

supervised learning. In the natural language process (NLP), the mainstream method is contextual methods [13, 14, 28]. To introduce contextual information, LM-LSTM [13] trains the model by predicting the next token using previous tokens. ELMo [28] trains the model by predicting a token using tokens from both directions. BERT [14] employs Transformer Encoder Layer and two pre-training tasks to train the model. Then, there are many variants of BERT, such as ALBERT [22], and StructBERT [34]. All of these models achieves better performance, and BERT-based methods have become the mainstream NLP method.

7 CONCLUSION

In this paper, we studied how to effectively represent EHR data for various downstream tasks. We first designed a novel architecture that is suitable for modeling EHR data, and we proposed pre-training for modeling EHR data. Then, we carefully devised three pre-training tasks to enable the model to handle various characteristics in EHR data, such as insufficiency and incompleteness. Extensive experimental results for four tasks demonstrated the effectiveness and robustness of the proposed model. We also introduced an interpretation method by sensitivity analysis and designed an interface to show the prediction results and interpretation.

ACKNOWLEDGMENTS

This work was supported by the National Key R&D Program of China (2019YFB2102103), the National Natural Science Foundation of China (Grant No. 92046010, 82161148011, 61872369), the Fundamental Research Funds for the Central Universities (Grant No. YWF-21-BJ-J-839), and Beijing Outstanding Young Scientist Program under (Grant No. BJJWZYJH012019100020098).

REFERENCES

- [1] Lei Jimmy Ba, Jamie Ryan Kiros, and Geoffrey E. Hinton. 2016. Layer Normalization. *CoRR* abs/1607.06450 (2016).
- [2] Tian Bai, Shanshan Zhang, Brian L. Egleston, and Slobodan Vucetic. 2018. Interpretable Representation Learning for Healthcare via Capturing Disease Progression through Time. In *KDD*. ACM, 43–51.
- [3] Inci M. Baytas, Cao Xiao, Xi Zhang, Fei Wang, Anil K. Jain, and Jiayu Zhou. 2017. Patient Subtyping via Time-Aware LSTM Networks. In *KDD*. ACM, 65–74.
- [4] Thomas A Buchanan, Anny H Xiang, et al. 2005. Gestational diabetes mellitus. *The Journal of clinical investigation* 115, 3 (2005), 485–491.
- [5] Mathilde Caron, Ishan Misra, Julien Mairal, Priya Goyal, Piotr Bojanowski, and Armand Joulin. 2020. Unsupervised Learning of Visual Features by Contrasting Cluster Assignments. In *NeurIPS*.
- [6] Zhengping Che, David C. Kale, Wenzhe Li, Mohammad Taha Bahadori, and Yan Liu. 2015. Deep Computational Phenotyping. In *KDD*. ACM, 507–516.
- [7] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey E. Hinton. 2020. A Simple Framework for Contrastive Learning of Visual Representations. In *ICML (Proceedings of Machine Learning Research, Vol. 119)*. PMLR, 1597–1607.
- [8] Yu Cheng, Fei Wang, Ping Zhang, and Jianying Hu. 2016. Risk Prediction with Electronic Health Records: A Deep Learning Approach. In *SDM*. SIAM, 432–440.
- [9] Edward Choi, Mohammad Taha Bahadori, Le Song, Walter F. Stewart, and Jimeng Sun. 2017. GRAM: Graph-based Attention Model for Healthcare Representation Learning. In *KDD*. ACM, 787–795.
- [10] Edward Choi, Mohammad Taha Bahadori, Jimeng Sun, Joshua Kulas, Andy Schuetz, and Walter F. Stewart. 2016. RETAIN: An Interpretable Predictive Model for Healthcare using Reverse Time Attention Mechanism. In *NIPS*. 3504–3512.
- [11] Sumit Chopra, Raia Hadsell, and Yann LeCun. 2005. Learning a Similarity Metric Discriminatively, with Application to Face Verification. In *CVPR (1)*. IEEE Computer Society, 539–546.
- [12] Kevin Clark, Minh-Thang Luong, Quoc V. Le, and Christopher D. Manning. 2020. ELECTRA: Pre-training Text Encoders as Discriminators Rather Than Generators. In *ICLR*. OpenReview.net.
- [13] Andrew M. Dai and Quoc V. Le. 2015. Semi-supervised Sequence Learning. In *NIPS*. 3079–3087.
- [14] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *NAACL-HLT (1)*. Association for Computational Linguistics, 4171–4186.
- [15] Alexey Dosovitskiy, Jost Tobias Springenberg, Martin A. Riedmiller, and Thomas Brox. 2014. Discriminative Unsupervised Feature Learning with Convolutional Neural Networks. In *NIPS*. 766–774.
- [16] Junyi Gao, Cao Xiao, Yasha Wang, Wen Tang, Lucas M. Glass, and Jimeng Sun. 2020. StageNet: Stage-Aware Neural Networks for Health Risk Prediction. In *WWW*. ACM / IW3C2, 530–540.
- [17] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross B. Girshick. 2020. Momentum Contrast for Unsupervised Visual Representation Learning. In *CVPR*. IEEE, 9726–9735.
- [18] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep Residual Learning for Image Recognition. In *CVPR*. IEEE Computer Society, 770–778.
- [19] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long Short-Term Memory. *Neural Comput.* 9, 8 (1997), 1735–1780.
- [20] Diederik P. Kingma and Jimmy Ba. 2015. Adam: A Method for Stochastic Optimization. In *ICLR (Poster)*.
- [21] Evangelia Kintiraki, Sophia Papakatsika, George Kotronis, Dimitrios G Goulis, and Vasilios Kotsis. 2015. Pregnancy-induced hypertension. *Hormones* 14, 2 (2015), 211–223.
- [22] Zhenzhong Lan, Mingda Chen, Sebastian Goodman, Kevin Gimpel, Piyush Sharma, and Radu Soricut. 2020. ALBERT: A Lite BERT for Self-supervised Learning of Language Representations. In *ICLR*. OpenReview.net.
- [23] Junyu Luo, Muchao Ye, Cao Xiao, and Fenglong Ma. 2020. HiTANet: Hierarchical Time-Aware Attention Networks for Risk Prediction on Electronic Health Records. In *KDD*. ACM, 647–656.
- [24] Fenglong Ma, Radha Chitta, Jing Zhou, Quanzeng You, Tong Sun, and Jing Gao. 2017. Dipole: Diagnosis Prediction in Healthcare via Attention-based Bidirectional Recurrent Neural Networks. In *KDD*. ACM, 1903–1911.
- [25] Ishan Misra and Laurens van der Maaten. 2020. Self-Supervised Learning of pre-training-Invariant Representations. In *CVPR*. IEEE, 6706–6716.
- [26] Phuoc Nguyen, Truyen Tran, Nilmini Wickramasinghe, and Svetla Venkatesh. 2017. DeepR: A Convolutional Net for Medical Records. *IEEE J. Biomed. Health Informatics* 21, 1 (2017), 22–30.
- [27] World Health Organization. 2017. *World Health Statistics 2017: Monitoring Health for the SDGs, Sustainable Development Goals*. World Health Organization. <https://books.google.com/books?id=JVXptAEACAAJ>
- [28] Matthew E. Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. 2018. Deep Contextualized Word Representations. In *NAACL-HLT*. Association for Computational Linguistics, 2227–2237.
- [29] Peter J Rousseeuw. 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics* 20 (1987), 53–65.
- [30] Nitish Srivastava, Geoffrey E. Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. 2014. Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* 15, 1 (2014), 1929–1958.
- [31] Laurens Van der Maaten and Geoffrey Hinton. 2008. Visualizing data using t-SNE. *Journal of machine learning research* 9, 11 (2008).
- [32] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is All you Need. In *NIPS*. 5998–6008.
- [33] Jingyuan Wang, Ze Wang, Jianfeng Li, and Junjie Wu. 2018. Multilevel Wavelet Decomposition Network for Interpretable Time Series Analysis. In *KDD*. ACM, 2437–2446.
- [34] Wei Wang, Bin Bi, Ming Yan, Chen Wu, Jiangnan Xia, Zuyi Bao, Liwei Peng, and Luo Si. 2020. StructBERT: Incorporating Language Structures into Pre-training for Deep Language Understanding. In *ICLR*. OpenReview.net.
- [35] Yan-Ting Wu, Chen-Jie Zhang, Ben Willem Mol, Andrew Kawai, Cheng Li, Lei Chen, Yu Wang, Jian-Zhong Sheng, Jian-Xia Fan, Yi Shi, et al. 2020. Early prediction of gestational diabetes mellitus in the Chinese population via advanced machine learning. *The Journal of Clinical Endocrinology & Metabolism* (2020).
- [36] Zhirong Wu, Yuanjun Xiong, Stella X. Yu, and Dahua Lin. 2018. Unsupervised Feature Learning via Non-Parametric Instance-level Discrimination. *CoRR* abs/1805.01978 (2018).
- [37] Mang Ye, Xu Zhang, Pong C. Yuen, and Shih-Fu Chang. 2019. Unsupervised Embedding Learning via Invariant and Spreading Instance Feature. In *CVPR*. Computer Vision Foundation / IEEE, 6210–6219.
- [38] Jun Zhang, Susan Meikle, and Ann Trumble. 2003. Severe maternal morbidity associated with hypertensive disorders in pregnancy in the United States. *Hypertension in pregnancy* 22, 2 (2003), 203–212.
- [39] Xianli Zhang, Buyue Qian, Shilei Cao, Yang Li, Hang Chen, Yefeng Zheng, and Ian Davidson. 2020. INPREM: An Interpretable and Trustworthy Predictive Model for Healthcare. In *KDD*. ACM, 450–460.
- [40] Xi Sheryl Zhang, Fengyi Tang, Hiroko H. Dodge, Jiayu Zhou, and Fei Wang. 2019. MetaPred: Meta-Learning for Clinical Risk Prediction with Limited Patient Electronic Health Records. In *KDD*. ACM, 2487–2495.