIPRED-NAvgSeg_H = 0.3659$
 $PSIPRED-CV_E = 0$
 $PSIPRED-NCOUNT_E^5 = 0$
 $PSIPRED-MaxSeg_E = 0$
 $PSIPRED-NAvgSeg_E = 0$
 $CV_{L---G} = 1$

2.3 Deep neural network

A standard neural network (NN) is composed of linked processors. The latter are called neurons. Each neuron produces a sequence of real-valued activations. In this part of the definition, we point at the fact that neuron activation may differ from a neuron to another. For

instance, Input neurons need sensors that perceive the environment; others get activated via well-arranged connections, with the help of former active neurons, favoring the process of activation. Another set of neurons may affect the environment due to their triggering activities. Looking for the exact weights that allow the NN display the wished comportment is considered as the credit assignment main function. In this context, such comportment will probably need long causal series of computational phases. The phases take on their charge to change the network aggregate activation, usually in a non-linear way. The essential about Deep Learning is to assign credit accurately through such phases [33].

After multiple experimental tests, the best feed-forward deep neural network model used is described in detail below. Figure 1 and the referred description offer an outline of the architecture used: as input data, we use only 9 features described previously.

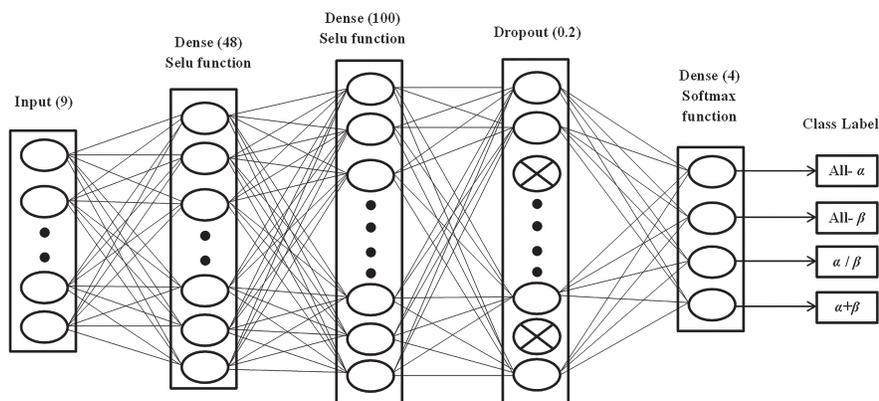


Figure 1. The proposed Deep Neural Network architecture

We have used a dense (48) layer (a fully-connected layer with 48 hidden units) which has been followed by a second dense (100) layer. In both layers, we have adopted the scaled exponential linear units (SELUs) function [34] as activation. In order to prevent our deep neural network from overfitting, we have applied a dropout [35] with a rate

of 0.2. Whereas the output layer is a fully-connected layer with four hidden units (dense (4)) that uses, as activation function, the softmax regression function, as we are dealing with a multiclass classification problem. Our model was run for 150 epochs (epoch = full pass over the training set), and the stochastic gradient descent (Adam) [36] was chosen as an optimization algorithm. Our method is implemented in Python programming language using Kears library under TensorFlow backend.

2.4 Performance measures

To obtain an accurate statistical prediction, three tests cannot be circumvented. It is the question of the independent dataset test, the subsampling test, and the jackknife test. The three cross-validation methods are largely used. They can easily examine the predictor effectiveness in a given practical situation [1], [30]. But according to the results, the jackknife test is considered the most objective. Its efficiency lays in the fact that it always gives the same result for a given dataset [8]. This is why the jackknife test has been chosen to test the validity of our method. Along the adopted test, every protein sequence in the dataset is picked out, in turn, as a test sample, when the predictor is trained by the rest of the protein sequences. In these terms and to measure the truthfulness and the strength of our predictor, five indexes have been selected for this purpose. The indexes in question are: sensitivity (*Sens*), precision (Prec), F-measure (*F*), Area Under ROC Curve (*AUC*) and Overall accuracy (*OA*). *AUC* is the area calculated under the Receiver Operating Characteristic (ROC). *AUC* is used to measure the average performance of the classification model. Its value is between 0 and 1. The value 1 is considered as perfect, whereas the value 0 seems to be untruthful. So, the more the value draws near 1 means that the classification model is more truthful [30]. Here are five formulas representing respectively the five indexes:

$$Sens = \frac{TP}{TP + FN};$$

$$\begin{aligned}
 \text{Prec} &= \frac{TP}{TP + FP}; \\
 F &= 2 \times \frac{\text{Prec} \times \text{Sens}}{\text{Prec} + \text{Sens}}; \\
 \text{AUC} &= \frac{1}{2} \times \left(\frac{TP}{TP + FN} + \frac{TN}{TN + FP} \right); \\
 \text{OA} &= \frac{TP + TN}{TP + FP + TN + FN}.
 \end{aligned}$$

In these formulas, the abbreviations *TP*, *FP*, *TN*, and *FN* stand respectively for the following: true positives, false positives, true negatives, and false negatives.

3 Results and discussion

3.1 Prediction performances of our method

The results of protein structural class prediction on the datasets 25PDB and FC699, obtained by jackknife test and performance evaluation indicators Sens, Prec, F-measure, AUC, and OA, are shown in Table 3.

Table 3. The prediction quality of our method on 25PDB and FC699 datasets

Dataset	Structural class	<i>Sens</i> (%)	<i>Prec</i> (%)	<i>F-measure</i>	<i>AUC</i>
25PDB	<i>all</i> - α	96.8	95.6	0.96	0.98
	<i>all</i> - β	96.3	93.4	0.95	0.95
	α/β	95.1	92.2	0.94	0.93
	$\alpha + \beta$	90.9	86.9	0.89	0.88
	OA	94.8			
FC699	<i>all</i> - α	98.9	99.1	0.99	0.99
	<i>all</i> - β	97.7	96.2	0.97	0.98
	α/β	92.8	93.4	0.93	0.98
	$\alpha + \beta$	97.8	94.5	0.96	0.94
	OA	96.8			

Respectively, the results of 94.8% and that of 96.8% represent the overall prediction accuracy of the datasets 25PDB and FC699. As seen in the same table after comparing the prediction accuracy of four protein structural classes, we have noticed that the performance evaluation indexes of $all - \alpha$ class are about 95.6% – 98.0% on the dataset 25PDB, whereas it was 98.9% – 99.1% on dataset FC699. On the other hand, the performance indexes of $all - \beta$ class are about 93.4% – 98.0% on both protein structural class datasets. This demonstrates that we archive a very high performance for the previous structural classes. The explanation laying behind the fact that $all - \alpha$ class and $all - \beta$ class have been so performing is that these proteins are helices or stands, in addition to being easily predictable since they are single and highly repetitive structures [37]. In parallel, the result of α/β and $\alpha + \beta$ classes have revealed so satisfactory. On 25PDB and FC699 datasets, the performances indexes of α/β are 92.2% – 98.0% and about 86.9% – 97.8% for $\alpha + \beta$ class.

3.2 Comparison with other prediction methods

In order to demonstrate the adopted method efficiency, we have compared the performance of the proposed method with the recently reported protein structural class prediction methods on the same datasets, as shown in Table 4.

Seeing the results displayed in Table 4 about the two datasets, our selected method proves to be the best performing on overall accuracies. It is striking, when analyzing the results, that the overall accuracies on 25PDB and FC699 datasets, respectively scored 1.4% and 0.3% higher than the previous top performed results. We have also noticed that the significant improvement on 25PDB is particularly made for the α/β and $\alpha + \beta$ classes, which tend to be difficultly predictable classes. In this way, we have improved the accuracy of α/β by 0.3% and by 1.3% on $\alpha + \beta$ class. Concerning FC699, our deep neural network overpasses the previous best-performing method by 8.8% on the $\alpha + \beta$ class.

Table 4. Performance comparison of different methods on two datasets

Dataset	Method	Accuracy (%)				
		<i>all - α</i>	<i>all - β</i>	<i>α/β</i>	<i>$\alpha + \beta$</i>	Overall
25PDB	[9]	92.6	80.1	74.0	71.0	79.7
	[18]	-	-	-	-	83.0
	[38]	94.1	87.1	84.1	74.4	85.0
	[39]	95.0	91.4	77.5	88.7	88.8
	[19]	98.9	89.6	85.6	78.9	88.4
	[2]	96.4	94.8	92.5	89.6	93.4
	[40]	95.6	89.5	88.1	87.0	90.1
	[30]	97.3	88.9	90.8	79.4	89.0
	[37]	95.7	97.7	94.8	84.4	93.1
	[20]	91.4	82.4	78.6	74.2	81.8
	This paper	96.8	96.3	95.1	90.9	94.8
FC699	[9]	-	-	-	-	87.5
	[41]	96.2	90.7	96.3	69.5	92.0
	[18]	-	-	-	-	94.5
	[38]	96.9	94.8	97.1	78.1	94.5
	[39]	98.5	98.1	97.6	81.7	96.5
	[19]	97.7	97.4	97.1	79.3	95.6
	[42]	96.9	89.2	89.4	89.0	90.4
	[20]	98.5	94.8	95.8	75.6	93.9
	This paper	98.9	97.7	92.8	97.8	96.8

4 Conclusion

Protein structural class prediction is considered as an important problem in bioinformatics. Despite all the attempts that have been done to solve this problem, the prediction accuracy improvement for low-similarity protein datasets is still a challenge. The main reason for the inadequate improvement is the unsatisfactory prediction accuracies for protein from α/β and $\alpha + \beta$ classes. In this paper, we have proposed to apply deep learning for the prediction of protein structural class. A 9-dimensional feature vector (8 features based on information extracted

from the secondary structure and one feature computed from the sequence) is used as input to a feed-forward deep neural network. The deep neural network and the jackknife test are employed to predict and evaluate the model over two benchmark datasets: 25PDB and FC699 datasets with sequence similarity lower than 25% and 40%, respectively. The results show that our approach outperforms the existing methods, especially for proteins from the α/β and $\alpha + \beta$ classes.

References

- [1] K.-C. Chou and C.-T. Zhang, "Prediction of Protein Structural Classes," *Critical Reviews in Biochemistry and Molecular Biology*, vol. 30, no. 4, pp. 275–349, jan 1995. [Online]. Available: <https://doi.org/10.3109/10409239509083488>
- [2] L. Zhang, L. Kong, X. Han, and J. Lv, "Structural class prediction of protein using novel feature extraction method from chaos game representation of predicted secondary structure," *Journal of Theoretical Biology*, vol. 400, pp. 1–10, 2016.
- [3] M. Levitt and C. Chothia, "Structural patterns in globular proteins," *Nature*, vol. 261, p. 552, jun 1976. [Online]. Available: <http://dx.doi.org/10.1038/261552a0><http://10.0.4.14/261552a0>
- [4] K.-C. Chou, "A novel approach to predicting protein structural classes in a (20-1)-D amino acid composition space," *Proteins: Structure, Function, and Genetics*, vol. 21, no. 4, pp. 319–344, apr 1995. [Online]. Available: <http://doi.wiley.com/10.1002/prot.340210406>
- [5] C. Chen, Y.-X. Tian, X.-Y. Zou, P.-X. Cai, and J.-Y. Mo, "Using pseudo-amino acid composition and support vector machine to predict protein structural class," *Journal of Theoretical Biology*, vol. 243, no. 3, pp. 444–448, dec 2006.
- [6] R. Y. Luo, Z. P. Feng, and J. K. Liu, "Prediction of protein structural class by amino acid and polypeptide composition," *European Journal of Biochemistry*, vol. 269, no. 17, pp. 4219–4225,

- sep 2002. [Online]. Available: <http://doi.wiley.com/10.1046/j.1432-1033.2002.03115.x>
- [7] K.-C. Chou and Y.-D. Cai, "Predicting protein structural class by functional domain composition," *Biochemical and Biophysical Research Communications*, vol. 321, no. 4, pp. 1007–1009, sep 2004.
 - [8] L. A. Kurgan, T. Zhang, H. Zhang, S. Shen, and J. Ruan, "Secondary structure-based assignment of the protein structural classes," *Amino Acids*, vol. 35, no. 3, pp. 551–564, oct 2008.
 - [9] L. Kurgan, K. Cios, and K. Chen, "SCPRED: Accurate prediction of protein structural class for sequences of twilight-zone similarity with predicting sequences," *BMC Bioinformatics*, vol. 9, no. 1, p. 226, 2008. [Online]. Available: <https://doi.org/10.1186/1471-2105-9-226>
 - [10] T. Liu, X. Zheng, and J. Wang, "Prediction of protein structural class for low-similarity sequences using support vector machine and PSI-BLAST profile," *Biochimie*, vol. 92, no. 10, pp. 1330–1334, 2010.
 - [11] Z.-X. Wang and Z. Yuan, "How good is prediction of protein structural class by the component-coupled method?" *Proteins: Structure, Function, and Genetics*, vol. 38, no. 2, pp. 165–175, feb 2000.
 - [12] L. Jin, W. Fang, and H. Tang, "Prediction of protein structural classes by a new measure of information discrepancy," *Computational Biology and Chemistry*, vol. 27, no. 3, pp. 373–380, jul 2003.
 - [13] H.-B. Shen, J. Yang, X.-J. Liu, and K.-C. Chou, "Using supervised fuzzy clustering to predict protein structural classes," *Biochemical and Biophysical Research Communications*, vol. 334, no. 2, pp. 577–581, 2005.
 - [14] Y. Cao, S. Liu, L. Zhang, J. Qin, J. Wang, and K. Tang, "Prediction of protein structural class with Rough Sets," *BMC Bioinformatics*, vol. 7, 2006.
 - [15] S. Jahandideh, P. Abdolmaleki, M. Jahandideh, and S. H. S. Hayatshahi, "Novel hybrid method for the evaluation of param-

- eters contributing in determination of protein structural classes,” *J Theor Biol*, vol. 244, 2007.
- [16] T.-L. Zhang, Y.-S. Ding, and K.-C. Chou, “Prediction protein structural classes with pseudo-amino acid composition: Approximate entropy and hydrophobicity pattern,” *Journal of Theoretical Biology*, vol. 250, no. 1, pp. 186–193, jan 2008.
- [17] J.-Y. Yang, Z.-L. Peng, and X. Chen, “Prediction of protein structural classes for low-homology sequences based on predicted secondary structure,” *BMC Bioinformatics*, vol. 11, no. 1, p. S9, 2010. [Online]. Available: <https://doi.org/10.1186/1471-2105-11-S1-S9>
- [18] L. Nanni, S. Brahnem, and A. Lumini, “Prediction of protein structure classes by incorporating different protein descriptors into general Chou’s pseudo amino acid composition,” *Journal of Theoretical Biology*, vol. 360, pp. 109–116, 2014.
- [19] J. Wang, C. Wang, J. Cao, X. Liu, Y. Yao, and Q. Dai, “Prediction of protein structural classes for low-similarity sequences using reduced PSSM and position-based secondary structural features,” *Gene*, vol. 554, no. 2, pp. 241–248, 2015.
- [20] S. Bankapur and N. Patil, “Enhanced Protein Structural Class Prediction using Effective Feature Modeling and Ensemble of Classifiers,” *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, p. 1, 2020.
- [21] S. Kaur, A. Sharma, and P. Singh, “Protein structural classes prediction based on convolutional neural network classifier with feature selection of hybrid pso-fa optimization approach,” *European Journal of Molecular & Clinical Medicine*, vol. 7, no. 10, pp. 252–265, 2020.
- [22] L. Deng and D. Yu, “Deep learning: Methods and applications,” pp. 197–387, 2013.
- [23] D. Erhan, A. Courville, and P. Vincent, “Why Does Unsupervised Pre-training Help Deep Learning ?” *Journal of Machine Learning Research*, vol. 11, pp. 625–660, 2010.

- [24] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, p. 436, may 2015. [Online]. Available: <http://dx.doi.org/10.1038/nature14539><http://10.0.4.14/nature14539>
- [25] C. Farabet, C. Couprie, L. Najman, and Y. Lecun, “Learning hierarchical features for scene labeling,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1915–1929, aug 2013.
- [26] G. Hinton, L. Deng, D. Yu, G. Dahl, A.-r. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. Sainath, and B. Kingsbury, “Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups,” *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 82–97, nov 2012.
- [27] M. Torrìsi, G. Pollastri, and Q. Le, “Deep learning methods in protein structure prediction,” *Computational and structural biotechnology journal*, vol. 18, pp. 1301–1310, jan 2020. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/32612753><https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7305407/>
- [28] J. J. Almagro Armenteros, C. K. Sønderby, S. K. Sønderby, H. Nielsen, and O. Winther, “Erratum: DeepLoc: prediction of protein subcellular localization using deep learning (Bioinformatics (Oxford, England) (2017)),” p. 4049, dec 2017.
- [29] L. A. Kurgan and L. Homaeian, “Prediction of structural classes for protein sequences and domains—Impact of prediction algorithms, sequence representation and homology, and test procedures on accuracy,” *Pattern Recognition*, vol. 39, no. 12, pp. 2323–2343, 2006.
- [30] Y. Liang and S. Zhang, “Predict protein structural class by incorporating two different modes of evolutionary information into Chou’s general pseudo amino acid composition,” *Journal of Molecular Graphics and Modelling*, vol. 78, pp. 110–117, 2017.
- [31] D. T. Jones, “Protein secondary structure prediction based on position-specific scoring matrices”¹Edited by G. Von Heijne,” *Journal of Molecular Biology*, vol. 292, no. 2, pp. 195–202, 1999.
- [32] D. W. A. Buchan and D. T. Jones, “The PSIPRED Protein

- Analysis Workbench: 20 years on,” *Nucleic Acids Research*, apr 2019. [Online]. Available: <https://doi.org/10.1093/nar/gkz297>
- [33] J. Schmidhuber, “Deep learning in neural networks: An overview,” *Neural Networks*, vol. 61, pp. 85–117, 2015.
- [34] G. Klambauer, T. Unterthiner, A. Mayr, and S. Hochreiter, “Self-normalizing neural networks,” in *Advances in Neural Information Processing Systems*, vol. 2017-Decem, jun 2017, pp. 972–981. [Online]. Available: <http://arxiv.org/abs/1706.02515>
- [35] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Dropout: A simple way to prevent neural networks from overfitting,” *Journal of Machine Learning Research*, vol. 15, pp. 1929–1958, 2014.
- [36] D. P. Kingma and J. L. Ba, “Adam: A method for stochastic optimization,” in *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, dec 2015. [Online]. Available: <http://arxiv.org/abs/1412.6980>
- [37] B. Yu, L. Lou, S. Li, Y. Zhang, W. Qiu, X. Wu, M. Wang, and B. Tian, “Prediction of protein structural class for low-similarity sequences using Chou’s pseudo amino acid composition and wavelet denoising,” *Journal of Molecular Graphics and Modelling*, vol. 76, pp. 260–273, 2017.
- [38] L. Kong and L. Zhang, “Novel structure-driven features for accurate prediction of protein structural class,” *Genomics*, vol. 103, no. 4, pp. 292–297, 2014.
- [39] J. Wang, Y. Li, X. Liu, Q. Dai, Y. Yao, and P. He, “High-accuracy prediction of protein structural classes using PseAA structural properties and secondary structural patterns,” *Biochimie*, vol. 101, pp. 104–112, 2014.
- [40] M. H. Olyaei, A. Yaghoubi, and M. Yaghoobi, “Predicting protein structural classes based on complex networks and recurrence analysis,” *Journal of Theoretical Biology*, vol. 404, pp. 375–382, 2016.
- [41] L. Kong, L. Zhang, and J. Lv, “Accurate prediction of protein

structural classes by incorporating predicted secondary structure information into the general form of Chou's pseudo amino acid composition," *Journal of Theoretical Biology*, vol. 344, pp. 12–18, 2014.

- [42] L. Liu, J. Cui, and J. Zhou, "A Novel Prediction Method of Protein Structural Classes Based on Protein Super-Secondary Structure," *Journal of Computer and Communications*, vol. 04, no. 15, pp. 54–62, nov 2016. [Online]. Available: <http://www.scirp.org/journal/doi.aspx?DOI=10.4236/jcc.2016.415005>

Rached Yagoubi, Abdelouahab Moussaoui,
Mohamed Bachir Yagoubi

Received August 26, 2020
Revised version October 30, 2021
Accepted January 11, 2022

Rached Yagoubi
Computer Science and Mathematics Laboratory,
Amar Telidji University of Laghouat,
Ghardaia Road, 03000 Laghouat, Algeria
E-mail: rached.yagoubi@gmail.com

Abdelouahab Moussaoui
Department of Computer Science,
Setif 1 University, El Bez Setif,
19000 Setif, Algeria
E-mail: moussaoui.abdel@gmail.com

Mohamed Bachir Yagoubi
Computer Science and Mathematics Laboratory,
Amar Telidji University of Laghouat,
Ghardaia Road, 03000 Laghouat, Algeria
E-mail: m.yagoubi@lagh-univ.dz