Neuro-Symbolic Digital Twins for Precision and Predictive Public Health*

Gayo DIALLO [1,*,†]

[1] Team AHeaD, Bordeaux Population Health Inserm U1219, Univ. Bordeaux, F-33000, Bordeaux, France

Abstract

Public health prioritizes community medical conditions and population health factors. Promoting population health and preventing disease outbreaks and epidemics are the main goals. Targeting populations based on territorial factors, socio-economic and environmental determinants, and phenotypic profiles is essential for developing precise preventive or health promotion measures. Digital Twins (DTs) technology enables data acquisition, hypothesis generation, and in-silico experiments and comparisons. Thanks to Internet of Things and Artificial Intelligence, digital twins can collect a wider range of real-time data from various sources in addition to traditional data sources like Electronic Health Records. Thus, comprehensive simulations of physical entities, their functionality, and their evolution can be created and maintained. This position paper proposes using DT technology, Public Health instruments, knowledge graphs, and AI to enable Precision and Predictive Public Health for population health. In particular, it introduces Neuro-symbolic DTs, which combine semantic reasoning supported by a knowledge graph, deep-learning's predictive power, and a DT's agility to simulate public health interventions in a virtual environment.

Keywords

Precision Public Health, Neuro-symbolic AI, Digital Twins, Knowledge Graphs

1. A new vision for digital public health

Public health, according to the World Health Organization (WHO), prioritizes the protection of the overall health of communities (population level health), which can range in size from a single locale to a whole nation or global region. More specifically, it prioritizes the management of ongoing medical conditions in a community, and factors that impact health at the population level. A main objective is then to promote population wellness, as well as to track, control, and prevent disease outbreaks and epidemics [1]. A population-based approach considers interventions at all practicable levels, including the entire population within a community, the systems affecting the health of those populations, and/or the at-risk individuals and families within those populations.

In this frame, and in the context of the current ICT/digital era and Artificial Intelligence (AI) advances, the objective of precision public health is to capitalize on advanced data and

AI4DT&CP@IJCAI 2023: The First Workshop on AI for Digital Twins and Cyber-Physical Applications in conjunction with 32nd International Joint Conference on Artificial Intelligence, 19th - 25th August 2023, Macao, S.A.R

*Corresponding author.
gayo.diallo@u-bordeaux.fr (G. D.)
http://www.gayodiallo.org/ (G. D.)
0000-0002-9799-9484 (G. D.)

analytical techniques in order to expedite the resolution of pressing public health dilemmas worldwide and solve global public health issues efficiently. While predictive modelling is to identify the likelihood of future events, in public health, advanced predictive models are now utilized to anticipate health events and screen high-risk individuals [2].

We hypothesize that it is possible to go far beyond the current practices and achieve better results in terms of public health studies impact thanks to the exploitation of Digital Twins (DTs) technology that will be used to ingest and process near real-time heterogeneous health related data. More importantly, the use of prior knowledge encoded into knowledge graphs [3] that will be used to tailor DTs activities, enabling both logical and statistical based inferences and discovery of new insights. A knowledge graph (KG) is a representation of knowledge related to a domain in a machine-readable form [3]. It is a directed labelled graph in which the labels have well-defined meanings. It is composed of nodes, edges and labels.

The ultimate goal of relying on DTs enabled by KGs will be to be able to target populations according to territorial factors, socio-economic and environmental determinants. It will also be phenotypic profiling in order to develop specific (or precision) preventive or health promotion measures. Establishing a virtual system for managing disease outbreaks is a vital research opportunity. For instance, Deren et al. have suggested an integrated system as a smart city component, drawing from their experience with COVID-19 in China. The model utilizes multiple elements, such as a spatio-temporal patient database, cloud computing, and AI location technology, to effectively respond to outbreaks [4].

2. The neuro-symbolic DTs overall framework

The overall framework of the neuro-symbolic DTs, which enables precision and predictive public health, is depicted in figure 1. It comprises three main components: a data and knowledge layer, the DTs core engine and a services layer.

Thanks to the capability of ingesting and processing in a timely manner various health related data and non health per-se, such as socio-economic determinants of health (SDoH) [5], contextualized by relevant metadata encoded into various knowledge sources, it will be possible to address various use cases according to different end-users needs.

2.1. Neuro-Symbolic Digital Twins

In the context of health, DTs are virtual copies of human organs, tissues, cells, population dynamics, or micro-environments that adapt to online data and predict the future of the corresponding physical entity [6]. They enable the acquisition and discovery of new information, the generation and testing of new hypotheses, and the execution of in-silico experiments and comparisons. In addition to the more traditional data described in Electronic Health Records (EHRs), where applicable, and through the Internet of Things [7] and AI approaches, DTs can be implemented to collect much more real-world and real-time data from a wide range of sources, and thus can establish and maintain more comprehensive simulations of the physical entities, their functionality, and the changes they undergo over time [8].

Due to the ability to understand complex structures and engage in contextual reasoning using specific knowledge, Neuro-symbolic AI [9, 10] paradigm could be a valuable technology for

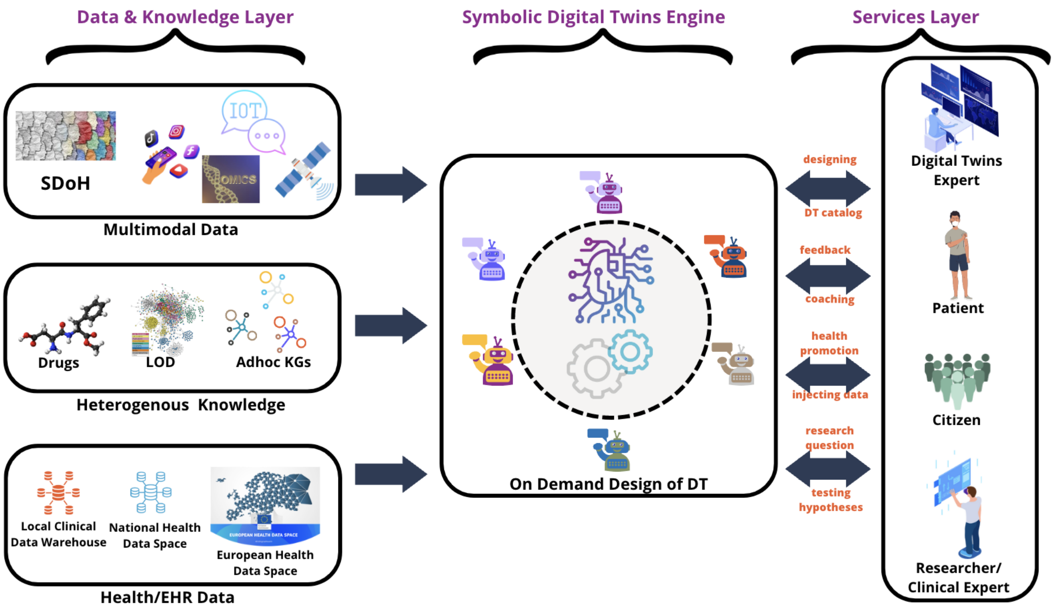

Figure 1: The overall framework of the neuro-symbolic Digital Twins based precision and predictive public health. The component at the centre of the framework, the Symbolic Digital Twins Engine, is the heart of the framework, with a repository of DTs that can be reused according to the questions to be addressed. The Data & Knowledge Layer is related to the different relevant types of data and knowledge sources, while the Services layer on the right are the main of use cases that address different end users need.

DTs. This capability facilitates comprehension of human language in DT contexts. The merging of neural networks and symbolic reasoning in neuro-symbolic AI provides an advantageous combination, empowering complex systems to incorporate domain knowledge and execute complex reasoning and decision-making based on said knowledge.

2.2. Components of the framework

The three main components of the framework are depicted in figure 1.

The core symbolic DTs engine. The Symbolic DTs engine, represented in the central panel of the figure, constitutes the core component of the framework. It includes a regularly updated repository of DTs, a set of tools for On-demand Twins designing, components for multimodal interaction (voice, text) with the DTs, etc. The engine is aware of the available DTs and is able to identify and suggest a suitable one to reuse or extend for a particular use case or research question.

The data and knowledge source layer. It is represented on the left panel of figure 1. In that part, there are the following data or knowledge sources: (i) the so-called non-traditional health data per-se are depicted at the top (Multimodal Data), such as that generated by social networks and online activities, wearables and sensors and data from satellites or mobile geolocalized devices (e.g. Call Data Records [11]); (ii) traditional health data that are used for public health activities, such as EHRs and clinical data warehouses that could be made available through platforms such as i2b2 [12], or, at a European level for instance, the European Health Data Space (EHDS) infrastructure for the secondary use of health data [13] ; (iii) knowledge generated either by Semantic Web technology [14] and the Linked Open Data cloud, or by standard health related coding systems (e.g., the International Classification of Diseases from the World Health Organization) and ad-hocs Knowledge Graphs developed for a specific purpose.

The services layer. The third main component of the framework, the services layer, is represented on the right part of figure 1. It is dedicated to providing a set of services towards different end-users of the framework. Four groups of end-users are identified: (i) DTs experts, (ii) patients, (iii) citizen or public and (iv) clinical experts or research scientists.

Digital Twins expert will be able to design and instantiate new DTs. Patients will be able to interact with their twins, either for giving (treatment) feedbacks or user profiling [15] and personalized coaching. Citizens will be subject to health promotion campaigns, for instance. Finally, a clinical expert or researcher in public health will be able to test hypotheses or simulate some effects of interventions [16].

3. Conclusion and Future Directions

Neuro-symbolic DTs offer a promising avenue for the advancement of public health research, allowing for a more effective implementation of digital solutions to various crucial concerns. These DTs have the potential to substantially improve pharmacovigilance and pharmaco-epidemiology including post-market drug monitoring, assessment of a treatment efficacy, but also diseases phenotyping, and other areas of public health research. In addition, the incorporation of neuro-symbolic DTs can improve health literacy among citizens [17]and promote data-driven decision-making through real-time reporting.

However, the adoption of neuro-symbolic DTs presents significant concerns and obstacles that require cautious considerations. An important aspect is the interaction between end-users and DTs, taking into account their diverse characteristics, which include expert designers, laypersons, and researchers. It is essential to investigate how natural language processing technologies, such as prompt engineering, text, and speech, can be utilized to facilitate seamless interaction with DTs for various user groups.

In addition, when utilizing AI-based prediction models, it is crucial to address the issue of supplying customized explanations for the outcomes and results generated by DTs. In public health applications, ensuring that users comprehend the reasoning behind the DT's predictions can increase transparency, trust, and acceptance of the technology.

It would also be interesting to investigate methods for empowering patients to actively contribute additional data to the DT system. Patients' participation in supplying the system with pertinent data has the potential to improve the accuracy and applicability of diagnostic tests in public health settings.

Ethical considerations also play a significant role in the public health planning application of digital twins technology. Identifying and addressing crucial ethical issues, such as equal access and averting bias, is of the utmost importance. It is essential to ensure that neuro-symbolic DTs are accessible securely to all individuals and communities, while mitigating both the risk of perpetuating inequalities and privacy violation.

In conclusion, neuro-symbolic digital twins hold a lot of promise for advancing public health research by allowing for a more efficient evaluation of a variety of crucial questions. To fully harness the potential of these technologies for the advancement of public health, it is essential to address challenges associated with user interaction, explanation of results, patient empowerment, and ethical considerations. Future research should concentrate on creating user-friendly interfaces, refining explanation techniques, examining patient engagement strategies, and establishing comprehensive ethical frameworks for the responsible use of neuro-symbolic DTs in public health planning.

Acknowledgments

The authors would like to thank the Agence Universitaire de la Francophonie (AUF) which partly funded this study, as part of the SAMIA project, agreement OPE 2023-0038.

References

[1] B. J. Turnock, Public health: what it is and how it works, sixth edition ed., Jones & Bartlett Learning, Burlington, Massachusetts, 2016.

[2] S. I. Ranapurwala, J. E. Cavanaugh, T. Young, H. Wu, C. Peek-Asa, M. R. Ramirez, Public health application of predictive modeling: an example from farm vehicle crashes, Injury Epidemiology 6 (2019) 31. URL: https://injepijournal.biomedcentral.com/articles/10.1186/s40621-019-0208-9. doi:10.1186/s40621-019-0208-9.

[3] A. Hogan, E. Blomqvist, M. Cochez, C. D'amato, G. D. Melo, C. Gutierrez, S. Kirrane, J. E. L. Gayo, R. Navigli, S. Neumaier, A.-C. N. Ngomo, A. Polleres, S. M. Rashid, A. Rula, L. Schmelzeisen, J. Sequeda, S. Staab, A. Zimmermann, Knowledge Graphs, ACM Computing Surveys 54 (2022) 1–37. URL: https://dl.acm.org/doi/10.1145/3447772. doi:10.1145/3447772.

[4] L. Deren, Y. Wenbo, S. Zhenfeng, Smart city based on digital twins, Computational Urban Science 1 (2021) 4. URL: http://link.springer.com/10.1007/s43762-021-00005-y. doi:10.1007/s43762-021-00005-y.

[5] D. L. Olstad, L. McIntyre, Reconceptualising precision public health, BMJ open 9 (2019) e030279. doi:10.1136/bmjopen-2019-030279.

[6] T. Sun, X. He, Z. Li, Digital twin in healthcare: Recent updates and challenges, DIGITAL HEALTH 9 (2023) 205520762211496. URL: http://journals.sagepub.com/doi/10.1177/20552076221149651. doi:10.1177/20552076221149651.

[7] I. Society, The Internet of Things (IoT): An Overview, 2015. URL: WWW.INTERNETSOCIETY.ORG.

[8] M. N. Kamel Boulos, P. Zhang, Digital Twins: From Personalised Medicine to Precision

Public Health, Journal of Personalized Medicine 11 (2021) 745. URL: https://www.mdpi.com/2075-4426/11/8/745. doi:10.3390/jpm11080745.
[9] M. Drancé, M. Boudin, F. Mougin, G. Diallo, Neuro-symbolic XAI for Computational Drug Repurposing:, in: Proceedings of the 13th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management, SCITEPRESS - Science and Technology Publications, Online Streaming, — Select a Country —, 2021, pp. 220–225. URL: https://www.scitepress.org/DigitalLibrary/Link.aspx?doi=10.5220/0010714100003064. doi:10.5220/0010714100003064.
[10] M. K. Sarker, L. Zhou, A. Eberhart, P. Hitzler, Neuro-symbolic artificial intelligence: Current trends, AI Communications 34 (2022) 197–209. URL: https://www.medra.org/servlet/aliasResolver?alias=iospress&doi=10.3233/AIC-210084. doi:10.3233/AIC-210084.
[11] E. Mutafungwa, F. Thiessard, P. Diallo, R. Gore, V. Jouhet, C. Karray, N. Kheder, R. Saddem, S. Ben Yahia, J. Hämäläinen, G. Diallo, Mobile Data as Public Health Decision Enabler: A Case Study of Cardiac and Neurological Emergencies, in: Data for Development (D4D) Challenge at NetMob 2015, Boston, USA, 2015.
[12] S. N. Murphy, M. E. Mendis, D. A. Berkowitz, I. Kohane, H. C. Chueh, Integration of clinical and genetic data in the i2b2 architecture, AMIA ... Annual Symposium proceedings. AMIA Symposium 2006 (2006) 1040.
[13] C. Safran, M. Bloomrosen, W. E. Hammond, S. Labkoff, S. Markel-Fox, P. C. Tang, D. E. Detmer, Toward a National Framework for the Secondary Use of Health Data: An American Medical Informatics Association White Paper, Journal of the American Medical Informatics Association 14 (2007) 1–9. URL: https://academic.oup.com/jamia/article-lookup/doi/10.1197/jamia.M2273. doi:10.1197/jamia.M2273.
[14] T. Berners-Lee, J. Hendler, O. Lassila, The Semantic Web. A New Form of Web Content That Is Meaningful to Computers Will Unleash a Revolution of New Possibilities, Scientific American 284 (2001) 1–5.
[15] P. Kostkova, G. Diallo, G. Jawaheer, User Profiling for Semantic Browsing in Medical Digital Libraries, in: S. Bechhofer, M. Hauswirth, J. Hoffmann, M. Koubarakis (Eds.), The Semantic Web: Research and Applications, volume 5021, Springer Berlin Heidelberg, Berlin, Heidelberg, 2008, pp. 827–831. URL: http://link.springer.com/10.1007/978-3-540-68234-9_67. doi:10.1007/978-3-540-68234-9_67.
[16] M. Pellegrino, G. Lombardo, S. Cagnoni, A. Poggi, High-Performance Computing and ABMS for High-Resolution COVID-19 Spreading Simulation, Future Internet 14 (2022) 83. URL: https://www.mdpi.com/1999-5903/14/3/83. doi:10.3390/fi14030083.
[17] I. Ouedraogo, B. M. J. Some, R. Benedikter, G. Diallo, Improving health literacy in rural Africa through mobile phones: a systematic literature review, Journal of Health Informatics in Africa 8 (2022) 26–31. URL: https://doi.org/10.12856/JHIA-2021-v8-i1-312. doi:10.12856/JHIA-2021-v8-i1-312.