

# Leveraging Patient Similarity Analytics in Personalized Medical Decision Support System

Araek Tashkandi<sup>1</sup> and Lena Wiese<sup>2</sup>

<sup>1</sup> King Abdulaziz University, Faculty of Computing and Information Technology, 21589 Jeddah, Kingdom of Saudi Arabia. [asatashkandi@kau.edu.sa](mailto:asatashkandi@kau.edu.sa)

<sup>2</sup> Georg-August-University Goettingen, Institute of Computer Science, Goldschmidtstr. 7, 37077 Goettingen, Germany. [wiese@cs.uni-goettingen.de](mailto:wiese@cs.uni-goettingen.de)

## 1 Abstract

Patient similarity analytics harness the information wealth of electronic medical record (EMR) for supporting medical decision making. Finding a group of patients having similar features (for example, similar lab results or similar diagnoses), helps medical staff with treatment decisions or health predictions. There are different approaches for patient similarity metrics (PSMs) as described by [3]. We implemented patient similarity for mortality prediction as [2]. The method is based on cosine similarity that exploits similarities between ICU patients along multiple dimensions. We applied the proposed method on a real-world EMR data set MIMIC-III [1] containing both demographic data (like gender) as well as lab results. SQL is mainly used for implementing the similarity calculation. Our computation and analysis are conducted in MonetDB. Based on the requirement of calculating our PSM by SQL, our hypothesis is that column-oriented database management systems will outperform the row-oriented ones. To test this assumption, we conducted the same analyses in PostgreSQL.

We intend to use other PSMs that are mentioned by [3] and compare them to ours. For enhancing the computational efficiency, various technologies will be considered. Since “pairwise PSM computation is very much parallelizable” [2], such technologies are the big data analytic platforms such as Apache Hadoop. We will conduct an evaluation of the different implementations of patient similarity algorithms along with the selected technologies to achieve an efficient prediction and computation combination.

## References

1. Johnson, A.E., Pollard, T.J., Shen, L., Lehman, L.w.H., Feng, M., Ghassemi, M., Moody, B., Szolovits, P., Celi, L.A., Mark, R.G.: MIMIC-III, a freely accessible critical care database. *Scientific data* 3 (2016)
2. Lee, J., Maslove, D.M., Dubin, J.A.: Personalized mortality prediction driven by electronic medical data and a patient similarity metric. *PloS one* 10(5), e0127428 (2015)
3. Sharafoddini, A., Dubin, J.A., Lee, J.: Patient similarity in prediction models based on health data: a scoping review. *JMIR medical informatics* 5(1) (2017)