EPiC Series in Computing

Volume 98, 2024, Pages 61–68

Proceedings of 39th International Conference on Computers and Their Applications

Artificial Intelligence Usage and Data Privacy Discoveries within mHealth

Jennifer Schulte, Patrick Engebretson*, and Mark Spanier†

Dakota State University, Madison, SD

jennifer.schulte@dsu.edu, pat.engebretson@dsu.edu, mark.spanier@dsu.edu

Abstract

Advancements in artificial intelligence continue to impact nearly every aspect of human life by providing integration options that aim to supplement or improve current processes. One industry that continues to benefit from artificial intelligence integration is healthcare. For years now, elements of artificial intelligence have been used to assist in clinical decision making, helping to identify potential health risks at earlier stages, and supplementing precision medicine. An area of healthcare that specifically looks at wearable devices, sensors, phone applications, and other such devices is mobile health (mHealth). These devices are used to aid in health data collection and delivery. This paper aims at addressing the current uses and challenges of artificial intelligence within the mHealth field as well as an overview of current methods to help provide patient data privacy during data collection and storage.

1 Introduction

Advancements in artificial intelligence continue to make an impact in day-to-day life. This includes providing integration solutions whose purpose is to supplement or improve various steps or processes. One industry benefiting from this is healthcare. Artificial intelligence and machine learning are being used to help clinicians with decision making, radiologists with diagnosis, and pharmaceutical companies design drugs. Within the healthcare industry is mobile health, or mHealth. This area of healthcare is often referred to as using mobile technologies such as text reminders, home monitoring systems, and wearable devices in medical care [1]. With many of these solutions, especially wearable devices, data is typically collected from biomedical sensors [2]. This data is then stored and used to analyze and train various artificial intelligence models. One of the added requirements that comes with storing and saving sensitive health data is protecting patient privacy.

* Contributor 1

† Contributor 2

The remainder of this paper will look at some of the advancements and challenges that mHealth is facing with the use of artificial intelligence. Additionally, various methods and models used to protect patient privacy, such as k- anonymity and federated learning, will be addressed.

2 Artificial Intelligence in mHealth

Combining artificial intelligence with mHealth practices has led to a fairly new research domain called AI-powered mHealth, or AIM. [3] Many areas within artificial intelligence are being used within mHealth due to the sheer amount of data that is now being generated. Part of this growth comes from the increase in mHealth devices, sensors, and research as well as a shift to implementing electronic health records. By using machine learning, as well as other artificial intelligence methods, this data can be processed into personalized output for a patient, the healthcare team, and others when deemed necessary.

2.1 Advancements and Benefits

Growth in mHealth and AIM research can be attributed to a few different things. First, mHealth encompasses the Internet of Medical Things (IoMT) which contains wearable and implantable devices [4]. This area, especially smart watches, has seen a substantial rise in the number of users in the past decade, including use within the healthcare field. For example, there has been an increase in intensive care unit capacity monitoring and constant patient monitoring within healthcare facilities as well as at home. Another significant factor working towards fueling the growth of AIM research is the Defense Advanced Research Projects Agency (DARPA) XAI program. The goal of this program is to support the development of AI systems and its understanding by end users. By supporting and helping grow the number of users who have knowledge of what AI is capable of and what various models do, the growth of AIM research and mHealth as an industry will only continue to increase. [3] Some interesting examples within AIM research geared towards specific chronic conditions and diseases include the following:

1) Asthma

a) An application (XGBoost) collects patient chest movements and utilizes supervised learning on the data to identify breathing patterns [1].

b) An application utilizing regression models on audio recordings are able to create inhaler flow profiles with 91% accuracy. This research is being conducted in hopes to improve patient inhaler use [1].

2) Parkinson's Disease

a) A supervised learning application uses accelerometers and gyroscopes to collect data for a recurrent neural network (RNN) to detect early signs of Parkinson's disease [5].

Other conditions such as diabetes, sleep apnea, and Alzheimer's Disease are also benefiting from new and continuing AIM research. Some additional benefits include automatic detection of chronic disease occurrence, helping facilitate emergency response, real-time interventions for individuals struggling with mental health, and even preventing medical errors from occurring [3].

2.2 Artificial Intelligence Models

There are varying types of artificial intelligence at play within the mHealth area. The use of machine learning on all types of data has already helped provide faster, life-saving diagnosis, precision medicine findings, and much more. At a high level, one can break down machine learning techniques by the type of data that is used. Regression is used with numerical labels, clustering is ideal for unlabeled data, and classification is used with a categorical output value. For mHealth wearable devices specifically, the most common models used include some sort of classification. [4] Below are some examples of use cases within the different learning methods most often seen within AIM.

a) Supervised Learning: One common trend for supervised learning within mHealth is to build classification models. An example from [5] highlights a use case where a deep learning model was used to assist in identifying anomalies in breathing in hopes to better diagnose both respiratory and pulmonary diseases. Deep learning models have also been used for similar studies.

b) Semi-Supervised Learning: An example utilizing semi-supervised learning consists of a study that used a long short-term memory network. This study collected data from an mHealth wearable device (heart rate sensor) application and looked to identify varying medical conditions such as high cholesterol and diabetes. Specifically related to the data, this seems to be a good choice of model due to the presence of unlabeled data versus labeled data. [5]

c) Reinforcement Learning: Perhaps the most interesting use of artificial intelligence comes from the use of reinforcement learning to help customize and optimize Parkinson's Disease medication regimens for individual users. In this specific example there was a study conducted with 26 patients, all who suffer from Parkinson's Disease, who wore a movement tracker on their wrist for a designated amount of time. After this time period was complete, the data was evaluated, and the participating physician would then modify the patient's medication as needed based on the model output. [5]

2.3 Challenges

The fact that this research is being conducted within the healthcare field brings an additional level of complexity due to state and federal rules and regulations. This can be seen most evident in the fact that even though there has been an increase in AIM research, there hasn't been a lot of adoption for mHealth devices that implement artificial intelligence. One leading reason for this is that many of these devices are still being developed and researched. Additionally, while the data collection and sensors are sufficient, there is a need for additional verification for the AI integration. The challenge here, and quite possibly a primary reason for a lack of adoption, is a lack of transparency in ML models. This is something that XAI is working towards improving by breaking the knowledge barrier through building a basic understanding of AI. [3] Speaking specifically to the use of machine learning,

> the data generated by mHealth devices for home monitoring are increasingly reliable and validated against existing gold-standard equipment. However, the validity of the information created by machine learning analysis has not yet reached the standards required by health services. Many more large-scale studies, akin to clinical trials, will be required to test the outputs of real-time analysis using mHealth and machine learning algorithms deployed in the real world. [1]

An example of an mHealth application that has overcome these challenges and has gained mainstream acceptance is the irregular heart rhythm notification on the Apple Watch. This feature wasn't approved by the U.S. Food and Drug Administration until 2018. Even after approval, the feature still comes with many warnings and precautions. [4]

In terms of the data itself, there are concerns about the amount, availability, and privacy. The remainder of this section will address some of these concerns as mHealth research has the risk of introducing biases into their machine learning models when faced with these challenges. [2]

Since AIM research is looking at validating sensors, devices, or an artificial intelligence model, the sample sizes for data collection are quite limited and very specific. Part of this decision comes down to the current availability of sensors and devices as well as the willingness of individuals who not only meet the requirements to be a part of a study but are also willing to participate in it. With the ability to share patient data safely and securely, the challenge of the small sample size could be alleviated. For the most part, healthcare providers do not share data outside of their network for research studies. This includes proprietary mHealth data and devices. Unfortunately, this leads to siloed data which becomes an obstacle for machine learning processes. To mitigate this problem and help ensure data reliability [4] suggests a wide range of clinical experiments, transparent results, and using cross validation techniques.

The core of these challenges comes down to data privacy and the question of how this highly sensitive data can be shared in such a way that individuals cannot be identified. This leads us to the next section concerning securing patient data.

3 Patient Data Privacy

Patient healthcare data contains sensitive information that is protected by federal and state regulations. Whether storing or sharing of the data occurs, the data needs to be altered in some way to ensure patient privacy. In this section two commonly used methods of protecting patient data will be discussed; federated learning and k-anonymity.

Within patient privacy there are three main attributes that are of interest. These are identifiers, quasi-identifiers, and sensitive attributes. In the mHealth sector an identifier would be an attribute that uniquely identifies a patient. For example, an identifier might be the patient's medical record number. A quasi-identifier is a collection of attributes that together would be capable of identifying a patient. Date of birth and address together would be quasi-identifiers. Lastly, sensitive attributes are often attributes that are ideally kept private such as a medical condition. [6]

3.1 K-anonymity

K-anonymity is a method of deploying data masking to ensure that data is anonymized when quasi-identifiers appear at least k times. By doing this, patients have the probability of being identified 1/k at most [7]. Another way of stating this is that k-anonymity requires that at least k individuals share the same attributes [8]. K should be thought of as a numerical value that denotes the level of anonymity. If k=7 with identifiable variables being age and gender, then the k- anonymized data set needs to have at least 7 records for each value combination of age and gender. Higher values of k can signify a low level of re-identification probability, but it can also show that too much data may have been lost in the anonymization process. [9]

To achieve anonymity, suppression and generalization techniques are often applied. Suppression is simply a deletion of values from a table which can occur at a record level, attribute level, or cell level within the dataset. Generalization on the other hand transforms data into more general forms. For example, a patient's address might be replaced with just a zip code, city, or country. An important distinction to make between the two is that suppression can lead to a substantial amount of data loss since information is just removed. [7]

Additionally, there are varying models and algorithms in which k-anonymity can be achieved. Some of the algorithms found during this research include the incognito mode algorithm, Samarati algorithm, and Sweeney algorithm. [10] For some comparison amongst this small selection of algorithms,

- The Samarati algorithm evaluates all nodes at the generalized level while the Sweeny algorithm evaluates many nodes at different levels.
- Samarti's algorithm utilizes minimal generalization and suppression with techniques such as binary searches.
- While Samarti's algorithm is relatively fast, Sweeney's algorithm is noted to be even faster. However, the output produced by the Sweeny algorithm is often not used in research

purposes due to the generalization and scarcity of data that is left after k-anonymization is reached.

Lastly, it is important to note that employing only k- anonymization techniques will leave your data sensitive to various attacks. One specific attack type is sensitive attribute disclosure. To help visualize this process, Table 1 and Table 2 are provided. The data in Table 1 is in its raw data format whereas Table 2 contains the k-anonymized data. Thinking like an adversary, if you know that Emily is a female born in the 1990s, she can easily be identified as an individual suffering from COPD. Due to vulnerabilities such as this, there are other methods and algorithms that can be used instead of k-anonymity such as l-diversity and t-closeness. [6] These are not discussed in detail as it is outside the scope of this paper.

Patient ID	Name	Date of Birth	Zip Code	Gender	Health Condition
A0002	Allen	10/20/1994	53021	Male	Asthma
A1697	Baker	12/31/1991	53020	Male	COPD
B6720	Carly	7/23/1993	53020	Female	COPD
B9104	Darrell	3/19/1994	53020	Male	Asthma
C0052	Emily	11/27/1991	53020	Female	COPD
C4527	Fred	4/15/1994	53019	Male	Asthma

Table 1: Raw data

Patient ID and Name	Date of Birth	Zip Code	Gender	Health Condition
Suppressed	10/**/199*	5302*	Male	Asthma
	//1991	53020	Male	COPD
	7/**/199*	5302*	Female	COPD
	//199*	53020	Male	Asthma
	11/**/1991	5302*	Female	COPD
	4/**/199*	53019	Male	Asthma

Table 2: K-anonymized data

3.2 Federated Learning

Federated Learning is an approach of machine learning that assists in securely sharing patient data by sharing models rather than data. Over time this has become the widely preferred data privacy framework compared to a centralized machine learning scheme.

> [Federated learning]-based methods typically involve a central server that holds a global model, and during each round of [federated learning], a random subset of users is chosen to participate in the training of the global model. The selected users train the models locally using their data and then send their trained models back to the server. The server then

> aggregates the models and updates the global model, and the process repeats for subsequent [federated learning] rounds. [11]

There are a few different types of federated learning models. Breaking these down into two main categories there is cross-device or cross-silo and horizontal or vertical.

- Cross-device federated learning happens across multiple devices. These devices independently train the models before aggregation.
- Cross-silo federated learning occurs when multiple centers are used to train models.
- Horizontal federated learning is when training nodes share the feature with varying sample space.
- Vertical federated learning signifies when training nodes share the same sample space but vary in features.

When a combination of cross-device and cross-silo methods are used, the federated learning method is referred to as hierarchical federated learning. [12] Figure 1 provides sample layout diagrams for cross-silo federated learning, cross-device federated learning, and hierarchical federated learning. Unfortunately, at this time many of the proposed federated learning frameworks are not designed for mobile and wearable device data [11].

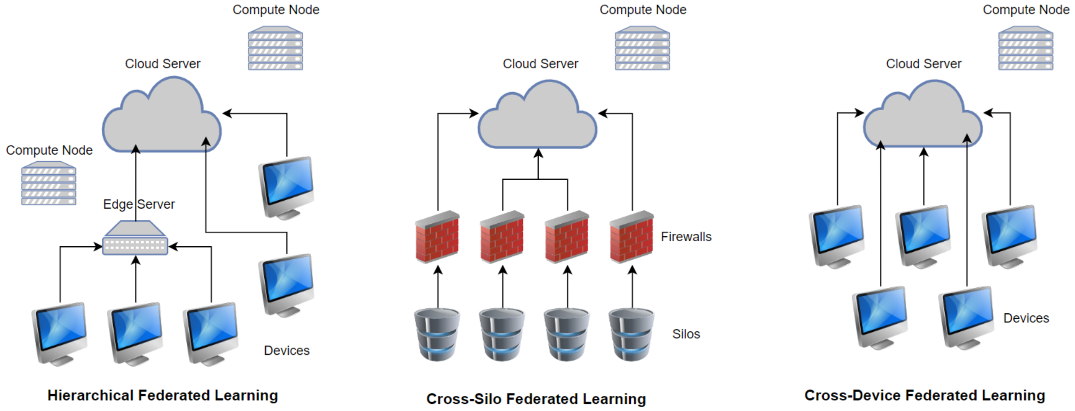

Figure 1: Federated Learning Diagrams

3.3 SHFL: A Suggested Framework

New frameworks to ensure patient data privacy continue to be designed and researched. One notable example is SHFL which is a k-anonymity-based secure hierarchical federated learning framework [11]. SHFL combines the benefits that both k-anonymity and federated learning have to offer into one solution and is specifically designed for smart healthcare systems. Unfortunately, federated learning struggles with scalability issues. This is due to the fact that the learning models are distributed to multiple devices and then collected again to train one local model. This suggested framework implements a hierarchical federated learning approach and a k-anonymity method that would hide both the location and identity of patients. The designers of SHFL claim that after detailed experiments and analysis of both performance and accuracy, SHFL outperforms state-of-the-art federated learning approaches. [11]

4 Conclusion

AIM research is still a relatively new field of research. As this field continues to grow there will be many more achievements, benefits, and challenges that come to the surface. Already AI has made such an impact within healthcare. Patients are experiencing more personalized care, physicians are benefiting from decision support systems, and new medications and treatments are continuously being brought to market.

The wearable devices and sensors, even at varying levels of development, are not the main cause of lack of mHealth AI integration. Instead challenges like the black box nature of artificial intelligence, lack of sufficient and fully representative datasets, and poor data sharing are all contributing to slow mainstream acceptance. Projects such as DARPA's XAI will continue to help lead the way in support of developing AIM research and fill the knowledge gaps that are required for artificial intelligence success.

Another crucial requirement that mHealth devices, and the healthcare field in general, need to continue developing and researching are data masking techniques, models, and frameworks. To share data for research or public purposes, individuals need to be anonymized. Specifically discussed in this paper were k-anonymization and federated learning. Each faces their own challenges such as scalability options and vulnerabilities to attack like sensitive attribute disclosure, however frameworks are being developed using multiple algorithms and methods to create more secure models. SHFL is an example of this by combining k-anonymity and hierarchical federated learning.

References

[1] K. C. Tsang, H. Pinnock, A. M. Wilson, and S. A. Shah, "Application of Machine Learning Algorithms for Asthma Management with mHea lth: A Clinica l Review," J. Asthma Allergy, vol. Volume 15, pp. 855–873, Jun. 2022, doi: 10.2147/JAA.S285742.

[2] T. Wang, Y. Du, Y. Gong, K.-K. R. Choo, and Y. Guo, "Applications of Federated Learning in Mobile Health: Scoping Review," J. Med. Internet Res., vol. 25, p. e43006, May 2023, doi: 10.2196/43006.

[3] P. Bhatt, J. Liu, Y. Gong, J. Wang, and Y. Guo, "Emerging Artificia l Intelligence–Empowered mHealth: Scoping Review," JMIR MHealth UHealth, vol. 10, no. 6, p. e35053, Jun. 2022, doi: 10.2196/35053.

[4] F. Sabry, T. Eltaras, W. Labda, K. Alzoubi, and Q. Malluhi, "Machine Learning for Healthcare Wearable Devices: The Big Picture," J. Healthc. Eng., vol. 2022, pp. 1–25, Apr. 2022, doi: 10.1155/2022/4653923.

[5] D. Nahavandi, R. Alizadehsani, A. Khosravi, and U. R. Acharya, "Applica tion of artificia l intelligence in wearable devices: Opportunities and challenges," Comput. Methods Programs Biomed., vol. 213, p. 106541, Jan. 2022, doi: 10.1016/j.cmpb.2021.106541.

[6] Graduate Student, Dept. of CSE, PESIT BSC, Bengaluru, 560100, India, S. Punagin, and A. Arya, "Privacy in the age of Pervasive Internet and Big Data Analytics – Challenges and Opportunities," Int. J. Mod. Educ. Comput. Sci., vol. 7, no. 7, pp. 36–47, Jul. 2015, doi: 10.5815/ijmecs.2015.07.05.

[7] A. Aristodimou, A. Antoniades, and C. S. Pattichis, "Privacy preserving data publishing of categorical data through k -anonymity and feature selection," Healthc. Technol. Lett., vol. 3, no. 1, pp. 16–21, Mar. 2016, doi: 10.1049/htl.2015.0050.

[8] I. E. Olatunji, J. Rauch, M. Katzensteiner, and M. Khosla, "A Review of Anonymization for Healthcare Data," Big Data, p. big.2021.0169, Mar. 2022, doi: 10.1089/big.2021.0169.

[9] K. El Emam and F. K. Dankar, "Protecting Privacy Using k- Anonymity," J. Am. Med. Inform. Assoc., vol. 15, no. 5, pp. 627–637, Sep. 2008, doi: 10.1197/jamia.M2716.

[10] A. K. Sangaiah, A. Javadpour, F. Ja'fari, P. Pinto, and H.-M. Chuang, "Privacy-Aware and AI Techniques for Healthcare Based on K- Anonymity Model in Internet of Things," IEEE Trans. Eng. Manag., pp. 1–15, 2023, doi: 10.1109/TEM.2023.3271591.

[11] M. Asad, M. Aslam, S. F. Jilani, S. Shaukat, and M. Tsukada, "SHFL: K-Anonymity-Based Secure Hierarchical Federated Learning Framework for Smart Healthcare Systems," Future Internet, vol. 14, no. 11, p. 338, Nov. 2022, doi: 10.3390/fi14110338.

[12] O. Rana et al., "Hierarchical and Decentralised Federated Learning," 2023, doi: 10.48550/ARXIV.2304.14982.