

# Improving Protein Secondary Structure Prediction Using an Ensemble of Recurrent and Convolutional Neural Networks with Evolutionary Profiles

M.M.M. Mufassirin

Department of Computer Science, Faculty of Applied Sciences, South Eastern University of Sri Lanka, Sri Lanka.

School of Information and Communication Technology, Griffith University, Australia.

[mufassirin@seu.ac.lk](mailto:mufassirin@seu.ac.lk)

Protein secondary structure prediction is a critical sub problem in computational biology and bioinformatics. The prediction of protein secondary structure has been extensively studied using various computational methods, including empirical and physics-based approaches and machine learning algorithms. With the advancement of deep learning methods, protein secondary structure classification accuracy has been substantially enhanced. Protein secondary structures are broadly classified into either 3-state (Q3) or 8-state (Q8) classes. This study proposes an approach that combines Convolutional Neural Network (CNN), bidirectional Long-Short-Term-Memory (BiLSTM), and evolutionary protein profile input features to improve secondary structure prediction accuracy. The proposed model was trained and validated using the DNSS2 dataset and tested on three independent test datasets, CB513, CASP11, and CASP12. The performance of the model was compared with five state-of-the-art approaches, and the impact of combining different input features on the model's performance was also evaluated. The proposed approach outperformed the state-of-the-art approaches, particularly for Q3 secondary structure prediction using PSSM, HMM, and 7PCP as input features. The ensemble of CNN and BiLSTM achieved the highest Q3 score of 85.35% and Q8 score of 75.51% on the test set. The approach presented in this study combines deep neural networks with optimized hyper-parameters and protein evolutionary profile features to improve secondary structure prediction accuracy, which is a novel contribution to the field. The proposed model significantly improved the accuracy of protein secondary structure prediction compared to five state-of-the-art methods. The approach can be useful in various fields, including drug discovery, protein engineering, and functional annotation.

Keywords: Convolutional Neural Network, Deep Learning, LSTM, Protein Secondary Structure, Protein Structure Prediction