

Our paper is conditionally accepted due to the ethical concern. In particular, the reviewers have raised concerns about data privacy. To address the issues, we have added an ethical statement in the paper. We also include the ethics meta review and our statement here for your convenience.

Ethics Meta Review

The authors should ensure that the data used in the presented work does not reveal private or personal information of individuals represented in the data. It should be clearly stated how personal identifying information has been removed at the appropriate point in the paper. Please refer to section B4 in the ACL ethics and responsible research charta¹.

Ethics Statement

All datasets used in this research are publicly available and are obtained according to each dataset's respective data usage policy. We avoid showing any direct excerpts of the data in the paper. We do not attempt to identify or deanonymize users in the data in any way during our research, thus preventing any bias in our methods toward any specific users.

More specifically, the proposed models are trained on the clinical notes of the public MIMIC-III database, which are already deidentified in accordance with Health Insurance Portability and Accountability Act (HIPAA) standards using structured data cleansing and date shifting. As such, all identifying data elements in HIPAA, including patient name, telephone number, address, and dates, are already removed (Johnson et al., 2016) from our training data to hinder attempts to retrieve personal information from our models. Similar to existing pre-trained and publicly available models for the clinical domain, i.e., ClinicalBERT (Huang et al., 2019) and BioClinicalBERT (Alsentzer et al., 2019), the proposed models serve as a resource to facilitate future research on clinical text.

References

Emily Alsentzer, John Murphy, William Boag, Weihung Weng, Di Jindi, Tristan Naumann, and Matthew McDermott. 2019. Publicly available clinical bert embeddings. In *Proceedings of the 2nd Clinical Natural Language Processing Workshop (ClinicalNLP)*, pages 72–78.

Kexin Huang, Jaan Altosaar, and Rajesh Ranganath. 2019. Clinicalbert: Modeling clinical notes and predicting hospital readmission. *arXiv preprint arXiv:1904.05342*.

Alistair EW Johnson, Tom J Pollard, Lu Shen, H Lehman Li-Wei, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi, and Roger G Mark. 2016. MIMIC-iii, a freely accessible critical care database. *Scientific data*, 3(1):1–9.

¹<https://aclrollingreview.org/responsibleNLPresearch>