UNIVERSITAS
GADJAH MADA
A Comparative Analysis of Protein Language Modelling for Protein Secondary Structure Prediction
Kayla Queenazima Santoso, Afiahayati S.Kom., M. Cs., Ph.D; Dr.Eng.Ir.Igi Ardiyanto, S.T., M. Eng., IPM., ASEAN Er
Universitas Gadjah Mada, 2025 | Diunduh dari http://etd.repository.ugm.ac.id/

REFERENCES

Adzhubei, A. A., Sternberg, M. J., & Makarov, A. A. (2013). Polyproline-II helix in proteins: structure and function. Journal of molecular biology, 425(12), 2100–2132. https://doi.org/10.1016/j.jmb.2013.03.018

Afiahayati, & Mulyana, S. (2008). Multiple sequence alignment menggunakan Hidden Markov Model [Seminar Nasional Informatika]. https://media.neliti.com/media/publications/173187-ID-multiple-sequence-alignment-menggunakan.pdf

Alanazi, W., Meng, D., & Pollastri, G. (2025). Porter 6: Protein Secondary Structure Prediction by Leveraging Pre-Trained Language Models (PLMs). International Journal of Molecular Sciences, 26(1), 130. https://doi.org/10.3390/ijms26010130

Ali, S., Chourasia, P., & Patterson, M. (2023). When Protein Structure Embedding Meets Large Language Models. Genes, 15(1), 25. https://doi.org/10.3390/genes15010025

amino-acid residue in a polypeptide. (2008). In International Union of Pure and Applied Chemistry (IUPAC) eBooks. https://doi.org/10.1351/goldbook.a00279

Anfinsen CB, Haber E, Sela M, White FH. (1961). The kinetics of formation ofnative ribonuclease during oxidation of the reduced polypeptide chain.Proc Natl Acad Sci USA, 47, 1309-1314.

Asgari E, Mofrad MRK (2015) Continuous Distributed Representation of Biological Sequences for Deep Proteomics and Genomics. PLOS ONE 10(11): e0141287. https://doi.org/10.1371/journal.pone.0141287

Brandes, N., Ofer, D., Peleg, Y., Rappoport, N., & Linial, M. (2022). ProteinBERT:

UNIVERSITAS
GADJAH MADA
A Comparative Analysis of Protein Language Modelling for Protein Secondary Structure Prediction
Kayla Queenazima Santoso, Afiahayati S.Kom., M. Cs., Ph.D; Dr.Eng.Ir.Igi Ardiyanto, S.T., M. Eng., IPM., ASEAN Er
Universitas Gadjah Mada, 2025 | Diunduh dari http://etd.repository.ugm.ac.id/

a universal deep-learning model of protein sequence and function. Bioinformatics, 38(8), 2102–2110. https://doi.org/10.1093/bioinformatics/btac020

Celik, I., Yadav, R., Duzgun, Z., Albogami, S., El-Shehawi, A. M., Fatimawali, Idroes, R., Tallei, T. E., & Emran, T. B. (2021). Interactions of the Receptor Binding Domain of SARS-CoV-2 Variants with hACE2: Insights from Molecular Docking Analysis and Molecular Dynamic Simulation. Biology, 10(9), 880. https://doi.org/10.3390/biology10090880

Chen, Y., Chen, G., & Chen, C. Y. (2024). PSSP-MFFNet: A Multifeature Fusion Network for Protein Secondary Structure Prediction. ACS Omega, 9(5), 5985–5994. https://doi.org/10.1021/acsomega.3c10230

Chen, Y., Chen, G., & Chen, C. Y. (2024). MFTrans: A multi-feature transformer network for protein secondary structure prediction. International Journal of Biological Macromolecules, 267, 131311. https://doi.org/10.1016/j.ijbiomac.2024.131311

Cuff, J. A., & Barton, G. J. (1999). Evaluation and Improvement of Multiple Sequence Methods for Protein Secondary Structure Prediction. Proteins: Structure, Function, and Genetics, 34, 508–519.

Elnaggar, A., Essam, H., Salah-Eldin, W., Moustafa, W., Elkerdawy, M., Rochereau, C., & Rost, B. (2023). Ankh: Optimized Protein Language Model Unlocks General-Purpose Modelling. bioRxiv. https://doi.org/10.1101/2023.01.16.524265

Elnaggar, A., Heinzinger, M., Dallago, C., Rihawi, G., Wang, Y., Jones, L., Gibbs, T., Feher, T., Angerer, C., Steinegger, M., Bhowmik, D., & Rost, B. (2020). ProtTrans: Towards Cracking the Language of Life's Code Through Self-Supervised Deep Learning and High Performance Computing. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2007.06225

Goldberg, Y. (2014). word2vec Explained: deriving Mikolov et al.'s negative-sampling word-embedding method. arXiv preprint arXiv:1402.3722.

UNIVERSITAS
GADJAH MADA
A Comparative Analysis of Protein Language Modelling for Protein Secondary Structure Prediction
Kayla Queenazima Santoso, Afiahayati S.Kom., M. Cs., Ph.D; Dr.Eng.Ir.Igi Ardiyanto, S.T., M. Eng., IPM., ASEAN Er
Universitas Gadjah Mada, 2025 | Diunduh dari http://etd.repository.ugm.ac.id/

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.

Habibie, F. (2021). GA-based feature selection for protein secondary structure prediction [Thesis]. Universitas Gadjah Mada.

Ho, C., Huang, Y., Chen, T., Lo, C., & Lo, W. (2021). Discovering the Ultimate Limits of Protein Secondary Structure Prediction. Biomolecules, 11(11), 1627. https://doi.org/10.3390/biom11111627

Houlsby, N., Giurgiu, A., Jastrzebski, S., Morrone, B., De Laroussilhe, Q., Gesmundo, A., Attariyan, M., Gelly, S., (2019). Parameter-efficient transfer learning for NLP.

Hu, B., Xia, J., Zheng, J., Tan, C., Huang, Y., Xu, Y., & Li, S. Z. (2022). Protein Language Models and Structure Prediction: Connection and Progression: A systematic survey. ArXiv, 2211.16742.

Hong, Y., Song, J., Ko, J., Lee, J., & Shin, W. H. (2022). S-Pred: protein structural property prediction using MSA transformer. Scientific reports, 12(1), 13891. https://doi.org/10.1038/s41598-022-18205-9

Huang, B., Kong, L., Wang, C., Ju, F., Zhang, Q., Zhu, J., Gong, T., Zhang, H., Yu, C., Zheng, W., & Bu, D. (2023). Protein Structure Prediction: Challenges, Advances, and the Shift of Research Paradigms. Genomics Proteomics & Bioinformatics, 21(5), 913–925. https://doi.org/10.1016/j.gpb.2022.11.014

Ismi, D. P., Pulungan, R., & Afiahayati, N. (2022). Deep learning for protein secondary structure prediction: Pre and post-AlphaFold. Computational and Structural Biotechnology Journal, 20, 6271–6286. https://doi.org/10.1016/j.csbj.2022.11.012

Ju, F., Zhu, J., Zhang, Q., Wei, G., Sun, S., Zheng, W., & Bu, D. (2021). Seq-SetNet: directly exploiting multiple sequence alignment for protein secondary structure prediction. Bioinformatics, 38(4), 990–996. https://doi.org/10.1093/bioinformatics/btab777

UNIVERSITAS
GADJAH MADA
A Comparative Analysis of Protein Language Modelling for Protein Secondary Structure Prediction
Kayla Queenazima Santoso, Afiahayati S.Kom., M. Cs., Ph.D; Dr.Eng.Ir.Igi Ardiyanto, S.T., M. Eng., IPM., ASEAN Er
Universitas Gadjah Mada, 2025 | Diunduh dari http://etd.repository.ugm.ac.id/

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. a. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., . . . Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. Nature, 596(7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2

Kalyan KS, Sangeetha S. S. (2020). ECNLP: A survey of embeddings in clinical natural language processing. J Biomed Inform. 101:103323. doi: 10.1016/j.jbi.2019.103323. Epub 2019 Nov 8. PMID: 31711972.

Kazm, A., Ali, A., & Hashim, H. (2024). Transformer Encoder with Protein Language Model for Protein Secondary Structure Prediction. Engineering Technology & Applied Science Research, 14(2), 13124–13132. https://doi.org/10.48084/etasr.6855

Kendrew, J. C., Dickerson, R. E., Strandberg, B. E., Hart, R. G., Davies, D. R., Phillips, D. C., and Shore, V. C. (1960) Structure of Myoglobin: A Three-Dimensional Fourier Synthesis at 2 Å. Resolution. Nature, 185(4711), 422–427.

Klausen, M. S., Jespersen, M. C., Nielsen, H., Jensen, K. K., Jurtz, V. I., Sønderby, C. K., Sommer, M. O. A., Winther, O., Nielsen, M., Petersen, B., & Marcatili, P. (2019). NetSurfP-2.0: Improved prediction of protein structural features by integrated deep learning. Proteins, 87(6), 520–527. https://doi.org/10.1002/prot.25674

Lin, Z., Akin, H., Rao, R., Hie, B., Zhu, Z., Lu, W., Smetanin, N., Verkuil, R., Kabeli, O., Shmueli, Y., Dos Santos Costa, A., Fazel-Zarandi, M., Sercu, T., Candido, S., & Rives, A. (2023). Evolutionary-scale prediction of atomic-level protein structure with a language model. Science (New York, N.Y.), 379(6637), 1123–1130. https://doi.org/10.1126/science.ade2574

Liu, H., Tam, D., Muqeeth, M., Mohta, J., Huang, T., Bansal, M., Raffel, C., (2022). Few-shot parameter-efficient fine-tuning is better and cheaper than in-context learning. ArXiv abs/2205.05638.

UNIVERSITAS
GADJAH MADA
A Comparative Analysis of Protein Language Modelling for Protein Secondary Structure Prediction
Kayla Queenazima Santoso, Afiahayati S.Kom., M. Cs., Ph.D; Dr.Eng.Ir.Igi Ardiyanto, S.T., M. Eng., IPM., ASEAN Er
Universitas Gadjah Mada, 2025 | Diunduh dari http://etd.repository.ugm.ac.id/

Lupo, U., Sgarbossa, D., & Bitbol, A. (2022). Protein language models trained on multiple sequence alignments learn phylogenetic relationships. Nature Communications, 13(1). https://doi.org/10.1038/s41467-022-34032-y

Ofer, D., Brandes, N., & Linial, M. (2021). The language of proteins: NLP, machine learning & protein sequences. Computational and Structural Biotechnology Journal, 19 1750–1758. https://doi.org/10.1016/j.csbj.2021.03.022

Peracha, O. (2023). PS4: A Next-Generation Dataset for Protein Single-Sequence Secondary Structure Prediction. BioTechniques, 76(2), 63–70. https://doi.org/10.2144/btn-2023-0024

Qin, Y., Chen, Z., Peng, Y., Xiao, Y., Zhong, T., and Yu, X. (2024). Deep learning methods for protein structure prediction. MedComm – Future Medicine, 3(3). https://doi.org/10.1002/mef2.96

Rao, R., Liu, J., Verkuil, R., Meier, J., Canny, J. F., Abbeel, P., Sercu, T., & Rives, A. (2021). MSA Transformer. Proceedings of the 38th International Conference on Machine Learning, PMLR 139. https://proceedings.mlr.press/v139/rao21a/rao21a.pdf

Rost, B. (2001). Review: Protein Secondary Structure Prediction Continues to Rise. Journal of Structural Biology, 134(2–3), 204–218. https://doi.org/10.1006/jsbi.2001.4336

Saharkhiz, S., Mostafavi, M., Birashk, A., Karimian, S., Khalilollah, S., Jaferian, S., Yazdani, Y., Alipourfard, I., Huh, Y. S., Farani, M. R., & Akhavan-Sigari, R. (2024). The State-of-the-Art Overview to Application of Deep Learning in Accurate Protein Design and Structure Prediction. Topics in Current Chemistry, 382(3). https://doi.org/10.1007/s41061-024-00469-6

Sak, H., W, A., Senior, & Beaufays, F. (2014). Long Short-Term Memory Based Recurrent Neural Network Architectures for Large Vocabulary Speech Recognition. arXiv (Cornell University). https://doi.org/10.48550/arxiv.1402.1128

UNIVERSITAS
GADJAH MADA
A Comparative Analysis of Protein Language Modelling for Protein Secondary Structure Prediction
Kayla Queenazima Santoso, Afiahayati S.Kom., M. Cs., Ph.D; Dr.Eng.Ir.Igi Ardiyanto, S.T., M. Eng., IPM., ASEAN Er
Universitas Gadjah Mada, 2025 | Diunduh dari http://etd.repository.ugm.ac.id/

Singh, J., Litfin, T., Paliwal, K., Singh, J., Hanumanthappa, A. K., & Zhou, Y. (2021). SPOT-1D-Single: improving the single-sequence-based prediction of protein secondary structure, backbone angles, solvent accessibility and half-sphere exposures using a large training set and ensembled deep learning. Bioinformatics (Oxford, England), 37(20), 3464–3472. https://doi.org/10.1093/bioinformatics/btab316

Smolarczyk, T., Roterman-konieczna, I., & Stapor, K. (2020). Protein Secondary Structure Prediction: A Review of Progress and Directions. Current Bioinformatics, 15, 90–107. https://doi.org/10.2174/1574893614666191017104639

Susanty, M., Mursalim, M. K. N., Hertadi, R., Purwarianti, A., & Rajab, T. L. (2024). Classifying alkaliphilic proteins using embeddings from protein language model. Computers in Biology and Medicine, 173, 108385. https://doi.org/10.1016/j.compbiomed.2024.108385

Wang, Q., Wang, B., Xu, Z., Wu, J., Zhao, P., Li, Z., Wang, S., Huang, J., & Cui, S. (2021). PSSM-Distil: Protein Secondary Structure Prediction (PSSP) on Low-Quality PSSM by Knowledge Distillation with Contrastive Learning. Proceedings of the AAAI Conference on Artificial Intelligence, 35(1), 617-625. https://doi.org/10.1609/aaai.v35i1.16141

Wang, L., Li, X., Zhang, H., Wang, J., Jiang, D., Xue, Z., & Wang, Y. (2025). A comprehensive review of protein language models. https://arxiv.org/html/2502.06881v1

Yang, W., Hu, Z., Zhou, L., & Jin, Y. (2021). Protein secondary structure prediction using a lightweight convolutional network and label distribution aware margin loss. Knowledge-Based Systems, 237, 107771. https://doi.org/10.1016/j.knosys.2021.107771

Yang, W., Liu, C., & Li, Z. (2023). Lightweight Fine-tuning a pretrained protein language model for protein secondary structure prediction. bioRxiv (Cold Spring Harbor Laboratory). https://doi.org/10.1101/2023.03.22.530066

UNIVERSITAS
GADJAH MADA
A Comparative Analysis of Protein Language Modelling for Protein Secondary Structure Prediction
Kayla Queenazima Santoso, Afiahayati S.Kom., M. Cs., Ph.D; Dr.Eng.Ir.Igi Ardiyanto, S.T., M. Eng., IPM., ASEAN Er
Universitas Gadjah Mada, 2025 | Diunduh dari http://etd.repository.ugm.ac.id/

Yang, W., Liu, Y., & Xiao, C. (2022). Deep metric learning for accurate protein secondary structure prediction. Knowledge-Based Systems, 242, 108356. https://doi.org/10.1016/j.knosys.2022.108356

Zhao, L., Li, J., Zhan, W., Jiang, X., & Zhang, B. (2024). Prediction of protein secondary structure by the improved TCN-BiLSTM-MHA model with knowledge distillation. Scientific Reports, 14(1). https://doi.org/10.1038/s41598-024-67403-0