This article has been accepted for publication, but has not been fully edited. Content may change prior to final publication.

A Framework for Edge-Assisted Healthcare Data Analytics using Federated Learning

Saqib Hakak and Suprio Ray
Faculty of Computer Science
University of New Brunswick
Fredericton, Canada
Email: {saqib.hakak, sray}@unb.ca

Wazir Zada Khan
Department of CS & IS
Jazan University
Jazan, Saudi Arabia
Email: wazirzadakhan@jazanu.edu.sa

Erik Scheme
Electrical and Computer Engineering
University of New Brunswick
Fredericton, Canada
Email: escheme@unb.ca

***Abstract*—With the emergence of wearable technology, IoT, and Edge computing, the nature of healthcare is rapidly shifting towards digital health aided by these ICT technologies. At the same time, consumer devices, such as smart, wearable fitness watches are gaining market share as a way to monitor physical activity and wellness. Despite these advances, and their ability to capture longitudinal behavioural patterns, these devices have yet to be fully leveraged within the healthcare system. If the user-generated data from such devices could be collected without compromising an individual's privacy, these insights could comprise part of a more holistic and preventative healthcare solution. In this article, we propose an Edge-assisted data analytics framework that uses Federated Learning to re-train local machine learning models using user-generated data. This framework could leverage pre-trained models to extract user-customized insights while preserving privacy and Cloud resources. We also identify some potential application scenarios and discuss research challenges to be explored within the proposed framework.**

***Index Terms*—Edge computing, Healthcare Analytics, Federated Learning, Wearable technology, Medical IoT**

I. Introduction

The constitution of the World Health Organization (WHO) in 1946 recognized healthcare as a fundamental human right. However, half of world's population still lacks access to basic healthcare services. According to a WHO report [4] 800 million people spend at least 10% of their budget on healthcare, and over 100 million people have been pushed into extreme poverty due to healthcare expenses. At the same time, the global population is aging rapidly and it is estimated that more than 1.5 billion people will be 65 years or older by 2050 [3]. Consequently, the leading causes of death are shifting towards chronic diseases and conditions, such as arthritis, cancer, COPD, diabetes, heart diseases and injury due to falls. It is not just the elderly, however, who are susceptible to chronic diseases; some of these conditions, particularly, obesity and diabetes are increasingly prevalent among younger populations. Diabetes affects nearly 1 in 10 American adults [5] and an increasing number of type 2 diabetes has been diagnosed among children. Correspondingly, the Centers for Disease Control (CDC) in the United States have stated that 75% of its annual healthcare spending is related to chronic diseases. As their prevalence continues to increase, management and prevention of chronic diseases is vitally important to the delivery and sustainability of quality healthcare services for our citizens.

Fortunately, the timely identification of symptoms and effective monitoring of known conditions can significantly mitigate the risks associated with chronic disease. Remote monitoring and telemedicine can drastically reduce re-admissions, emergency room visits, and their associated costs. According to a 2016 report [2], the preventable hospital readmission costs in the U.S. was $17 billion annually. The Internet of Things (IoT) can play a major role in reducing the burden on the acute care system and improving the quality of life of chronically ill patients. Medical IoT sensors can help monitor blood pressure and glucose levels, heart rate and body temperature. These sensors can send timestamped readings to data collection devices in an IT platform for persistent storage and analysis. Real-time monitoring involves running data analytics and machine learning (ML) algorithms on these data to detect patterns and automatically generate alarms when an anomaly is detected. Due to advances in Wireless Sensor Network (WSN) technology, declining prices and miniaturization of smart sensors, IoT technology is rapidly making inroads in many spheres of our lives. With the adoption and maturation of wearable fitness tracking devices such as the Fitbit, medical IoT is finding applications in wellness and health monitoring. Consequently, it is projected that the value of improved health outcomes due to remote disease monitoring and wellness management could be up to $1.6 trillion annually by 2025 [1].

While IoT technologies hold great promise in the support of affordable healthcare, globally, they are not without challenges. With the rapid adoption of medical IoT and fitness devices, there will likely be billions of such devices in the near future. These devices will generate high volumes of raw data at exceptional rates that may challenge current storage and computational capacity. Although today's Cloud computing platforms provide a model to support large scale data processing, their centralized computing framework requires the migration of all data to a remote data center, where the the data can be analyzed and processed with ML algorithms. This movement of data, especially sampled time-series sensor data, incurs significant latency and monetary costs due to "pay-as-you-go" business models. To address some of the limitations of the Cloud, Edge computing [10], [16] has been proposed as a new paradigm that extends traditional Cloud computing [15]. In the Edge computing paradigm, Edge resources such as micro data centers, routers and storage arrays

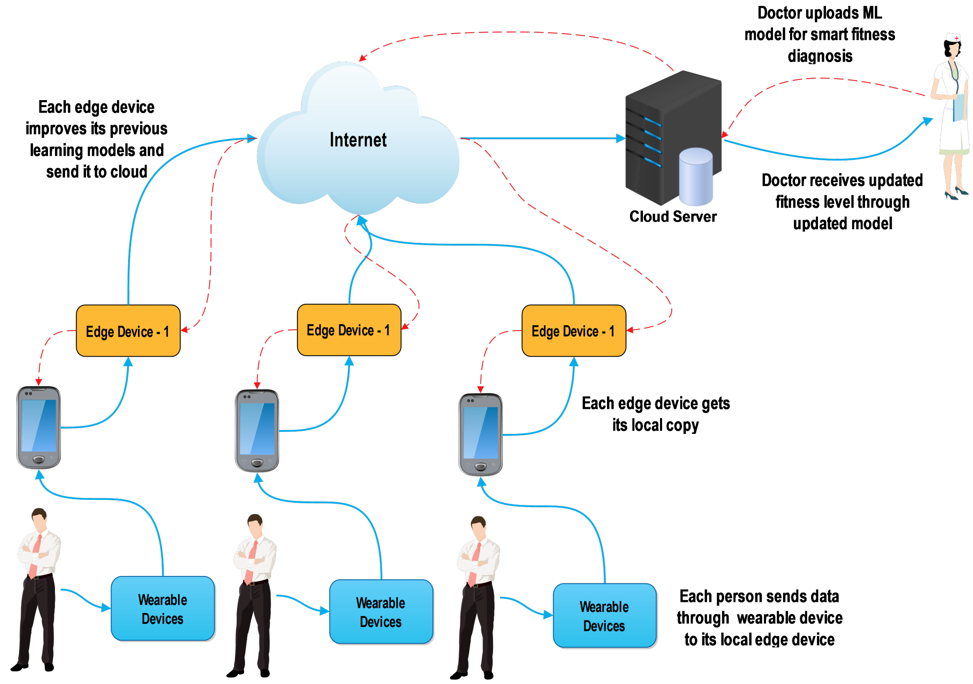

Fig. 1. A general overview of the proposed framework, as used with ubiquitous fitness trackers.

will be available at the Edge of the network. These resources will have significantly less capacity compared to their Cloud counterparts, but will be more closely located to the end users.

Due to this proximity, Edge computing offers significant opportunity to enable data analytics in a cost- and time-effective manner. In contrast to these benefits, existing ML algorithms that are trained and tested on a central data repository will no longer remain viable. Hence, new approaches are needed to adapt ML algorithms without requiring access to all data, to take advantage of the Edge. Recently, Distributed ML and Federated Learning [13] have been proposed that do not require a central data repository.

In this paper, we propose a conceptual framework for efficient healthcare data analytics based on user-generated data that leverages the Edge computing paradigm. Because healthcare data is highly sensitive in nature, data privacy is of great concern. Our approach will utilize privacy-preserving distributed ML techniques, and in particular, Federated Learning. We discuss the main ideas of our proposed framework and highlight some potential challenges that remain to be addressed.

The remainder of the paper is organized as follows. In section II, we review key recent works in the area of Federated Learning and Edge computing. The proposed framework is presented in Section III. In section IV, we highlight potential applications and research challenges of the proposed work before concluding the work in Section V.

II. BACKGROUND AND RELATED WORK

Li et al. [11] recently explored the applications of Federated Learning in healthcare, computer science, industrial engineering, and mobile devices. They provided an explanation of the concept of Federated Learning, created a categorization, and highlighted future research areas for optimization.

Oueida et al. [12] proposed an Edge computing-based framework for efficient resource management in emergency departments for supporting emerging smart healthcare systems. The proposed framework was validated using PRN (resource preservation net) Petri net via demonstration in two emergency departments (ED) in a general hospital for non-consumable medical resources. Both EDs had unique resources, but shared billing and radiology facilities. The efficiency and performance of the proposed framework were analyzed using average patient time, length of stay, and resource utilization. The lifecycle of patients in the emergency department (e.g., arrival to discharge) is also explained with the help of real-life scenarios. The results indicate that by integrating edge computing with PRN, the proposed framework shows improvements in patient waiting time, length of stay, and resource utilization.

Dong et al. [8] proposed an Edge-assisted decentralized healthcare monitoring system for the Internet of Medical Things (IoMTs). The proposed system was designed to ensure efficient resource management and allocation of wireless channel resources. The problems of allocating the wireless channels in intra-WBANs and reducing the system-wide cost were modeled using cooperative game model. Optimization (i.e., Pareto optimal point) was achieved using Nash bargaining and the performance of the proposed system was evaluated using the number of MUs and system-wide cost metrics. As with Li et al. [11], the authors also highlighted open research issues and challenges.

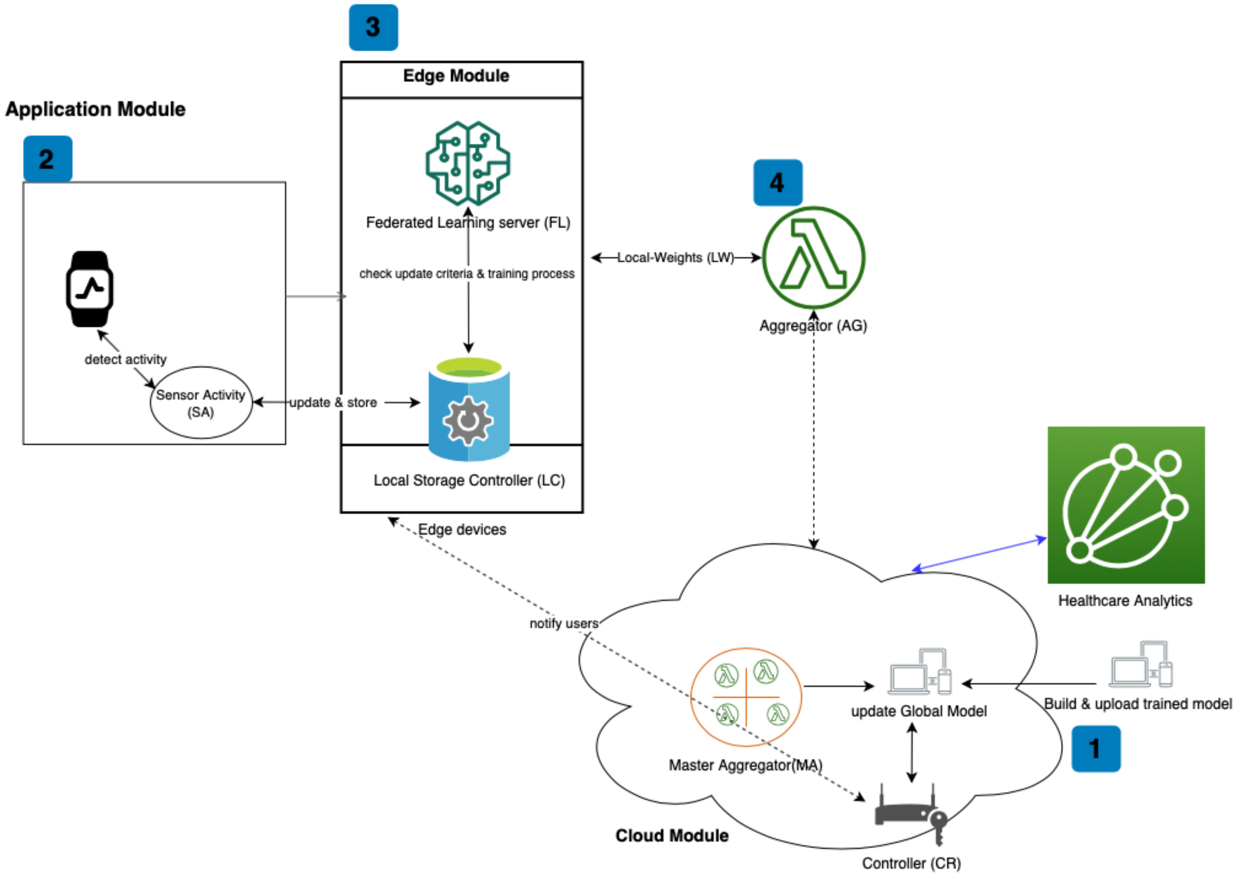

Fig. 2. Division of Sub-Modules in the Edge-Assisted Health Analytics framework.

III. Proposed Framework

In this work, we extend on these previous concepts by proposing a personalised Edge-based Federated Learning approach. Figure 1 presents a general overview of the proposed framework, whose architecture is further detailed in Figure 2. The proposed model could help healthcare practitioners by providing data-driven insights for the diagnosis or prediction of diseases by analysing mobility levels and behaviours using data generated from wearable-devices. Additional potential applications are discussed further in section IV. The proposed framework is divided into the following three modules, namely, the *Cloud Module*, the *Edge Module* and the *Application Module*, as follows:

- **Cloud Module**: This module will be managed by a model-owner, such as a designated healthcare IT manager, who will be responsible to coordinate various tasks on the Cloud such as registering patients, maintaining the database, and uploading trained models to the Cloud. The Cloud Module consists of two main components - the *Controller* (CR) and the *Master Aggregator* (MA). The CR will notify users/model owners about the availability of potentially updated global models. Similarly, the MA will aggregate the local-weights (LW) from the Edge devices using available algorithms (such as Federated Averaging). It will then update the global model by issuing the notification to the model-owner via the CR and transfer it again for inference to the local Edge nodes. Initially, the model-owner will train the global model on available public datasets. This global model could include data of any nature, such as, disease prediction based on fitness level or diagnosis, based on the application module. Once the model is pre-trained, the model owner will upload the global model to the Cloud via an application programming interface (API). Once the model is uploaded, the CR (whose main role will be to update both model owner and patients about the models) will notify users to upload the model.
- **Edge Module**: The Edge Module will improve the overall learning process of the global model, as local copies are re-trained on their respective Edge device. The Edge Module consists of three components; 1) the *Federated Learning server* (FL), 2) the *Local-storage Controller* (LC) and 3) an *Aggregator* (AG). When the users download the Global Model (GM), the FL server will start the learning process by first checking its eligibility criteria, which is based on the status of the Edge device, such as battery strength, processing capabilities, or available local samples. If the eligibility criteria are met, it will start incorporating the new data samples from the LC. The

LC controller will be responsible for storing data samples from end-user devices (such as smart-fitness watches) during any eligible activity as specified in the application module. Periodically, the updated encrypted local-weights from the local devices will be aggregated locally by the AG. Some threshold condition (such as, time interval) can be set here to trigger sending the values generated by the AG to the MA. For secure local model and global update aggregation, homomorphic encryption [6] could be used, as it allows calculations to be computed over the encrypted data without the need for decryption.

- **Application Module**: This module supports the incorporation of any smart-input device from which health-related data can be generated. In the example scenario provided, the input application module is a wearable device like a smart-fitness tracker watch, from which data are collected and stored in the LC of the Edge module for processing by its respective Edge device using Federated Learning algorithms. An important aspect of the application module is the Sensor Activity (SA) notification system. The SA issues updates to the LC in the Edge module when some physical or meaningful sensor activity is detected. As soon as the SA detects such activity, the LC will start collecting data. Again, basic thresholds or more intelligent decision logic could be implemented to initiate action. For example, data storage limits in the LC could be specified, triggering the FL to start the learning process before space runs out.

IV. Applications and Challenges of the Proposed Framework

A. Applications

Although primarily focused on healthcare, the proposed framework is general and could be employed and integrated for various applications such as smart-city, smart-grid, industrial IoT (IIoT), or the Internet of Medical Things (IoMT). A few health-related example applications may include:

- **Disease Management/Prevention**: The proposed framework could be extended to develop a distributed disease management system based on personal health data. Using a range of different evaluation parameters such as fitness level, sleep patterns, number of active physical hours, heart-beats etc, insights could be extracted about the health or condition of a person, or a group of individuals (such as by geographic location). Based on regularly updated learning models, appropriate measures could be taken to inform the user(s) about possible precautionary measures or prevention strategies. For such personalised health-diagnostics, especially those that leverage behavioural information, identifying at risk persons while preserving privacy remains an interesting and active research area.
- **Addiction/Mental Health Tracking**: The proposed framework could also be used to monitor addiction or mental health related behaviours. Gamified apps and chatbots could be used to motivate users to stay engaged, patient and mindful, or intervene when destructive patterns of behaviours are noted. Although such interventions have gained traction in recent years, informing them based on user-generated data and behaviours remains an unmet goal.
- **Real-time Health Monitoring**: With the emergence of technologies such as 5G, it is could be possible to extend this framework to enable real-time diagnostics using Edge computing. Real-time or near real-time constraints may be challenged by the reliability of connectivity, the heterogeneity of Edge devices, and preservation of privacy, and the regulation of decisions and liability remain an ongoing issue.

B. Research Challenges

As an emerging area of focus, the proposed framework could address a number of research challenges:

- Cyberattacks: There has been a surge of cyberattacks during this recent COVID-19 pandemic [9]. Numerous attacks, including inferential attacks, model poisoning attacks [14] and privacy concerns may make this framework vulnerable. Additionally, the Edge devices themselves may be vulnerable to various denial-of-service (DoS) or even physical attacks [7]. Hence, securing this model against such attacks is an essential research area.
- Data bias: There is a strong potential for bias to be introduced, as less-constrained or regulated data are added via the local sources. Although, different approaches exist to mitigate bias in ML models, more research is needed. Similarly, the implications of incorrectly labeled or corrupted data, as well as the provenance of the data, should be explored.
- Constrained nature of Edge devices: As most Edge devices are resource-constrained, it is possible that these devices may become exhausted quickly in terms of storage and computational resources. Consequently, novel methods of managing the usage and optimization of local storage and computation within this context could be explored.
- Acceptance of technology: The key benefit and differentiator of this approach is the leveraging of user-generated data. For the model to work properly, it will require regular updates for training purposes. Consequently, without some minimum level of user compliance, the performance of the model will drastically decrease. Therefore, it is important for users to remain engaged and actively accept the technologies. How to motivate users to accept emerging technologies for their own health benefit remains a real challenge.

V. Conclusion

In this paper, we propose a conceptual framework for leveraging Edge computing to support healthcare analytics based on user-generated data. The confluence of Cloud, Edge computing, IoT, Wearables, and Federated Learning will soon enable end users to play a more integrated role in their own health management. The empowerment and accountability of patients in the monitoring and prevention of diseases remains a critical goal in the sustainability of modern healthcare systems. To this end, the proposed model could facilitate the effective and scalable integration of user-generated wellness and behavioural data.

REFERENCES

[1] Unlocking the potential of the internet of things. "https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/the-internet-of-things-the-value-of-digitizing-the-physical-world", 2015.

[2] Connected care platforms help high-risk patients avoid the ER. "https://mhealthintelligence.com/news/connected-care-platforms-help-high-risk-patients-avoid-the-er", 2016.

[3] Global health and aging. "https://www.who.int/ageing/publications/global_health.pdf", 2016.

[4] Tracking universal health coverage: 2017 global monitoring report. "https://www.who.int/healthinfo/universal_health_coverage/report/2017/en/", 2017.

[5] S. N. Bhupathiraju and F. B. Hu. Epidemiology of obesity and diabetes and their cardiovascular complications. *Circulation Research*, 118(11), 1026.

[6] K. Bonawitz, V. Ivanov, B. Kreuter, A. Marcedone, H. B. McMahan, S. Patel, D. Ramage, A. Segal, and K. Seth. Practical secure aggregation for privacy-preserving machine learning. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, pages 1175–1191, 2017.

[7] M. Caprolu, R. Di Pietro, F. Lombardi, and S. Raponi. Edge computing perspectives: architectures, technologies, and open security issues. In *2019 IEEE International Conference on Edge Computing (EDGE)*, pages 116–123. IEEE, 2019.

[8] P. Dong, Z. Ning, M. S. Obaidat, X. Jiang, Y. Guo, X. Hu, B. Hu, and B. Sadoun. Edge computing based healthcare systems: Enabling decentralized health monitoring in internet of medical things. *IEEE Network*, 2020.

[9] S. Hakak, W. Z. Khan, M. Imran, K.-K. R. Choo, and M. Shoaib. Have you been a victim of covid-19-related cyber incidents? survey, taxonomy, and mitigation strategies. *IEEE Access*, 8:124134–124144, 2020.

[10] W. Z. Khan, E. Ahmed, S. Hakak, I. Yaqoob, and A. Ahmed. Edge computing: A survey. *Future Generation Computer Systems*, 97:219–235, 2019.

[11] L. Li, Y. Fan, M. Tse, and K.-Y. Lin. A review of applications in federated learning. *Computers & Industrial Engineering*, page 106854, 2020.

[12] S. Oueida, Y. Kotb, M. Aloqaily, Y. Jararweh, and T. Baker. An edge computing based smart healthcare framework for resource management. *Sensors*, 18(12):4307, 2018.

[13] N. Rieke, J. Hancox, W. Li, F. Milletari, H. Roth, S. Albarqouni, S. Bakas, M. N. Galtier, B. Landman, K. Maier-Hein, et al. The future of digital health with federated learning. *arXiv preprint arXiv:2003.08119*, 2020.

[14] R. Tomsett, K. Chan, and S. Chakraborty. Model poisoning attacks against distributed machine learning systems. In *Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications*, volume 11006, page 110061D. International Society for Optics and Photonics, 2019.

[15] B. Varghese, P. Leitner, S. Ray, K. Chard, A. Barker, Y. Elkhatib, H. Herry, C.-H. Hong, J. Singer, F. P. Tso, et al. Cloud futurology. *Computer*, 52(9):68–77, 2019.

[16] B. Varghese, S. Ray, B. N. Gohil, and S. Vega. Research challenges in query processing and data analytics on the edge. CASCON '19, page 317–322, 2019.