AI-BASED GENERATION OF MULTIFARIOUS MEDICAL DATA FOR *IN SILICO* CLINICAL TRIALS

Rajarajeswari Ganesan[1] , Frans N. van de Vosse[1], Wouter Huberts[1,2]

1. Department of Biomedical Engineering, Eindhoven University of Technology, The Netherlands;

2. Department of Biomedical Engineering, CARIM School for Cardiovascular Diseases, Maastricht University, The Netherlands

Background

For universal acceptance of a medical product, *In Silico* Clinical Trials need to test it on a sample population which captures the heterogeneity in the population. Heterogeneity connotes different age groups, gender, and inclusion of pathological conditions. Sample population should preferably include equal number of patients in each of these groups to be statistically significant. However, it is difficult to procure sufficiently large and properly balanced sample sets fulfilling these requirements. Deep Generative Models (DGMs) act as a great solution to all these challenges as they can create well-balanced synthetic datasets. Architecture of DGMs is dependent on the kind of data to be generated. In this study, we therefore developed a new DGM strategy to capture the facets of medical data.

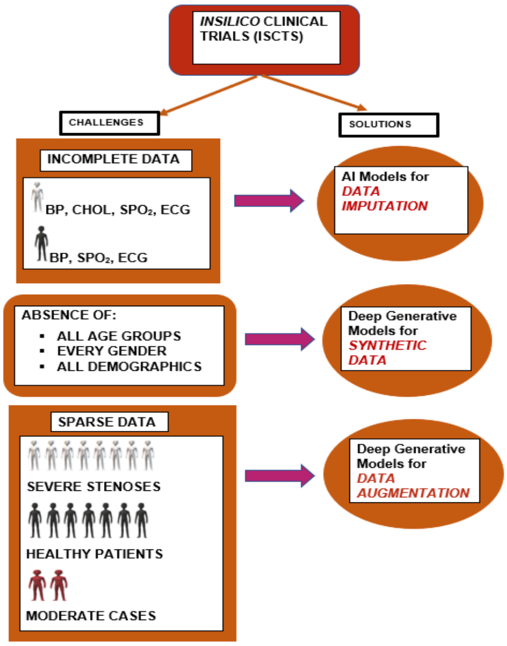

Figure 1: Challenges for ISCTs and their respective solutions

Method

Medical data can be classified into three classes: (a)Tabular Data (or One-Dimensional data) ; (b) Medical Images (or Two-Dimensional data) ; (c) Mesh geometries (or Three-Dimensional data). We identified the key requirements for each data type and the important patient-specific parameters for ISCTs. Each DGM framework has been designed based on these requirements to ensure that the generated data is clinically valid.

Results

Our observational and contextual assessment showed that class-based attribute learning in DGMs lead to creation of tenable synthetic samples.

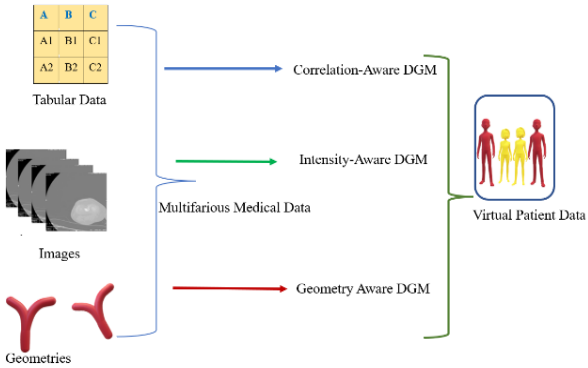

Figure 2: Overview of Deep Generative Models for different medical data

Correlation-aware[1] DGM intends to retain the dependencies between parameters for generation of plausible new samples. Intensity-aware[2] DGM captures the spatial relation in images for distinct formation of bone and soft tissues. Geometry-aware[3] DGM learns the coordinate (in)dependent features from mesh structures to generate new geometries.

Conclusion

Our results show that the DGMs such as Generative Adversarial Networks (GANs),Variational auto-encoders (VAEs) or a hybrid model should comprise facet-based learning potentiality for physiologically relevant sample generation. Further utilization of this surrogate dataset for AI models makes the models robust against privacy attacks.

References

1. Xu, L., & Veeramachaneni, K. (2018). Synthesizing tabular data using generative adversarial networks. arXiv preprint
2. Li, Yi, and Wei Ping. (2018). "Cancer metastasis detection with neural conditional random field." arXiv preprint
3.Chadebec, C., Mantoux, C., & Allassonnière, S. (2020). Geometry-aware Hamiltonian variational autoencoder arXiv:2010.11518.

Acknowledgements

This study was supported by the European Commission through the H2020 project "In Silico World: Lowering barriers to ubiquitous adoption of In Silico Trials" (topic SC1-DTH-06-2020, grant ID 101016503)."