

Medical Named Entity Recognition (MedNER): Deep learning model for recognizing medical entities (drug, disease) from scientific texts

Ullah Miah, M. Saef^a; Sulaiman, Junaida^b; Sarwar, Talha Bin^b; Islam, Saima Sharleen^c; Rahman, Mizanur^b; Haque, Md. Samiul^d

^a Faculty of Science and Technology, Aiub, Dhaka, Bangladesh

^b Faculty of Computing, Universiti Malaysia Pahang, Pekan, Malaysia

^c University of Trento, Department of Sociology and Social Research, Trento, Italy

^d Institute of Information Technology, University of Dhaka, Dhaka, Bangladesh

ABSTRACT

Medical Named Entity Recognition (MedNER) is an indispensable task in biomedical text mining. NER aims to recognize and categorize named entities in scientific literature, such as genes, proteins, diseases, and medications. This work is difficult due to the complexity of scientific language and the abundance of available material in the biomedical sector. Using domain-specific embedding and Bi-LSTM, we propose a novel NER model that employs deep learning approaches to improve the performance of NER on scientific publications. Our model gets 98% F1-score on a curated data-set of Covid-related scientific publications published in multiple web of science and pubmed indexed journals, significantly outperforming previous approaches deployed on the same data-set. Our findings illustrate the efficacy of our approach in reliably recognizing and classifying named entities (drug and disease) in scientific literature, opening the way for future developments in biomedical text mining.

KEYWORDS

Bi-LSTM; Deep learning in healthcare; Medical named entity recognition; Medical natural language processing; MedNER; Named entity recognition; NER; NLP

REFERENCES

- [1]P. Bose, S. Srinivasan, W. C. Sleeman IV, J. Palta, R. Kapoor, and P. Ghosh, "A survey on recent named entity recognition and relationship extraction techniques on clinical texts," *Applied Sciences*, vol. 11, no. 18, p. 8319, 2021.
- [2]B. Song, F. Li, Y. Liu, and X. Zeng, "Deep learning methods for biomedical named entity recognition: a survey and qualitative comparison," *Briefings in Bioinformatics*, vol. 22, no. 6, p. bbab282, 2021.
- [3]A. Graves and J. Schmidhuber, "Framewise phoneme classification with bidirectional lstm and other neural network architectures," *Neural networks*, vol. 18, no. 5-6, pp. 602–610, 2005.
- [4]H. Wei, M. Gao, A. Zhou, F. Chen, W. Qu, C. Wang, and M. Lu, "Named entity recognition from biomedical texts using a fusion attention-based bilstm-crf," *IEEE Access*, vol. 7, pp. 73 627–73 636, 2019.
- [5]M. Jiang, T. Sanger, X. Liu et al., "Combining contextualized embeddings and prior knowledge for clinical named entity recognition: evaluation study," *JMIR medical informatics*, vol. 7, no. 4, p. e14850, 2019.
- [6]K. Xu, Z. Zhou, T. Hao, and W. Liu, "A bidirectional lstm and conditional random fields approach to medical named entity recognition," in *International Conference on Advanced Intelligent Systems and Informatics*. Springer, 2017, pp. 355–365.
- [7]X. Chen, C. Ouyang, Y. Liu, and Y. Bu, "Improving the named entity recognition of chinese electronic medical records by combining domain dictionary and rules," *International Journal of Environmental Research and Public Health*, vol. 17, no. 8, p. 2687, 2020.
- [8]L. Li, J. Zhao, L. Hou, Y. Zhai, J. Shi, and F. Cui, "An attention-based deep learning model for clinical named entity recognition of Chinese electronic medical records," *BMC Medical Informatics and Decision Making*, vol. 19, no. 5, pp. 1–11, 2019.
- [9]M. S. U. Miah, J. Sulaiman, T. B. Sarwar, A. Naseer, F. Ashraf, K. Z. Zamli, and R. Jose, "Sentence boundary extraction from scientific literature of electric double layer capacitor domain: tools and techniques," *Applied Sciences*, vol. 12, no. 3, p. 1352, 2022.
- [10]UMP Library, "e-Resource UMP Lib," <https://login.ezproxy.ump.edu.my/login>, 2022. [Online]. Available: <https://login.ezproxy.ump.edu.my/login>
- [11]M. S. U. Miah, J. Sulaiman, and T. B. Sarwar, "Matrec: Material and process named entity recognition dataset for edlc," <https://data.mendeley.com/datasets/46p8vp9xts>, 2022.
- [12]J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. Minneapolis, Minnesota: Association for Computational Linguistics, Jun. 2019, pp. 4171–4186. [Online]. Available: <https://aclanthology.org/N19-1423>
- [13]I. Beltagy, K. Lo, and A. Cohan, "Scibert: A pretrained language model for scientific text," *arXiv preprint arXiv:1903.10676*, 2019.
- [14]L. Rokach, "Decision forest: Twenty years of research," *Information Fusion*, vol. 27, pp. 111–125, 2016.
- [15]T. K. Ho, "Random decision forests," vol. 1, 1995, pp. 278–282.
- [16]I. Rish et al., "An empirical study of the naive bayes classifier," in *IJCAI 2001 workshop on empirical methods in artificial intelligence*, vol. 3, no. 22, 2001, pp. 41–46.