



















## 7 CONCLUSIONS

In this paper, we propose a general risk prediction framework PRIME, which can integrate prior medical knowledge into all the existing predictive models to improve the predictive performance. Specifically, we employ two state-of-the-art deep learning architectures—recurrent neural networks (RNN) and convolutional neural networks (CNN)—as the basic predictive models. To model the discrete and heterogeneous prior medical knowledge, posterior regularization technique is used. However, different from existing posterior regularization, we use a log-linear model to estimate the desired distributions of diseases. The benefit of the proposed approach is that it can automatically learn the weights for different prior medical knowledge. We validate the proposed framework on three real medical datasets. Experimental results show that the proposed PRIME outperforms existing risk prediction models. Finally, we qualitatively analyze the reasonableness of the weights learned by the proposed PRIME.

## ACKNOWLEDGMENTS

The authors gratefully thank **Ran Huo** who is an MD candidate from Southern Medical University for helpful discussions. The authors would like to thank the anonymous referees for their valuable comments and suggestions, and NVIDIA Corporation with the donation of the Titan Xp GPU. This work is supported in part by the US National Science Foundation under grants IIS-1553411, IIS-1747614, IIS-1218393 and IIS-1514204. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

## REFERENCES

- [1] Inci M. Baytas, Cao Xiao, Xi Zhang, Fei Wang, Anil K. Jain, and Jiayu Zhou. 2017. Patient Subtyping via Time-Aware LSTM Networks. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'17)*. 65–74.
- [2] Zhengping Che, Yu Cheng, Zhaonan Sun, and Yan Liu. 2016. Exploiting Convolutional Neural Network for Risk Prediction with Medical Feature Embedding. In *Proceedings of NIPS Workshop on Machine Learning for Health (NIPS-ML4HC'16)*.
- [3] Zhengping Che, Yu Cheng, Shuangfei Zhai, Zhaonan Sun, and Yan Liu. 2017. Boosting Deep Learning Risk Prediction with Generative Adversarial Networks for Electronic Health Records. In *Proceedings of the IEEE International Conference on Data Mining (ICDM'17)*. 787–792.
- [4] Zhengping Che, David Kale, Wenzhe Li, Mohammad Taha Bahadori, and Yan Liu. 2015. Deep Computational Phenotyping. In *Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'14)*. 507–516.
- [5] Yu Cheng, Fei Wang, Ping Zhang, and Jianying Hu. 2016. Risk Prediction with Electronic Health Records: A Deep Learning Approach. In *Proceedings of the 2016 SIAM International Conference on Data Mining (SDM'16)*. 432–440.
- [6] Kyunghyun Cho, Bart Van Merriënboer, Dzmitry Bahdanau, and Yoshua Bengio. 2014. On the Properties of Neural Machine Translation: Encoder-decoder Approaches. *arXiv preprint arXiv:1409.1259* (2014).
- [7] Edward Choi, Mohammad Taha Bahadori, Elizabeth Searles, Catherine Coffey, Michael Thompson, James Bost, Javier Tejedor-Sojo, and Jimeng Sun. 2016. Multi-layer representation learning for medical concepts. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'16)*. 1495–1504.
- [8] Edward Choi, Mohammad Taha Bahadori, Le Song, Walter F Stewart, and Jimeng Sun. 2017. GRAM: Graph-based Attention Model for Healthcare Representation Learning. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'17)*. 787–795.
- [9] Edward Choi, Mohammad Taha Bahadori, Jimeng Sun, Joshua Kulas, Andy Schuetz, and Walter Stewart. 2016. Retain: An Interpretable Predictive model for Healthcare Using Reverse Time Attention Mechanism. In *Proceedings of Advances in Neural Information Processing Systems (NIPS'16)*. 3504–3512.
- [10] Ronan Collobert, Jason Weston, Léon Bottou, Michael Karlen, Koray Kavukcuoglu, and Pavel Kuksa. 2011. Natural Language Processing (Almost) From Scratch. *Journal of Machine Learning Research (JMLR)* 12, Aug (2011), 2493–2537.
- [11] Luc Djoussé and J Michael Gaziano. 2008. Alcohol consumption and heart failure: a systematic review. *Current atherosclerosis reports* 10, 2 (2008), 117–120.
- [12] Kuzman Ganchev, Jennifer Gillenwater, Ben Taskar, et al. 2010. Posterior Regularization for Structured Latent Variable Models. *Journal of Machine Learning Research (JMLR)* 11, Jul (2010), 2001–2049.
- [13] Joyce C Ho, Joydeep Ghosh, Steve R Steinhilber, Walter F Stewart, Joshua C Denny, Bradley A Malin, and Jimeng Sun. 2014. Limestone: High-throughput Candidate Phenotype Generation via Tensor Factorization. *Journal of Biomedical Informatics* 52 (2014), 199–211.
- [14] Joyce C Ho, Joydeep Ghosh, and Jimeng Sun. 2014. Marble: High-throughput Phenotyping from Electronic Health Records via Sparse Nonnegative Tensor Factorization. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'14)*. 115–124.
- [15] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long Short-term Memory. *Neural computation* 9, 8 (1997), 1735–1780.
- [16] George Hripcsak and David J Albers. 2012. Next-generation Phenotyping of Electronic Health Records. *Journal of the American Medical Association (JAMA)* 307, 1 (2012), 117–121.
- [17] Zhiting Hu, Xuezhe Ma, Zhengzhong Liu, Eduard Hovy, and Eric Xing. 2016. Harnessing Deep Neural Networks with Logic Rules. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (ACL'16)*. 2410–2420.
- [18] Jau-Huei Lin and Peter J Haug. 2008. Exploiting Missing Clinical Data in Bayesian Network Modeling for Predicting Medical Problems. *Journal of Biomedical Informatics* 41, 1 (2008), 1–14.
- [19] Roderick JA Little and Donald B Rubin. 2014. *Statistical Analysis with Missing Data*. Vol. 333. John Wiley & Sons.
- [20] 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 *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'17)*. 1903–1911.
- [21] Fenglong Ma, Chuishi Meng, Houping Xiao, Qi Li, Jing Gao, Lu Su, and Aidong Zhang. 2017. Unsupervised Discovery of Drug Side-effects from Heterogeneous Data Sources. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'17)*. ACM, 967–976.
- [22] Franz Josef Och and Hermann Ney. 2002. Discriminative Training and Maximum Entropy Models for Statistical Machine Translation. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics (ACL'02)*. 295–302.
- [23] Rimma Pivovarov, David J Albers, Jorge L Sepulveda, and Noémie Elhadad. 2014. Identifying and Mitigating Biases in EHR Laboratory Tests. *Journal of Biomedical Informatics* 51 (2014), 24–34.
- [24] Qiuling Suo, Fenglong Ma, Giovanni Canino, Jing Gao, Aidong Zhang, Pierangelo Veltri, and Agostino Gnasso. 2017. A Multi-task Framework for Monitoring Health Conditions via Attention-based Recurrent Neural Networks. In *Proceedings of the AMIA 2017 Annual Symposium (AMIA'17)*.
- [25] Qiuling Suo, Fenglong Ma, Ye Yuan, Mengdi Huai, Weida Zhong, Jing Gao, and Aidong Zhang. 2017. Personalized Disease Prediction Using A CNN-Based Similarity Learning Method. In *Proceedings of The IEEE International Conference on Bioinformatics and Biomedicine (BIBM'17)*. 811–816.
- [26] Qiuling Suo, Fenglong Ma, Ye Yuan, Mengdi Huai, Weida Zhong, Jing Gao, and Aidong Zhang. 2018. Deep Patient Similarity Learning for Personalized Healthcare. *IEEE Transactions on NanoBioscience* (2018).
- [27] Fei Wang, Noah Lee, Jianying Hu, Jimeng Sun, and Shahram Ebadollahi. 2012. Towards Heterogeneous Temporal Clinical Event Pattern Discovery: A Convolutional Approach. In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'12)*. 453–461.
- [28] Xiang Wang, Fei Wang, Jianying Hu, and Robert Sorrentino. 2014. Exploring Joint Disease Risk Prediction. In *AMIA Annual Symposium Proceedings (AMIA'14)*. 1180–1187.
- [29] Ye Yuan, Guangxu Xun, Fenglong Ma, Qiuling Suo, Hongfei Xue, Kebin Jia, and Aidong Zhang. 2018. A Novel Channel-aware Attention Framework for Multi-Channel EEG Seizure Detection via Multi-view Deep Learning. In *Proceedings of the 2018 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI'18)*. IEEE, 206–209.
- [30] Matthew D Zeiler. 2012. ADADELTA: an adaptive learning rate method. *arXiv preprint arXiv:1212.5701* (2012).
- [31] Jiacheng Zhang, Yang Liu, Huanbo Luan, Jingfang Xu, and Maosong Sun. 2017. Prior Knowledge Integration for Neural Machine Translation Using Posterior Regularization. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (ACL'17)*, Vol. 1. 1514–1523.
- [32] Jiayu Zhou, Fei Wang, Jianying Hu, and Jieping Ye. 2014. From Micro to Macro: Data Driven Phenotyping by Densification of Longitudinal Electronic Medical Records. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'14)*. 135–144.