A Study on Federated Learning Systems in Healthcare

by

Arthur Smith M.D.

Submitted in Partial Fulfillment of the Requirements

for the Degree of

Master

of

Computing and Information Systems

YOUNGSTOWN STATE UNIVERSITY

August, 2021

A Study on Federated Learning Systems in Healthcare

Arthur Smith M.D.

I hereby release this thesis to the public. I understand that this thesis will be made available from the OhioLINK ETD Center and the Maag Library Circulation Desk for public access. I also authorize the University or other individuals to make copies of this thesis as needed for scholarly research.

Signature:

Dr.Arthur Smith, Student Date

Approvals:

Dr.Alina Lazar, Thesis Advisor Date

Dr.John Sullins, Committee Member Date

Dr.Yong Zhang, Committee Member Date

Dr. Salvatore A. Sanders, Dean of Graduate Studies Date

ABSTRACT

Advances in modern medicine occur because of research, and the most relevant research comes from clinical trials that include a large and diverse patient population. What of the possibility that this could be accomplished by collaborations between different hospital systems across the U.S.? Within each of these systems, there exists a broad and unique range of characteristics (e.g. community or academic, rural or urban, number of beds, specialist availability and services) each with a distinctive patient population. The patient health care record contained within each however, is governed by privacy restrictions as well as a data governance by that particular hospital system. There also exists an abundance of information from the wearable health device sphere: personal fitness trackers such as wrist monitors as well as the physician prescribed cardiac, glucose and oxygen sensors. What if we could collect all this independent and wide-ranging data while maintaining the patient's privacy? The envisioned possibility here in obtaining such a diverse, unbiased collection has the potential to bring enormous advances to modern day medicine, ultimately improving patient care, which as physicians, is our ultimate goal.

Acknowledgements

I would like to thank my thesis advisor Dr.Alina Lazar of the Department of Computer Science and Information Systems at Youngstown State University. Dr. Lazar has consistently guided me in the correct direction in regards to both my Masters education and this thesis. She has demonstrated patience and understanding as I ventured into my Computer Science studies. Her guidance has helped me more than I can express.

I would also like to thank the committee members Dr.Yong Zhang and Dr.John R Sullins for their time and advice during my thesis process.

Finally, I must express my gratitude to my wife Charisse for providing me with unfailing support and continuous encouragement throughout my many years of study, my Masters degree attainment, and through the process of researching and writing this thesis. Thank you.

Table of Contents

List of Figures **1**

1 Introduction **2**

2 How Federated Learning Systems Work **4**

2.1 Architecture 5

3 Privacy Concerns **7**

4 Security of Federated Learning **8**

5 Research Applications in Practice **17**

6 AI and the Future of Medicine **18**

7 Why Study Federated Learning **19**

8 Conclusion **21**

9 References **22**

List of Figures

1 Aggregation Model 6

2 Peer to Peer Model 6

1 Introduction

The Physician-Patient relationship is indisputably the cornerstone of medical care. In it lies the sanctity of the trust of confidentiality. We are bound by an ethical and a moral code to keep this relationship intact. The Hippocratic Oath states ''What I may see or hear in the course of the treatment or even outside of the treatment in regard to the life of men, which on no account one must spread abroad, I will keep to myself, holding such things shameful to be spoken about'' [1]. Patient confidentiality is also federal law. As health care providers we all know and understand some semblance of HIPAA laws. In 1996 *The Health Insurance Portability and Accountability Act* was established by Congress to prevent sensitive patient data from being disclosed without the patients or their legal guardians' approval.

Privacy conventions instill a sense of trust in the healthcare system that we as providers will maintain our patients dignity through confidentiality. Some circumstances such as billing, provider to provider communications, and research do indeed necessitate some sharing of the patient data.This is all done of course with patient consent. While billing does little if anything to improve a patients condition, improving provider communications and medical research will.

Consider medical research across broad domains where hospitals, pharmaceutical companies, and primary care offices can collaborate on clinical trials, patient screening and prevention, and specialized topics such as addictions, pre-natal care, end of life care, and special populations. AI/Machine Learning holds so much promise to decrease morbidity and mortality. However, in order to accomplish this, a diverse and non-biased patient population needs to be aggregated and adequately studied. Imagine connecting urban hospitals in Los Angeles to rural hospitals in Arkansas to

primary care offices in Florida to evaluate what etiologies are involved in diabetes, why some treatments work for some and not others, and how to develop best practices for care.

I believe that the answers to many of our medical problems are out there if we were able to perform a meticulous analysis of our patient populations. We must first instill in our institutions that the methods are safe to utilize, we also must assure our patients that their medical data remains private.

In this paper we will explore a simple, patient confidentiality protecting, yet very functional methodology known as *Federated Learning*, its topology, and two cutting edge privacy mechanisms that can help to secure it to an even greater extent. First let's discuss the background of the methodology. The concept of Federated Learning was first described by Professor Patrick Hill in 1985 as ''Federated Learning Communities''. Professor Hill was discouraged by the current state of higher learning with what he described as ''a fragmentation between departments and the student'' and suggested the development of independent ''Federated Learning'' structures where people with the same ideas come together and learn from each other [2].

Fast forward to 2017 when H. Brendan McMahan and his colleagues at Google also used the term ''Federated Learning'' in the Computer Science literature with their work *Communication Efficient Learning of Deep Networks from Decentralized Data*. Here they describe it as: ''The learning task is solved by a loose federation of participating devices (clients) which are coordinated by a central server'', where ''each client has a local training data set which is never uploaded to the server''. Instead, ''each client computes an update to the current global model maintained by the server'' and ''only this update is communicated ''. Their paper introduced the ''Federated

Averaging Algorithm‘’ where the server sends a model to a selected number of devices (e.g. n cell phones), the devices in turn update this model with their data,then send the updated model back to the server which then takes the *average* of all such updates before sending the model back to the next batch of devices for another round.This will continue until the training process is completed [3].

2 How Federated Learning Systems Work

Merriam-Websters dictionary defines the term Federate as *united in alliance* [4]. With that in mind let's envision Federated Learning as a system of ‘’data islands‘’ that each participate in a sort of united collaboration with each other through a central mechanism. The data islands are able to share their some of their information with the central mechanism, but for privacys sake, not with each other. For our purposes here we will refer to this ‘’collaboration‘’ as the training of a Machine Learning (ML) model, a central mechanism such as a server which updates, averages, and redistributes the model, and the ‘’data islands‘’ as hospitals, wearable Internet of Things devices, cellular phones, etc. Let us also concretely define Machine Learning as a subset of Artificial Intelligence where we teach a ‘’machine‘’ (e.g., Computer System) to ‘’learn‘’ by inputting known data (e.g Chest x rays with pneumonia) in order to correctly identify the next chest x-ray in which we have no diagnosis (Pneumonia-yes or no?). Learning will be either classification or prediction.

The training of our machine learning model however, is a data intensive process, requiring sufficient amounts of information to obtain accurate results. In fact, as with

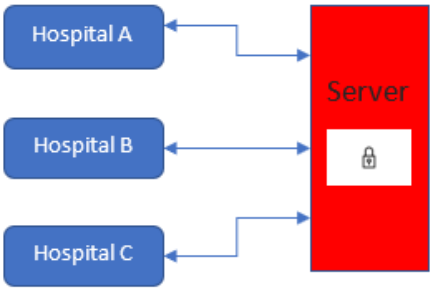
any study, the larger and more diversified the data set, the greater its validity. Therefore, in order to expose a model to a sizeable and diverse population, we would utilize a Federated Learning System. Here each system would be able update the model by using their own private data. The end result here being a more robust machine learning model better able to solve the complexities of disease processes, ultimately providing high quality information to physicians for patient care.
While there are several designs that can be used, let's look at the two basic structures below.

2.1 Architecture

There are two basic schemes in Federated Learning Systems:

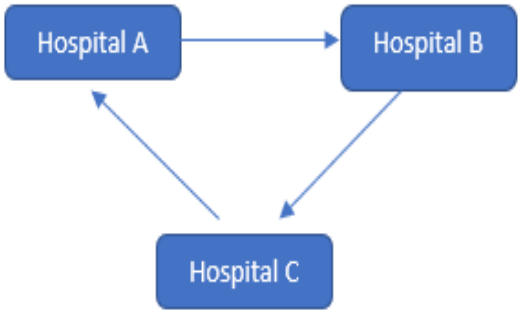
A.) *Aggregation Server*: This approach initially proposed in 2017 by McMahan et al is one where, as described previously, a machine learning model gets distributed to each of the ''data islands'' which we will hereby refer to as the Clients (Hospitals, IoT devices) from a cloud server, in which they update the model with their own independent data. The updated model then returns back to the central server where the data is aggregated from the clients, averaged, then redistributed. See Fig. 1.

Figure 1: Aggregation Model



..

Figure 2: Peer to Peer Model



.

B.) *Peer to Peer*: This architecture,in contrast to the Aggregation Server, is a mainly decentralized design. In other words, each of the clients are connected to one another rather than a common source such as the central server.Each client will update and aggregate the model themselves before a re-distribution to their neighbor. See Fig. 2 [5].

3 Privacy Concerns

As mentioned previously, patient privacy is paramount. Adversaries, with ransomware attacks aside, also attempt to gather data on patients in order to exploit for their own vested interests. Such a concern would be the leakage of a patients sensitive information to a 3rd party with financial interests (insurance companies, for- profit data mining companies). A major example of this occurred on May 14, 2013, at Howard University Hospital, where one of the hospitals medical technicians was charged by federal prosecutors for violating the Health Insurance Portability and Accountability Act. This individual, apparently over a 17 month period, collected patients names, addresses and Medicare numbers in order to sell their information [6]. Therefore, it is imperative that our Federated Learning Systems are resistant to any privacy leaks either in our data models, the network, or the even as above, from those entrusted with patient confidentiality.

An easily understandable privacy policy explaining the use of a patients anonymous data for research, can bring Federated Learning into a more mainstream acceptance and utilization through trust. However the data leakage from the electronic health records and what appears to be the ability for malicious actors to gain access to the hospital IT systems will have to be addressed with greater diligence.

In the following section we will discuss the concept of security in relation to Federated Learning Systems

4 Security of Federated Learning

Unfortunately, many growing threats exist today. What were once the spam and phishing of emails of the past, have now evolved into an alarming 1.5 trillion dollar annual profit worldwide using malware, especially ransomware, as well as many other types of attacks [7]. These attacks could range from infiltrating and changing the ML model known as ''poisoning'', to a data-mining of the medical records by for profit data mining entities, or just simply the accidental leakage of data to an honest source.

Based on the methodology of Federated Learning, a great degree of security is implicit, however as with any other type of data communications, it contains vulnerabilities. These vulnerabilities unfortunately exist in every system as there is no such perfect application, hardware or software available to date. Any system however, can and should have without question, the most recent and robust measures in place to help ensure the utmost in confidentiality and authorizations. Until suitable safety measures are in place and the general public while patients feel confident, they may be reluctant to agree to any informed consent that consist of new research methods. Therefore, leading security measures should be instituted to provide this protection. Two of the most cutting-edge security measures available to Federated Learning systems are: *Differential Privacy* and *Homomorphic Encryption*.

We will explore each in depth here to increase our understanding of their theory and mechanisms.

Differential Privacy: A privacy method developed by Cynthia Dwork and her colleagues at Microsoft in 2006, and described in their paper *Calibrating Noise to Sensitivity in Private Data Analysis*. The idea of ''Privacy From Pertubation'' where the true answer plus noise is added to the result to protect the participants in a statistical database from harm helped form the basis for this method [8]. With a Differential Privacy guarantee, a patients participation in a database is protected by the assurance that any analytical evaluation on that database will not reveal information specific to them [9]. It is a mathematical definition of privacy expressed as:

$$Pr[A(D1) \in S) \leq \epsilon * Pr[A(D2) \in S) + \delta$$

Where D1 and D2, are both identical databases except for differing by one row. (ϵ) is the Privacy Loss Parameter, and (δ) is what controls the probability a breach would occur, and since it is so small a number, it may be may be simply set to 0 [9]. In other words the probability of the two results with one database being 1 row less using differential privacy are such that they should be indistinguishable or as equally likely to occur [8]. This means that an adversary (insurance underwriter) who evaluates the output will not be able to tell whether or not any one particular individuals data was used, or even what it contained [10]. However, Dwork does explain that ''differential privacy ensures that only a limited amount of risk is incurred by joining a database'' [11]. Therefore,as stated earlier, and we digress some, there is no 100 percent privacy guarantee. This Privacy Loss Parameter (ϵ), is also called the ''Privacy Budget'', and is a real number between 0 and 1. The lower the value of the parameter, the more indistinguishable the results, and therefore the more each

individuals data is protected. However, a lower value also equals lower accuracy. A higher budget value on the other hand will offer less privacy, but it provides better accuracy. There is therefore a trade off in choosing the number [9]. Finally a LaPlace mechanism adds the necessary noise to the sensitive data.

It is beyond the scope of this paper to continue to further elucidate Differential Privacy. A simple way of looking at Differential Privacy is one where 2 databases are different by any one arbitrary row (Patient Data), and this difference (Differential) would not change the outcome of any statistical computation on such database. The ''Privacy Budget'' mentioned earlier is a selected distance between the two databases, and would be adjusted accordingly to balance accuracy and privacy.

To demonstrate, a small practical example was created using the *PyDP* package from **GitHub** which offers a Python API into **Google** *(DP Library)* [12].

We simulated a hypothetical database that each of the 50 states contributed to as mandated by the federal government to evaluate the number of seat belt related automobile crash deaths reported in one week. Fearing financial reductions in highway funds if the number of new cases is above a certain number (e.g. 20), the states collectively used differential privacy to submit their data. The privacy budget is a number between 0 and 1 determined by the DP Library. Below is a list of the States and their respective crashes in which there was a fatality due to the non-use of a seat belt.

State	Case Fatalities
Alabama	23
Alaska	31
Arizona	24
Arkansas	18
California	3
Colorado	6
Connecticut	38
Delaware	2
Florida	48
Georgia	20
Hawaii	8
Illinois	25
Indiana	43
Iowa	42
Kansas	30
Kentucky	28
Louisiana	35
Maine	21
Maryland	15
Massachusetts	44
Michigan	34
Minnesota	19

State	Case Fatalities
Mississippi	26
Missouri	12
Montana	33
Nebraska	22
Nevada	45
New Hampshire	5
New Jersey	27
New Mexico	50
New York	13
North Carolina	3
North Dakota	49
Ohio	17
Oklahoma	14
Oregon	47
Pennsylvania	16
Rhode Island	39
South Carolina	7
South Dakota	29

State	Case Fatalities
Tennessee	9
Texas	6
Utah	4
Vermont	46
Virginia	10
Washington	41
West Virginia	32
Wisconsin	37
Wyoming	1

The Public vs. the Private Counts Above the Federal Limit:

A) Public Count Above:

```
def count above(limit: int) -> int:
return df[df.New Cases > limit].count()[0]
```

B) Private Count Above:

```
def private count above(privacy  budget: float, limit: int) -> int:
x = Count(privacy budget, dtype="int")
return x.quick result(list(df[df.New Cases > limit]['New Cases']))
```

C) Results:

```
Public count above 20: 28
Private count above 20: 27
```

The difference in value corresponds to the privacy preserved [12].

The Public vs. the Private Mean

A) Public Mean:

```
def public mean()-> float:
return statistics.mean(list(df['New Cases']))
```

B) Private Mean:

```
def private mean(privacy budget:float)-> float:
x= Bounded Mean (privacy budget,0,1,100)
return x.quick result(list(df['New Cases']))
```

C) Results:

```
Public Mean: 24.16
Private Mean: 23.14
```

The difference in value corresponds to the privacy preserved [12].

The difference in values above reflect the ''Differential Privacy'' between the two databases. Furthermore, securing patient information in a medical database with Differential Privacy would add the guarantee that a patients inclusion in such a database should bring no harm to the patient. Further research continues into this robust privacy mechanism and its application to Federated Learning.

Homomorphic Encryption: 1st described by Rivest, Adelman, and Dertouzos from MIT in 1978 in their paper *On Data Banks and Privacy Homomorphisms*, this lattice based cryptography method allows for computations to be performed on encrypted data, without the need for an initial decryption [13]. For this reason, it is considered by many to be ''The Holy Grail'' of cryptography. The idea is to delegate the complexities of computation to another party without that party performing any type of decryption. After the computation has been completed, the results are sent back to the owners of the data, who then use their secret key to unlock the results. When it was first proposed, it was in theory only. Since it's initial proposal, researchers have developed several schemes to make it more practical. Let's first examine the fundamentals. The Oxford Concise Dictionary of Mathematics defines *Homomorphism* as *a mapping between two similar algebraic structures which preserves the relational properties of elements in the two structures* [14].
For example:

$$f(xy) = f(x) * f(y)$$

A practical analogy of Homomorphic Encryption would be one of an impenetrable jewelry ''glove box'' where one can assemble the precious material inside via the use of gloves that extend into the box, without have any direct access to the material. After the jewelry is assembled, it is sent back to the jeweler (the owner of the "data") [15].
Rivest et. al. described a Fully Homomorphic Encryption (FHE) theory where any arbitrary data can be added and multiplied. However this was at first discovered to

be impractical, therefore researchers developed Somewhat Homomorphic Encryption (SHE) schemes that were able to either calculate the addition of, or multiplication of cipher texts, but not both. Further extensive research is being done, and several libraries are readily available such as **Microsoft** *(SEAL)* and **IBM** *(HElib)* to implement this privacy mechanism into Federated Learning system. An extensive list can be found at *The Awesome Homomorphic Encryption List* [16].

Updating Federated Learning models via the cloud using Homomorphic Encryption would be an extensively secure and robust mechanism for protecting patient privacy.

5 Research Applications in Practice

While the current literature regarding the use of Federated Learning in Healthcare systems is limited due to the fact it is still innovative, several recent studies have been quite successful.

A 2020 study by Lee and Shin developed a FL framework with the APIs in Django and AWS,using the benchmark data-sets MNIST, MIMIC-3, and ECG from Physionet, and after comparing their FL model to the more traditional centralized machine learning model, discovered a very competitive performance between the two with of course better privacy from the federated learning [17]. Another study from IBM in 2019 developed a FL framework using predictive classifier that evaluated data from 1 million patients to learn a global adverse drug reaction (ADR) model that focused on two parameters: 1) Chronic opiate use disorder that developed from acute prescribing, and 2) Extra-pyramidal symptoms that resulted from taking anti-psychotic medicines (i.e. persons who suffer from schizophrenia). The results also showed a very compara-

ble performance between the federated framework and the centralized learning model [18]. Researchers from Boston University in conjunction with Massachusetts General Hospital in 2019 developed a predictive algorithm known as *iterative cluster Primal Dual Splitting*, which evaluated the electronic health records in a Federated Learning methodology in order to predict hospitalization admissions during a one year period for patients with heart disease [19]. Finally another from 2019 from NVIDIA and Kings College of London used Federated Learning to evaluate brain tumors via deep neural networks. 285 subjects in all were used through the BraTS 2018 dataset of MRIs. They too demonstrated comparable performances between the distributive and the centralized models [20].

Although the exact methods of the above studies are beyond the scope of our paper here, it is important to understand that such research attests that Federated Learning can not only uphold our patients privacy, but also accurately and correctly evaluate the medical data in regards to research. Further research is ongoing.

6 AI and the Future of Medicine

In 2020 alone, 4 billion dollars were invested by venture capitalists in the AI/Medicine research space, and at present there are over 1000 companies currently working in this field [21].

Buzzwords such as ''Deep Learning'' and ''Neural Networks'' permeate much of the AI/ML literature today, and healthcare research has made it a point to be involved. Disciplines such as Radiology for example use AI for fracture, stroke, and lung cancer detection, Pathology uses AI digitized tissue slides to increase diagnostic accu-

racy [22], Ophthalmology employs AI to detect diabetic retinopathy [23], and Psychiatry has even used AI for depression screening utilizing natural language processing methods to discover depression as well as those at increased risk for suicide [24]. A remote cardiovascular monitoring company known as LIVMOR recently received FDA clearance for its HALO, a wearable Atrial Fibrillation monitoring device [25], while Apple as well has also received FDA approval for their algorithm to detect irregular cardiac rhythms from their Apple Watches in patients over the age of 22 [26]. Actually, as of 2020, there are 29 algorithms have received FDA approval recently including sleep disorder monitoring algorithms [27].

On a side note as much as machine learning has made its way into health care, we must first we must consider whether that particular implementation is indicated: Does it provide patient benefit and safety? What about the providers benefit (more time spent with patient,less documentation time, improved diagnostic accuracy, and increased job satisfaction ? What about regulations regarding what AI can and cannot be used for? Who is responsible for missed diagnosis from algorithms? We must be very thoughtful and demand transparency and sound validations from the AI studies. If hospitals become ''united in alliance'' with research,and the patient care remains at the forefront, we believe these questions can be answered confidently.

7 Why Study Federated Learning

With the guiding principle of privacy in mind, and as we move into the future of medicine, there remains a vast infrastructure of diverse medical information that holds the possibility to reduce morbidity and mortality if only suitably obtained and

studied. One of the largest electronic health record suppliers, Epic, has records on 225 million patients in the U.S., or about two thirds of the countrys population [28]. There are several other electronic health record providers operating in each and every hospital in the United States. According to The American Hospital Association, there are 6,090 hospitals with 36,241,815 annual admissions [29].

Data governance however, both on a private as well as public level, strictly limits the capability of obtaining such disparate quality data. Currently, considerable amounts of clinically relevant Artificial Intelligence (AI) research are being done by data scientists on curated and redacted data sets as the ones above using the benchmarks, that, unlike real-time clinical data, do not reflect the diversity, or the inherent biases present in our patient populations . How then can we uphold the oath to our patients confidentiality, both ethically and legally, while employing the vast potential AI offers to help to improve morbidity and mortality? We believe the paradigm of a Federated Learning System (FLS) may hold the key.

We are motivated by the potential to be able to study the extensive medical data set using machine learning, meanwhile respecting patient privacy and data governance. Data sharing between distinct hospital systems is successful as patient records over the last few years are becoming more readily available to other institutions via the electronic health record for continuity of care should a patient present between the different hospitals. With the availability of a chart containing a patients medications, surgeries and especially allergies, the out-of-system Emergency Medicine physician, who would not otherwise have had quick access to this record, is now able to more appropriately and safely, treat his or her patient(s).

8 Conclusion

The ability for biomedical researchers to obtain and explore the data contained in the exhaustive amount of records would expose patterns that could reflect on the etiology of a number of disorders, all the while maintaining patient privacy and data governance.

Artificial Intelligence and its machine learning methods, especially using a Federated Learning System, appears to hold much promise for research in medicine. Especially when having the ability to adopt robust security measures. Furthermore, its potential for extensive collaborations between multiple clients including hospital networks, biotech firms, pharmaceutical researchers, and as mentioned previously, the IoT devices, cannot be understated.The benefits of investment and cooperation between these entities with regards to quality research, and how it would equate to an improved healthcare system should continue to be further explored.

This successful integration between patient care and secure technology, we believe, would significantly impact our goals of easing the heavy burdens of morbidity and mortality.

9 References

[1] Greek Medicine - The Hippocratic Oath. Publisher: U.S. National Library of Medicine.

[2] Patrick Hill. Transcribed from a speech given at the Inaugural Conference on Learning Communities of the Washington Center for Improving the Quality of Under- graduate Education, Olympia, WA, October 22, 1985. page 27.

[3] H. Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas. Communication efficient learning of deep networks from decentralized data. 2017.

[4] Merriam-Webster's Dictionary. Federate, 2014. Accessed: 2021-07-23.

[5] Nicola Rieke, Jonny Hancox, Wenqi Li, Fausto Milletari, Holger R Roth, Shadi Albarqouni, Spyridon Bakas, Mathieu N Galtier, Bennett A Landman, Klaus Maier-Hein, et al. The future of digital health with federated learning. *NPJ digital medicine*, 3(1):1–7, 2020.

[6] FBI. Former howard university hospital employee pleads guilty to selling personal information about patients. https://archives.fbi.gov/archives/washingtondc/press-releases/2012/former-howard-university-hospital-employee-pleads-guilty, 2012. Accessed: 2018-07-23.

[7] Ivana Vojinovic. Ransomware statistics in 2020: From random barrages to targeted hits. https://dataprot.net/statistics/ransomware-statistics, 2020. Accessed: 2021-07-23.

[8] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In *Proceedings of the Third Conference on Theory of Cryptography*, TCC'06, page 265 to 284, Berlin, Heidelberg, 2006. Springer-Verlag.

[9] Alexandra Wood, Micah Altman, Aaron Bembenek, Mark Bun, Gaboardi Marco, James Honaker, Kobbi Nissim, David R O'Brien, Thomas Steinke, and Salil Vadhan. Differential privacy: A primer for a non-technical audience. *Vand. J. Ent. & Tech. L.*, 21:209, 2018.

[10] Sameer Wagh, Xi He, Ashwin Machanavajjhala, and Prateek Mittal. Dp-cryptography: Marrying differential privacy and cryptography in emerging applications. *arXiv preprint arXiv:2004.08887*, 2020.

[11] Cynthia Dwork. Differential privacy: A survey of results. In *International conference on theory and applications of models of computation*, pages 1–19. Springer, 2008.

[12] PyDP. Introduction to pydp. https://pydp.readthedocs.io/en/latest/introduction.html#further-reading, 2021. Accessed: 2021-07-23.

[13] Ronald L Rivest, Len Adleman, Michael L Dertouzos, et al. On data banks and privacy homomorphisms. *Foundations of secure computation*, 4(11):169–180, 1978.

[14] Clapham,Nicholson. Homomorphism. Oxford Concise Dictionary of Mathematics by Oxford University Press, 1999.

[15] Craig Gentry. Computing arbitrary functions of encrypted data. *Communications of the ACM*, 53(3):97–105, 2010.

[16] Alexander Wood, Kayvan Najarian, and Delaram Kahrobaei. Homomorphic encryption for machine learning in medicine and bioinformatics. *ACM Computing Surveys (CSUR)*, 53(4):1–35, 2020.

[17] Geun Hyeong Lee and Soo-Yong Shin. Federated learning on clinical benchmark data: Performance assessment. *Journal of medical Internet research*, 22(10):e20891, 2020.

[18] Olivia Choudhury, Yoonyoung Park, Theodoros Salonidis, Aris Gkoulalas-Divanis, Issa Sylla, et al. Predicting adverse drug reactions on distributed health data using federated learning. In *AMIA Annual symposium proceedings*, volume 2019, page 313. American Medical Informatics Association, 2019.

[19] Theodora S Brisimi, Ruidi Chen, Theofanie Mela, Alex Olshevsky, Ioannis Ch Paschalidis, and Wei Shi. Federated learning of predictive models from federated electronic health records. *International journal of medical informatics*, 112:59–67, 2018.

[20] Wenqi Li, Fausto Milletarì, Daguang Xu, Nicola Rieke, Jonny Hancox, Wentao Zhu, Maximilian Baust, Yan Cheng, Sébastien Ourselin, M Jorge Cardoso, et al. Privacy-preserving federated brain tumour segmentation. In *International workshop on machine learning in medical imaging*, pages 133–141. Springer, 2019.

[21] Nupur Garg MD. Bullish on ai. https://www.acepnow.com/article/if-physicians-embrace, Jan 2021. Accessed 2021-07-23.

[22] Miao Cui and David Y Zhang. Artificial intelligence and computational pathology. *Laboratory Investigation*, 101(4):412–422, 2021.

[23] Michael D Abràmoff, Philip T Lavin, Michele Birch, Nilay Shah, and James C Folk. Pivotal trial of an autonomous ai-based diagnostic system for detection of diabetic retinopathy in primary care offices. *NPJ digital medicine*, 1(1):1–8, 2018.

[24] Sarah Graham, Colin Depp, Ellen E Lee, Camille Nebeker, Xin Tu, Ho-Cheol Kim, and Dilip V Jeste. Artificial intelligence for mental health and mental illnesses: an overview. *Current psychiatry reports*, 21(11):1–18, 2019.

[25] Ken Persen. K201208. https://www.accessdata.fda.gov/cdrh_docs/pdf20/K201208, September 2020. Accessed 2021-07-23.

[26] FDA. Den180044. https://www.accessdata.fda.gov/cdrh_docs/pdf18/den180044, September 2018. Accessed 2021-07-23.

[27] Stan Benjamens, Pranavsingh Dhunnoo, and Bertalan Meskó. The state of artificial intelligence-based fda-approved medical devices and algorithms: an online database. *NPJ digital medicine*, 3(1):1–8, 2020.

[28] Katie Jennings. The billionaire who controls your medical records. https://www.forbes.com/sites/katiejennings/2021/04/08/billionaire, April 2021. Accessed 2021-07-23.

[29] American Hospital Association. American hospital association fast facts on u.s. hospitals. https://www.aha.org/statistics/fast-facts-us-hospitals, 2021. Accessed 2021-07-23.