Using Electronic Health Records to Predict the Onset of Hypertension with LSTMs

Suparno Datta[1,2], Ariane Morassi Sasso[1,2], Nina Kiwit[1], Subhronil Bose[1], Jan Philipp Sachs[1,2], Girish Nadkarni[2,3,4], Riccardo Miotto[2,3] and Erwin Böttinger[1,2]

Background

Hypertension has long been recognized as an important **risk factor for cardiovascular disease** and mortality, having a prevalence of 22.1% worldwide and 49.6% in the United States alone [1, 2].

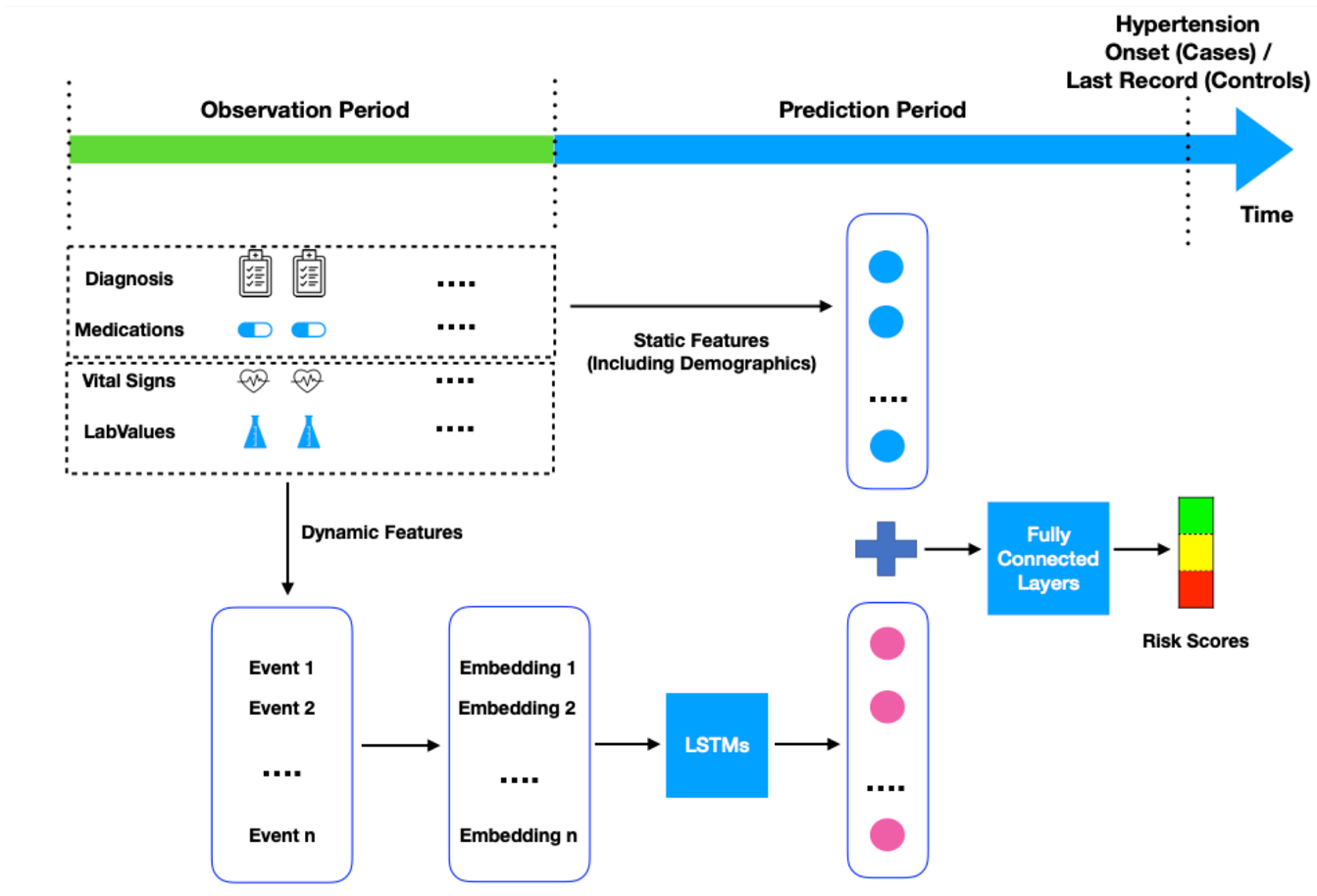
Aims

- Use long short-term memory (**LSTM**) networks that can deal with longitudinal electronic health records (**EHR**) data, to predict the onset of hypertension.
- Validate this approach against the best-performing model (**XGBoost**) reported in the literature [3].

Methods

Dataset. Data from **233,895 adult patients** in the Mount Sinai Health System in the United States (IRB-19-02369). The cohort was created using a validated phenotyping algorithm for hypertension [4] and was divided into **retrospective** (before 2017) and **prospective** subcohorts based on the date hypertension was diagnosed.

Models. The models were trained on the retrospective cohort using **cross-validation** and then applied to the prospective cohort to assess their performance. We report the model prediction performances based on data up to **a year before the onset of hypertension**. The figure below shows the design of the experiment and the architecture of the LSTMs.



Results

The **LSTM** networks achieved area under the receiver operating characteristic curve (**AUROC**) values of **0.98** in the **retrospective cohort** and **0.94** in the **prospective cohort** for a **time window of 1 year**. Meanwhile, the **XGBoost** model which was the same one used by Ye et al. [3] achieved **0.96 and 0.87** respectively, as seen in the table. **Features** related to **lipid disorders**, **renal disorders** and **type 2 diabetes** were identified as the **most informative** features for **predicting incident hypertension**.

Prediction Period	Model	Retrospective			Prospective		
		AUROC	AUPRC	F1	AUROC	AUPRC	F1
1 year	XGBoost	0.96	0.92	0.90	0.87	0.78	0.77
1 year	LSTM	**0.98**	**0.96**	**0.93**	**0.94**	**0.87**	**0.80**

Conclusion

These findings suggest that **deep learning models** based on **longitudinal EHR data** can help physicians to **identify patients at high risk** of developing hypertension with high discriminative performance. Those patients could be targeted by further intervention strategies such as lifestyle changes, aiming at a reduction of their risk of progression towards hypertension.

Future Work

Further steps will include both a **prospective clinical validation** and a validation of the models on EHR data from other health systems, preferably based **on standardized data** formats such as the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM).

References

[1] B. Zhou, J. Bentham, M. Di Cesare, H. Bixby, and G. Danaei et al., "Worldwide trends in blood pressure from 1975 to 2015: a pooled analysis of 1479 population-based measurement studies with 19·1 million participants," The Lancet, vol. 389, no. 10064, pp. 37–55, jan 2017.

[2] M. D. Chobufo, V. Gayam, J. Soluny, E. U. Rahman, S. Enoru, J. B. Foryoung, V. N. Agbor, A. Dufresne, and T. Nfoar, "Prevalence and control rates of hypertension in the usa: 2017–2018," International Journal of Cardiology Hypertension, vol. 6, p. 100044, 2020.

[3] C. Ye, T. Fu, S. Hao, Y. Zhang, and O. Wang et al., "Prediction of incident hypertension within the next year: Prospective study using statewide electronic health records and machine learning," Journal of Medical Internet Research, vol. 20, no. 1, p. e22, 2018.

[4] G. N. Nadkarni, O. Gottesman, J. G. Linneman, H. Chase, R. L. Berg, S. Farouk, R. Nadukuru, V. Lotay, S. Ellis, G. Hripcsak et al., "Development and validation of an electronic phenotyping algorithm for chronic kidney disease," in AMIA Annual Symposium Proceedings, vol. 2014. American Medical Informatics Association, 2014, p. 907.

[1]Hasso Plattner Institute, University of Potsdam, Potsdam, Germany

[2]Hasso Plattner Institute for Digital Health, Icahn School of Medicine at Mount Sinai, New York, NY, USA

[3]Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA

[4]Department of Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA

Contact: suparno.datta@hpi.de, ariane.morassi-sasso@hpi.de

Hasso Plattner Institute for Digital Health at Mount Sinai

Digital Engineering · Universität Potsdam