Research proposal

Background

The number of research works published in the field of healthcare and machine learning (ML) is growing continuously and is estimated to be over 7500 papers in 2020. However, only less than 1% of those projects are making it to clinical practice [1]. Although many other industries such as finance, retail, insurance, energy utilities, etc have adopted and benefitted from ML applications, healthcare did not witness any similar successes in deploying ML models [2]. Among the reasons for this observation is the fact that healthcare data by its very nature is highly complex, high-dimensional, and of inconsistent quality, which is why it is usually not handled well by general-purpose commercial machine learning platforms [2]. Prior research shows that another difficult aspect of dealing with this data is the lack of any standardized protocols for data export [3]. Therefore, downstream systems need to take care of adapters capable of receiving data over customized protocols for different groups of medical devices. The complexity of healthcare data leads to its underuse for many important secondary tasks in the clinical setting.

In particular, one of the most underutilized types of resources in healthcare data today is high-speed physiological data [4]. It is collected by a variety of electronic sensors and devices and is generated as part of the patient care process. Examples of this data include electrocardiogram (ECG), electroencephalogram (EEG), pulse oximetry data, etc [5]. These data streams are helping clinicians with patient treatment and facilitating diagnosis, especially in the intensive care environment [6, 7]. The common practice in many hospitals is to drop the vast majority of such data after it is collected by the monitoring systems, even though it incapsulates highly valuable information that can help clinicians provide better care [8]. One of the reasons for this underutilization is the lack of open-source systems supporting real-time analysis of streaming physiological data [9]. The scarcity of such systems can be explained by the complexity of handling such data. Since it is originally collected for use by bedside clinicians and most medical institutions do not store or process this data for any secondary purposes [8], there have not been many developments in this area.

However, the literature indicates that there is general public support for the secondary use of physiological streaming data, which is based on the findings of a household survey that investigated Australian and Canadian citizens' perceptions of such physiological data capture and re-use [4]. Moreover, it has been shown that this data has the potential for many useful applications, especially if used by ML solutions that have been shown to be tremendously advantageous in improving the overall efficiency of healthcare systems while lowering the cost of care [7]. Several projects have already been proposed that utilize the physiological streaming waveform data to improve clinical care. A specific example is an ML model capable of predicting the probability of cardiac arrest for patients based on their physiological signals [10]. Another ML solution is analyzing ECG waveform data to classify heart rhythms [11]. A framework has been proposed that performs real-time analysis of physiological data to monitor people's health conditions [12]. It is also not only ML applications that benefit from this type of data. Automated clinical alerting and decision support platforms are also based on physiological data analysis [13].

The importance of handling physiological streaming data and the scarcity of available efficient tools have led to this research project, which fills the gap between the availability of underused highly insightful data and solutions that rely on it to improve clinical care. It proposes a platform that would enable a seamless integration of data consumers with a data source by encapsulating all of the inherent complexity of physiological streaming signals and providing a convenient application protocol interface.

Objectives for the research project

The aim of the research is to propose a solution to the real-time physiological data stream processing of clinical bedside signals. This goal consists of several objectives that include (1) researching the best combination of functionality based on the needs of downstream systems, (2) implementing a platform that supports several data consumers, and (3) testing the performance and verifying the robustness of the solution. Along with the system for streaming data processing, a general approach and framework need to be provided that would make the process of clinical model deployment easier and faster.

Although different data consumers might have a variety of unique data processing requirements, our hypothesis is that it is possible to combine all of the requirements to develop a single system that could be flexible and robust enough to serve all of the streaming data processing needs.

The research questions include but are not limited to discovering which features need to be included in the system through user studies, determining the interface, and identifying how to process, window, and align the data in a way that is robust and efficient.

Timeline

	Month 1	Month 2	Month 3	Month 4	Month 5	Month 6
Research of existing solutions to physiological data stream processing and identification of important features						
Implementation of a minimum viable solution satisfying the needs of one ML consumer						
Extension of the system to support several other ML workflows						
End-to-end system testing and identification of areas of improvement						
Profiling and optimizing the pipeline for better performance and increased security						
System deployment in the clinical environment of the Hospital for Sick Children (SickKids)						

Literature cited

[1] Drysdale, Erik & Dolatabadi, Elham & Chivers, Corey & Liu, Vincent & Saria, Such & Sendak, Mark & Wiens, Jenna & Brudno, Michael & Hoyt, Amelia & Mazwi, Mjaye & Mamdani, Muhammad & Singh, & Allen, Vanessa & Mcgregor, Carolyn & Ross, Heather & Szeto, Antonio & Anand, Amol & Verma, & Wang, Bo & Goldenberg, Anna. (2019). Implementing AI in healthcare. https://doi.org/10.13140/RG.2.2.30793.70241
[2] Arora, A., Nethi, A., Kharat, P., Verghese, V., Jenkins, G., Miff, S., Chowdhry, V., & Wang, X. (2019). ISTHMUS: Secure, Scalable, Real-time and Robust Machine Learning Platform for Healthcare. arXiv. https://doi.org/10.48550/arXiv.1909.13343
[3] N. Stylianides, M. D. Dikaiakos, G. Panayi and T. Kyprianou, "Intensive Care Window: Real time monitoring and analysis in the intensive care environment," 2009 9th International Conference on Information Technology and Applications in Biomedicine, 2009, pp. 1-4, https://doi.org/10.1109/ITAB.2009.5394437
[4] McGregor, C., Heath, J., & Choi, Y. (2015). Streaming Physiological Data: General Public Perceptions of Secondary Use and Application to Research in Neonatal Intensive Care. Studies in health technology and informatics, 216, 453–457. https://doi.org/10.3233/978-1-61499-564-7-453
[5] Sahoo SS, Jayapandian C, Garg G, Kaffashi F, Chung S, Bozorgi A. Heart beats in the cloud: Distributed analysis of electrophysiological "Big Data" using cloud computing for epilepsy clinical research. J Am Med Inform Assoc. 2014;21(2):263-71. https://doi.org/10.1136/amiajnl-2013-00215
[6] Goodwin AJ, Eytan D, Greer RW, Mazwi M, Thommandram A, Goodfellow SD, Assadi A, Jegatheeswaran A, Laussen PC. A practical approach to storage and retrieval of high-frequency physiological signals. Physiol Meas. 2020 Apr 20;41(3):035008. https://doi.org/10.1088/1361-6579/ab7cb5
[7] Mohd Javaid, Abid Haleem, Ravi Pratap Singh, Rajiv Suman, Shanay Rab, Significance of machine learning in healthcare: Features, pillars and applications, International Journal of Intelligent Networks, Volume 3, 2022, Pages 58-73, ISSN 2666-6030, https://doi.org/10.1016/j.ijin.2022.05.002
[8] D. Sow, A. Biem, J. Sun, J. Hu and S. Ebadollahi, "Real-time prognosis of ICU physiological data streams," 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology, 2010, pp. 6785-6788, https://doi.org/10.1109/IEMBS.2010.5625983
[9] J. R. Sutton, R. Mahajan, O. Akbilgic and R. Kamaleswaran, "PhysOnline: An Open Source Machine Learning Pipeline for Real-Time Analysis of Streaming Physiological Waveform," in IEEE Journal of Biomedical and Health Informatics, vol. 23, no. 1, pp. 59-65, Jan. 2019, https://doi.org/10.1109/JBHI.2018.2832610
[10] Tonekaboni, Sana & Mazwi, Mjaye & Laussen, Peter & Eytan, Danny & Greer, Robert & Goodfellow, Sebastian & Goodwin, Andrew & Brudno, Michael & Goldenberg, Anna. (2018). Prediction of Cardiac Arrest from Physiological Signals in the Pediatric ICU.
[11] Goodfellow, S.D., Goodwin, A., Greer, R., Laussen, P.C., Mazwi, M., & Eytan, D. (2018). Towards Understanding ECG Rhythm Classification Using Convolutional Neural Networks and Attention Mappings. MLHC.
[12] D. Apiletti, E. Baralis, G. Bruno and T. Cerquitelli, "Real-Time Analysis of Physiological Data to Support Medical Applications," in IEEE Transactions on Information Technology in Biomedicine, vol. 13, no. 3, pp. 313-321, May 2009, https://doi.org/10.1109/TITB.2008.2010702
[13] J. B. Smith, A. T. Reisner, S. Edla, J. Liu, S. Liddle and J. Reifman, "A platform for real-time acquisition and analysis of physiological data in hospital emergency departments," 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2014, pp. 2678-2681, https://doi.org/10.1109/EMBC.2014.6944174
[14] D. Eytan, D. Shubin, M. Zhang, D. Ehrmann, S. Goodfellow, “Time-efficient labelling framework for biological waveform data using semi-supervised learning and interactive visualization,” Machine Learning for Healthcare 2021.