Computers in Biology and Medicine 158 (2023) 106848

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

Privacy-preserving artificial intelligence in healthcare: Techniques and applications

Nazish Khalid [a], Adnan Qayyum [a], Muhammad Bilal [b], Ala Al-Fuqaha [c], Junaid Qadir [d,*]

[a] Information Technology University, Lahore, Pakistan
[b] Big Data Enterprise and Artificial Intelligence Lab (Big-DEAL), University of the West England, Bristol, United Kingdom
[c] Hamad bin Khalifa University, Doha, Qatar
[d] Qatar University, Doha, Qatar

ARTICLE INFO

Keywords:
Privacy
Privacy preservation
Electronic health record (EHR)
Artificial intelligence (AI)

ABSTRACT

There has been an increasing interest in translating artificial intelligence (AI) research into clinically-validated applications to improve the performance, capacity, and efficacy of healthcare services. Despite substantial research worldwide, very few AI-based applications have successfully made it to clinics. Key barriers to the widespread adoption of clinically validated AI applications include non-standardized medical records, limited availability of curated datasets, and stringent legal/ethical requirements to preserve patients' privacy. Therefore, there is a pressing need to improvise new data-sharing methods in the age of AI that preserve patient privacy while developing AI-based healthcare applications. In the literature, significant attention has been devoted to developing privacy-preserving techniques and overcoming the issues hampering AI adoption in an actual clinical environment. To this end, this study summarizes the state-of-the-art approaches for preserving privacy in AI-based healthcare applications. Prominent privacy-preserving techniques such as Federated Learning and Hybrid Techniques are elaborated along with potential privacy attacks, security challenges, and future directions.

1. Introduction

The term artificial intelligence (AI), first coined by John McCarthy in 1956, refers to the capability of computers to perform tasks similar to those performed by humans. In other words, AI simulates human intellect through computer programs mimicking human actions artificially. However, AI requires lots of data and computing to realize its full potential. While computing power has undoubtedly aided in the revival of AI, data has enabled AI to achieve all of its current accomplishments. In recent years, AI-empowered software solutions have seen widespread adoption. Notably, AI is becoming a de facto standard for processing large amounts of data to support complex decisions, which is not just difficult but rather impossible for humans in certain fields. The amount of data created today significantly outpaces humans' capacity to consume, comprehend, and use it to inform non-trivial decisions in a timely manner. Henceforth, AI has many applications across different fields. It is hard to find one industry that will not benefit from this great innovation of our times [1].

In healthcare, AI is unlocking new possibilities by advancing medicine in entirely unimaginable ways and solving some of the grand global healthcare challenges. For example, AlphaFold, a recent AI-powered protein structure prediction algorithm solved the protein folding problem that hampered crucial advancements in biology and medicine for the past 50 years [2]. Likewise, innovations like In Silico Trailing allows pharmaceutical companies to simulate clinical trials for drug discovery on wider population models with greater control and fewer resource constraints to create great drug products.[1] There are enormous applications of AI in healthcare. Fig. 1 presents major healthcare specialities where researchers have been attempting to apply AI-based digital solutions. On the downside, there are ethical concerns about the potential misuse of these innovations. It took six hours for drug discovery AI to identify 40,000 potentially lethal molecules and most potent nerve agents [3]. Regardless, the transformations AI could bring to healthcare are unanimously agreed upon, ranging from advancing rapid diagnosis, personalizing care, and reducing unnecessary outpatient appointments that could save billions to the economy [4].

AI algorithms, specifically created through machine learning (ML), require large amounts of high-quality data to learn to perform pattern-matching tasks at human-level performance. The fact that data drives

* Corresponding author.
E-mail addresses: msee20010@itu.edu.pk (N. Khalid), adnan.qayyum@itu.edu.pk (A. Qayyum), muhammad.bilal@uwe.ac.uk (M. Bilal), aalfuqaha@hbku.edu.qa (A. Al-Fuqaha), jqadir@qu.edu.qa (J. Qadir).

[1] https://sk-pharma.com/the-rise-of-in-silico-trials-in-the-pharmaceutical-industry/

https://doi.org/10.1016/j.compbiomed.2023.106848
Received 6 December 2022; Received in revised form 21 March 2023; Accepted 30 March 2023
Available online 5 April 2023

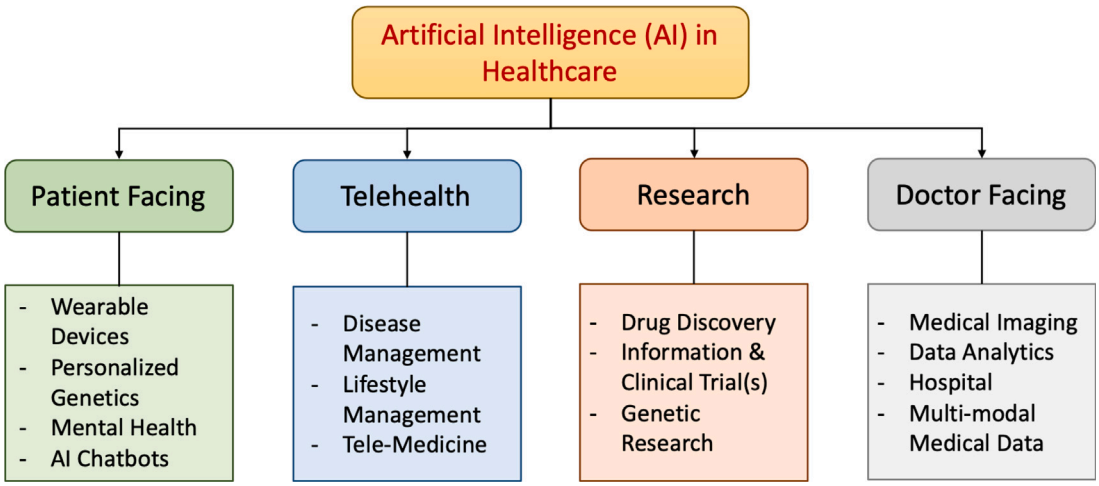

Fig. 1. Illustration of artificial intelligence technology landscape in healthcare.

these algorithms creates enormous concerns regarding data privacy, especially, when data required for AI training encode sensitive and confidential patient information. Any leakage or misuse of data of any sort could result in massive damage to patients, healthcare providers, and software vendors. Most healthcare datasets predominantly contain patient information and there are stringent regulations such as European Union General Data Protection Regulation (GDPR) which shall be respected at all costs while working with these datasets. Google recently faced a class-action style lawsuit for breaching UK data protection law for its AI solution that was created to identify patients at risk of acute kidney injury. There is a huge drive in the healthcare community to revitalize prevailing data management practices and create novel methods for healthcare data to promote AI research while preserving data privacy. An attack on healthcare data sources or AI models that allows the opponent to obtain sensitive and confidential data such as location, health records, or identity information, is highly unwanted and is a significant concern for the users' privacy.

To accrue the potential of AI in healthcare, notwithstanding its human surpassing performance, data privacy and security concerns shall be fully addressed [5]. Many recent studies have highlighted privacy issues in deploying AI-based systems, especially in healthcare. For instance, Hall et al. [6] noted that providers and patients would lose faith in telehealth solutions if the underlying data and technologies do not have proper security and privacy preservation. Tom et al. [7] explained the need for data protection in the age of AI-enabled ophthalmology. The authors focused on the balance between innovation and privacy. Mamdouh et al. [8] highlighted the privacy concerns related to the internet of things (IoT), primarily focusing on how privacy will be affected by the use of IoT in healthcare. Similarly, the privacy and security aspects of medical IoTs have been analyzed in [9].

In this paper, we present a comprehensive survey of existing literature on data privacy for developing healthcare AI systems. We provide a detailed overview of privacy challenges data owners face while sharing datasets with researchers to allow translational AI research in more secure ways. We discuss different attack types and ways they can compromise user privacy. In addition, we also elaborate on potential solutions to overcome these privacy issues. The salient contributions of this paper are the following:

(1) We present a diverse overview of privacy concerns associated with using AI when they are used in developing healthcare applications.
(2) We develop a pipeline of machine learning (ML) techniques for healthcare and show how they can be attacked at each step to compromise the privacy of the developed system.
(3) We present a taxonomy of privacy preservation techniques that can be used to withstand privacy threats.
(4) Finally, we discuss the limitations and pitfalls of existing privacy-preserving techniques and highlight different open research questions that require further development.

Related Surveys: Torkzadehmahani et al. [14] provided a survey on privacy-preserving AI and presented its target application in biomedicine. The authors described different preserving techniques in detail and also highlighted their limitations. Kaissis et al. [13] have provided an overview of different techniques to preserve and secure the AI-based system targetting medical imaging and discussed future aspects of medical imaging. Churi et al. [12] give a brief review on privacy preservation in data publishing, focusing on the healthcare domain. Tanuwidjaja et al. [11] proposed a survey on deep learning (DL) techniques for privacy preservation, their explanation, the challenges, and the pros and cons of each technique. Abouelmehdi et al. [10] presented a survey on security and privacy in extensive data healthcare. The authors described different privacy preservation techniques, challenges, and privacy laws. In this paper, a comprehensive review of privacy-preserving techniques for healthcare is presented. Table 1 provides a comprehensive comparison of this paper with the existing survey and review articles.

Organization of this paper: Section 2 presents an overview of privacy and AI, privacy and healthcare, and associated challenges. Section 3 provides a brief explanation of the types of attacks on ML and DL systems. Section 4 describes the different privacy-preserving techniques, focusing on healthcare domain applications. Section 5 discusses the limitation of using privacy-preserving techniques. Section 6 describes the future direction in this domain. Finally, we conclude the paper in Section 7.

2. Background

2.1. Historical perspective of privacy in medicine

The process of maintaining the security and confidentiality of patient records is known as medical privacy or health privacy. It involves both the security of medical records and the confidentiality of conversations between healthcare professionals. The terms can also refer to the physical privacy of patients from other patients and providers while in a medical facility, and to modesty in medical settings. Modern concerns include the degree of disclosure to insurance companies, employers, and other third parties. Patient care management systems (PCMS) and electronic health record (EHR) have brought about new privacy concerns, which must be balanced with efforts to cut back on

Table 1
Comparison of this paper with existing privacy-focused healthcare surveys.

Reference	Year	Scope				Privacy attacks		Privacy preserving techniques				Challenges	Future directions	Insights and pitfalls
		Applications	Healthcare	ML	DL	Data	Model	Cryptographic	Non cryptographic	Hybrid	Federated learning			
[10]	2018	Big data related	√	√	×	×	×	≈	×	×	×	√	×	×
[11]	2019	All	√	×	√	×	×	√	√	√	×	×	√	×
[12]	2019	Data publishing	√	≈	×	×	×	×	×	×	×	×	≈	×
[13]	2020	Medical imaging	√	√	×	≈	√	√	√	×	√	×	≈	×
[14]	2022	Biomedicine	√	√	√	×	×	√	√	√	√	√	√	×
This paper	2022	Healthcare	√	√	√	√	√	√	√	√	√	√	√	√

√ = discussed, × = not discussed, ≈ = partially discussed, ML = privacy preservation on ML algorithms, DL = privacy preservation on DL algorithms.

duplication of services and medical errors [15]. Many countries have passed laws that aim to protect people's privacy, including Australia, Canada, Turkey, the United Kingdom, the United States, New Zealand, and the Netherlands. Though many of these laws are more effective in theory than in practice. The United States passed the Health Insurance Portability and Accountability Act (HIPAA) in 1996 to strengthen the law to protect healthcare facilities. Also in 2018, the GDPR replaced the Data Protection Directive [16]. In 2012, the European Union (EU) Commission proposed a European Data Protection Regulation to replace the EU Data Protection Directive. The regulation permits EU residents to ask search engines to delink their personal information from the results of a search for their name [17]. On May 25, 2018, GDPR took effect in the EU, and the European Economic Area (EEA), as data protection and privacy regulation. It also applies to personal data transfers outside of the EU and EEA [18–20].

2.2. AI and privacy

In the context of ML and big data, privacy refers to safeguarding against adversarial attacks whose primary goal is to infer sensitive information from the victim, resulting in inadvertent data leaking [5]. Big data has altered the digital world as its effects get more pervasive, with a rising number of enterprises relying on big data analytics for the fulfillment of their everyday operations. In the digital era, our capacity to manage how our data is kept, updated, and shared between parties is critical to our privacy. Over the past years, with the introduction of powerful internet-based data mining tools, privacy has become a pressing social problem. Data privacy and control over personal information are becoming increasingly crucial as a result of the big data explosion and the AI age; the critical components of privacy protection and AI add to the risks to individual privacy [21]. Advanced AI methods like DL are naturally excellent at analyzing massive data sets, and it is probably one of the most efficient ways to analyze large amounts of data in an acceptable period [22].

2.2.1. Data exploitation

The term data exploitation can be defined as the illegal use of individuals' private data. Many consumer products, ranging from smart home appliances to computer software, include features that make them vulnerable to data mining empowered by AI models. The majority of people are unaware of how much data their apps and devices generate, analyze, or trade, resulting in privacy violations. Despite these concerns, recent years have witnessed increasing demand for remote monitoring systems and applications like health-monitoring systems that include different wearable devices being used for data collection related to important medical parameters of individual health such as blood pressure, glucose level, and heart rate, etc. As the demand and dependence on digital technology are new and reliant in this modern era, the possibility of data exploitation increases day by day, resulting in the compromise of the user's privacy [23].

2.2.2. Identification and tracking

In the context of privacy, the terms identification and tracking refer to the uninformed and illegal use of users' private data for identification and tracking purposes. Many consumer products, from computer software to offices, schools, and home appliances, include AI-based features that make them susceptible to privacy concerns and result in the breach of customers' privacy. Attackers can transform data into a weapon by using AI as a misinformation tool. Similarly, in the case of surveillance, which is a binary term originating from the French verb "to keep an eye on". The worst scenario, in this case, is that the people using these devices themselves have no idea how they are sharing their data without their explicit consent, ultimately worsening privacy concerns. Sharing becomes a virtue with digital technology. Simultaneously, governments and corporations have never had a better capacity to monitor people's behavior. Slick AI systems rely on data, and the rise of authoritarianism throughout the world implies that massive data collection might spell tragedy. Identification management is critical for following and serving patients from the minute they walk into a hospital, visit a general practitioner, or visit a healthcare business. Accurate data on the patient is necessary at all times through several applications. However, if this accurate and sensitive data is breached or compromised by an adversary, it results in a privacy violation.

2.2.3. Risks of biometric recognition

Biometrics are the individual personal characteristics that are required to be protected and remain confidential in biometric recognition systems. Nowadays, AI is getting increasingly competent at doing voice [24] and face recognition [25], which are the two fundamental means used for biometric recognition. Biometric information includes a person's face, fingerprints, voice, and iris. Because these criteria are sensitive, no one is obligated to provide them, and no service can be rejected for the same reason. Anonymity in public spaces can easily be jeopardized by using these techniques. With these recognition systems, law enforcement can easily find people without giving people a reason to be suspicious. Voice recognition is frequently utilized in healthcare because it allows patients to get care from the comfort of their own homes, as chatbots, or conversational agents, are computer programs that replicate human text or voice conversations. They are becoming more common in a variety of industries, including healthcare. Chatbots have the potential to improve patient care by offering improved accessibility, personalization, and efficiency [26]. In healthcare, there are several advantages to employing voice recognition technology. However, data required for the operation of such applications can cause serious privacy-related issues. The sensitive data of such applications, if compromised, results in the exposure of complete biometric information and the medical history of patients.

2.2.4. Prediction and profiling

AI's capabilities are not simply restricted to data or analysis. It may also be used to sort, score, categorize, evaluate, and rank individuals using the collected information as input to train AI models. This is frequently done without the individual's classified's agreement, and they frequently have little capacity to influence or contest the findings of these assignments. China's social score system exemplifies how such data might be used to restrict access to finance, employment, housing, and social services [27]. Electronic medical records and evidence-based medicine are two trends that define our era of medicine for patient profiling [28]. However, attacks on electronic health records will result in privacy concerns. Using its advanced techniques, AI can infer or anticipate sensitive information from non-sensitive data. For example, keyboard typing patterns may be used to infer emotional states including uneasiness, confidence, melancholy, and worry. Even more worrisome, data such as activity logs, location data (COVID-19 trace and track apps are good examples), and similar metrics may be used to identify a person's political opinions, ethnic identification, sexual orientation, and even overall health [29].

2.3. Healthcare data breaches

Data breach provides adversary access to confidential, sensitive, or protected information of the user. Medical identity theft and even medical data breaches are increasing at disproportionate rates as cybercrime spreads across industries [30]. Even though all forms of identity theft can cause significant financial harm, medical identity theft can have a direct impact on the patient's physical health. Even in the healthcare professions, the influence of cybercrime has reached extraordinary levels and is proving to be highly destructive. According to the identity theft resource center, 51 healthcare/medical data breach instances were reported in the first few months of 2014 [30].

Over the last few years, there have been a lot of reported privacy-related issues, and data breaches are common in the healthcare sector. The first case of an apparent data privacy breach was observed in 2005 [31]. From 2005 to 2019, the total healthcare data breached was 249.9 million [31]. According to several practitioners, the overall number of people affected by healthcare data breaches was 249.09 million from 2005 to 2019. In the previous five years alone, 157.40 million people have been affected [31]. The most damaging data breach observed in the healthcare domain happened in January 2015, when Anthem released the news that 78.8 million patients' records had been hacked, including their names, ID numbers, and health records [31,32]. Names, Social Security numbers, home locations, and birth dates were the compassionate data stolen from the cyber attack. The victims were mostly Anthem health plan customers, though some were not because Anthem also handled the paperwork for several insurance carriers. In 2015, Excellus suffered a healthcare data breach that affected 10.5 million people [33]. In the same year, the University of California, Los Angeles Health, and Premera Blue Cross [34]. In 2018, AccuDoc Solutions was breached, and the personal healthcare information of 2.7 million people was affected. It is the largest breach in that year [35]. In 2019, American Medical Collection Agency [36], Domition Dental [37] similarly faced data breach. In 2020, Florida Healthy Kids Corporation, 20/20 Eye Care Network, Inc, Forefront Dermatology, S.C. and Eskenazi Health faced major healthcare breaches [38].

The online healthcare system faces a significant challenge in preserving patient privacy [39]. The challenge is to provide digital services protecting patients, Health Information Systems (HIS), and security protection. The crucial and essential stored information in HIS is a wellspring for data breaches and system hacking. The researchers are working to propose a solution for making HIS secure and private [40]. At the same time, the quality of healthcare services has increased with the EHR system. Automated, precise, and timely medical facilities decrease the possibility of data breaches in EHRs. AI-based healthcare systems can be made safe and private by making them hard to hack while they are being developed. This can be done using different techniques to protect privacy (which will be discussed later).

Fig. 2. Challenges towards building privacy-preserving AI in healthcare.

2.4. Challenges in ensuring privacy

Privacy preservation is a tedious task in itself. With the involvement of AI algorithms, privacy-related concerns have increased (as different privacy attacks can be realized on the AI models). A taxonomy of different challenges hindering efficient privacy preservation is presented in Fig. 2.

2.4.1. Adaptability

The privacy-preserving machine learning (PPML) techniques are application-specific, i.e., they are specifically designed for particular ML algorithms and cannot be generalized for all methods. Since ML is an emerging field, new algorithms are introduced every day. Therefore, developers must develop novel methods for preserving privacy concerning the new algorithms [41]. The distributed approach proposed by Papernot et al. [42] or local differential privacy (LDP) is often used as these approaches have been shown to work in the majority of applications. Since 2017, Microsoft has used LDP to capture the number of seconds a user has spent using a certain app on Windows 10. Systems, Applications, and Products in Data Processing (SAP) use local differential privacy to minimize the difficulty and expense of managing a privacy budget [43]. Most privacy-protection techniques cannot be directly adapted to the new algorithm because they are tailored to specific use cases.

2.4.2. Scalability

A problem is also faced when a high-processing-power algorithm is designed to ensure privacy, which gives a good result when tested on small data but takes more time and power for large datasets. ML is advancing towards low processing power, high communication cost, and better speed. While the PPML techniques are obtruded, excessive computational power and communication costs are massive limitations in modern-day usage. For instance, the literature argues that homomorphic encryption is computationally very expensive [44]. The solution to this problem is distributed/parallel processing usage and only transferring the most important information to the algorithm [45].

2.4.3. Legibility

Informing data owners of their data collection is referred to as legibility. The data owners should be provided with complete information about where their data is held and how their privacy is protected. Most companies, like Facebook, Google, and Amazon, employ differential privacy for personal data they store in their systems. It is still challenging to assure users that their data privacy is preserved [45].

2.4.4. AI ethics

The ethical implications of AI-powered solutions are enormous. With the broader adoption of AI across industries, algorithms make crucial decisions affecting human lives. AI ethics advocates ensuring algorithms are fair, unbiased, and transparent. There should be a mechanism to understand the algorithm's underlying decision-making logic. There is a trade-off between developing a highly accurate algorithm and devising a less precise but ethically correct one. In most cases, the confidentiality of patients' data needs to be respected, which negatively affects model performance as crucial patients' sensitive data like genetic biomarkers were not allowed to be utilized during the model training [13]. For a more detailed discussion on the related concepts, we refer interested readers to a recent survey on ethical and trustworthy ML for healthcare [46].

2.4.5. Authentication and access control

Accredited users can get the patient's medical history or confidential information from the EHR stored inside the HIS. The security of authentication data is of paramount importance. If an adversary gets access to such information, the consequences could be severe, as it is pretty challenging to identify such an unaccredited intruder [47].

2.4.6. Data integrity

The accuracy of data plays a crucial role in the delivery of reliable medical AI solutions. Unauthorized intruders holding such vital information might change data, compromising data integrity. Data poisoning usually produces these modifications, resulting in inaccurate results. Thus, protecting data from any poisoning attacks is crucial but a very challenging task [47].

2.4.7. Robustness

The protection of data from tampering is also paramount as models use it to make crucial care decisions like differential diagnosis. An intruder might change the contents of the data and might influence the model's output. A robust mechanism to secure healthcare data should be built into the system to protect against similar attacks, e.g., the EHR system [48]. The patient owns his medical record and has complete access to it. For example, in a medical hospital, each patient's data is kept safe and only the special person assigned to the task has access to the confidential information. The adversary attacks the data and modifies the details of the patient, resulting in invalid owner access. The challenge is to prevent the unauthorized data owner from protecting the privacy of the data owner [49].

2.4.8. Tradeoff between privacy and utility

Assume that a healthcare system wants to disclose its data but first has to de-identify it. They aim to disclose data that is as accurate as possible (minimize utility loss) while also avoiding prejudice (minimize fairness loss). We can observe that there can be variance in the trends among both fairness and utility loss when we look at the outcomes from both artificial and real-world data episodically. This variance might be due to a variety of data properties. Therefore, it is very important to develop methods to address the trade-off between privacy and utility. It will eventually help to improve fairness and privacy [50].

3. Privacy attacks on ML

In this section, we will provide a comprehensive overview of different privacy attacks that can be realized on ML-based systems. Specifically, we have categorized such attacks into two dimensions, i.e., attacks on the data (i.e., data privacy) and attacks on the model (i.e., model privacy). The ML pipeline is presented in Fig. 3 which illustrates these attacks at different stages of the ML pipeline. The complete definition of the attacks in ML is described in the section below.

3.1. Data privacy

In healthcare, data privacy is of utmost importance, and in many cases, the user wants to secure the data before developing or deploying ML-based systems. Let us take an example of a medical study or a model trained for hospital specialists from the EHR. If the data owner and the calculation party are different, the private data would be sent to the computation party over a secure channel. It would, however, most likely be stored in its original form on the compute server, meaning it would not be encrypted or altered. The confidential data would be vulnerable to insider and outsider attacks, making this the most severe threat. The privacy includes the features, membership, and exact values of the data. The data privacy attack has three types: re-identification, reconstruction attacks, and property inference attacks, which are described next.

3.1.1. Re-identification/de-anonymization

A re-identification attack is the reverse of the de-identification method. In this attack, the source is providing data, and the data is identified. For example, attributes of heart disease are collected using the records of different patients' medical history if the attack is performed on the results of this system in patients' identification, including their names and medical records.

3.1.2. Reconstruction attacks

A reconstruction attack is a privacy attack in which a significant portion of the raw dataset is constructed again. Reconstruction is most effortlessly perceived by considering the dataset as an assortment of lines, one for every person. Assume that each column contains a significant amount of non-private recognizing data as well as a mysterious piece, one for each individual [51]. For example, whether or not a person has the gene for Alzheimer's illness. The main target of the reconstruction attack is to find the mysterious bits of all the people in the dataset.

3.1.3. Property inference attacks

The capacity to separate dataset properties that were not explicitly encoded as features or were not connected to the learning task is called a property inference attack. The ratio of women and men in the healthcare dataset is not the dataset's attribute. Face recognition using neural networks is a good example of a property inference attack. For example, when the model was trained, the dataset did not include any information about how many people wore glasses, which would be excluded by the attacker.

3.2. Model privacy

In the ML model, privacy concerns include securing both the model parameters and training algorithms. A standard ML model serves practice through cloud providers such as Google Cloud Platform, Microsoft Azure, or Amazon SageMaker. Models, in that case, are deployed via the ML API services, and users get charged per API usage. If their models are revealed, they will suffer a significant loss. The attackers can target the infrastructure where the model has been deployed or the model's parameters. Similarly, the other companies that are providing online

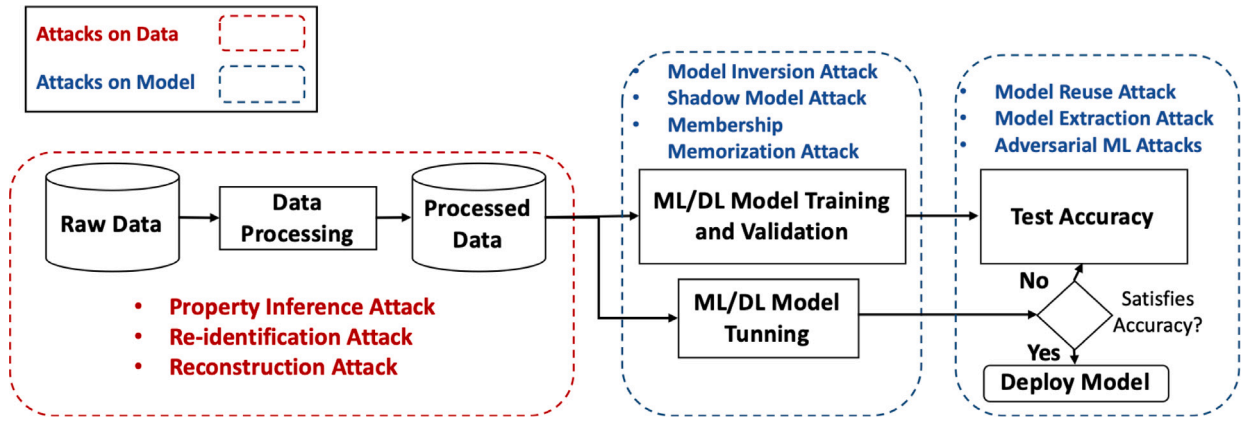

Fig. 3. Pipeline of the privacy attacks in ML techniques.

services, if their model architecture or even parameters are revealed, will suffer a great loss [52]. In the literature, the security and privacy-related implications of cloud-hosted ML models have already been studied and different types of attacks have been formulated [53]. This raises the issue of model privacy for AI-based healthcare applications, as patient privacy must not be compromised at any cost.

3.2.1. Model extraction attack

Model extraction is also known as black-box attacks, and the adversary aims to reconstruct the model by extracting the information. The substitute model is generated using this information, which performs similarly to the original model under the attack. The substitute model's main task is to create a model that accurately matches the original model during the testing phase. The data is taken from the input distribution and during the learning process.

3.2.2. Membership inference attack

Shokri et al. [54] proposed a membership inference attack, in which the attacker's main aim is to determine if an input y is part of the training set or not. The episode is realized on the DL model trained in the supervised learning strategy while assuming black-box settings. For instance, recognizing a person's participation in an emergency clinic's wellness examination training set reveals whether this individual was previously a patient there or not. White-box attacks can also be a threat if the attacker has access to the model's parameters and gradients, which will allow for a more constructive white-box inference attack that will significantly reduce the accuracy.

3.2.3. Model inversion attack

In the model inversion attack, the attacker aims to construct training data from the model predictions, exposing the privacy of the sensitive records. For example, a malicious person tries to recover the secret dataset used to train a supervised neural network in model inversion attacks. If a model inversion attack works, it should produce samples that are realistic and varied and that accurately describe each class in the private dataset.

3.2.4. Shadow model attack

A shadow model attack is very similar to an inference model attack and can be considered a sub-type of membership inference attack. In this attack, the adversary attempts to learn the features and statistical properties of the model using the shadow model (SM). SM imitates the target model, but we know the training dataset in this case. The adversary can target both the white or black-box attacks. Then the training of the attack model is done on the input and output labels of SM [54].

3.2.5. Adversarial ML attacks

In adversarial ML attacks, the trained ML/DL models are fed with such samples that contain carefully crafted imperceptible adversarial perturbations [53]. The key objective of such attacks is either to mislead the model's decisions or to achieve the intended outcomes. An untargeted adversarial attack is the most generic type of attack where the only objective is to cause the classifier to increase classification error. On the other hand, the targeted adversarial attack is a more challenging attack in which the aim is to get an input sample misclassified into the target class.

3.2.6. Membership memorization attack (MMA)

Song et al. [55] proposed a new attack on the models that memorize a lot during their training. The MMA determines whether or not a particular data record was included in the model's training dataset. When an opponent has complete knowledge of a record, discovering that it was used to train a specific model indicates information leaking through the model. It can, in rare situations, directly result in a data breach.

3.2.7. Model-reuse attacks

Model-reuse attacks in which deliberately created primitive models ("adversarial models") infect host ML systems and cause them to fail in a very predictable manner when targeted inputs ("triggers") are used. The system consists of a feature extractor and a classifier. The attacker wants to implement the backdoor logic, so they create the malicious feature extractors. In the model reuse attack, the attacker has trigger input and one of the target classes, and the extractor does not know the classifier or tuning part.

3.3. Overview of privacy attacks in healthcare

In the literature, different types of privacy attacks have been realized on the AI models trained using health data. For instance, Alam et al. [56] proposed a person re-identification attack that is the most concerning issue in publicly shared data by HIPPA. The author uses temporal and spatial information separately and then uses that information to identify the person using the designed framework. The proposed framework uses the Multi-Modal Siamese Convolutional Neural Network (mmSNN) model. The proposed framework shows that physicians provide group (PPG) based breathing rate and heart rate in conjunction with hand gesture contexts to be utilized by attackers to re-identify the users from HIPAA-compliant wearable data. The author uses the datasets Gamer's Fatigue Dataset, Restaurant Data, Older Adult Data, and Healthy Adult Fatigue Data. The method achieved 65% accuracy from all the datasets to identify the person. In a similar study, Karmaker et al. [57] presented a broad probabilistic re-identification framework that can be used to evaluate the likelihood of compromises based on

Table 2
Privacy attack in the healthcare domain.

Reference	Year	Attack type	Attack	Technique	Performance	Dataset	Target area
[56]	2021	Data	RI	Used multimodal Siamese neural network to learn spatial and temporal information separately and then used them to identify the person.	The results show that the proposed framework provides 65% accuracy on all datasets to reidentify a person.	Gamer's Fatigue, Restaurant data, Older Adult, and Healthy Adult Fatigue datasets.	Reidentification of person publicly shared healthcare datasets.
[57]	2018	Data	RI	Probabilistic naive re-identification the the framework that may be used to evaluate the likelihood of compromises based on explicit assumptions in specific scenarios	Compared to medical condition attributes, demographic attributes were shown to be more likely to be disclosed. Meningitis was discovered to be the most commonly disclosed anemia was the least common in social media data.	De-identified medical research data set, the HCUP National (Nationwide) Inpatient Sample (NIS)	Disclosing medical and demographic attributes on the social media.
[58]	2021	Model	AD	Hierarchical position selection, which uses RL framework to pick the attacked positions, and substitute the selection which uses a score-based method to identify substitutes.	The victim model is HiTANet, MedAttacker consistently achieves the best attack success rate, with success rates of 3.08%, 2.20%, and 1.74% higher than the other methods.	Heart failure, Kidney diseases, and Dementia.	Electronic health care system.
[59]	2020	Model	AD	A new framework of AD attack with the use of ML as an adversary with only a rudimentary understanding of data distribution. The model can change patient status in the healthcare system.	The adversarial attack is done on Whitebox and BlackBox. On decision tree highest accuracy drop is observed that is 32.27% using the HopSkipJump while the success rate is 15.68%.	Social Media health data	Smart healthcare system.
[60]	2020	Model	AD	Different AD attacks are implemented on the COVID-19 detection approaches that use DL approaches.	The tests reveal that DL models that ignore protective mechanisms against adversarial perturbations are nevertheless vulnerable to adversarial attacks.	COVID-19	Medical domain
[61]	2022	Model	MI	The problem from the standpoint of the data owner, who wants to estimate the risk of the disclosure before releasing any health data.	The partial synthetic data is vulnerable to the attack at a very high rate than fully synthetic data.	Datasets derived from several health data resources	EHR
[62]	2021	Model	MI	Realistic inference attack on the DL a model trained on 3D neuroimaging.	Properly identified if an MRI scan was used in model training with a 60% to over 80% success rate	MRI images	3D neuro-imaging
[63]	2021	Model	MI	MI framework is used to test the empirical privacy leakage	Membership inference attacks on CLMs result in non-trivial privacy leakages of up to 7%, according to the author's findings.	MIMIC III, UMM, VHA	Clinical language processing (CLMs)
[64]	2019	Model	MI	The mimic model behaves similarly to the public model in terms of prediction, and is used to reveal the discrepancies in prediction between the training and testing datasets	Attack performance against XGBoost-trained ML models, logistics, and an online cloud platform. Achieve inference accuracy & the precision of 73% and 84% on average, and 83% and 91% at best, using genuine data	Weibo dataset	Health data
[65]	2022	Model	Mi	Inversion framework that builds on the fundamentals of gradient-based model inversion attacks.	Outperform existing gradient-based approaches both in a quantitative and qualitative manner.	BraTS	Brain tumor segmentation

Re-identification (RI) attack, Model inversion attack = MI, Membership inference attack = MI, and Adversarial Attack = AD.

explicit assumptions in specific scenarios. The authors proposed a set of beliefs that can be used to produce a first-cut risk estimate for practical case studies. The Naive Re-identification Framework (NRF) is the name given to the framework based on these assumptions. The author uses NRF to investigate and quantify the risk of re-identification that arises from releasing de-identified medical data in the context of publicly available social media data as a case study. The results of this case study show that NRF can be used to get a reasonable estimate of re-identification risk, compare the risks of different social media, and look at how dangerous it is for people to share information about their demographics and medical conditions on social media.

To exploit the ML classifiers used in an intelligent healthcare system (IHS), a new type of adversarial attack where an adversary with only a rudimentary understanding of data distribution, and the AI model can realize both targeted and untargeted attacks is presented in [59]. Adversarial attacks adjust medical device readings in the IHS, resulting in a change in patient status (disease-affected, average condition, activities, etc.). On an IHS, the attack employs five adversarial ML algorithms (HopSkipJump, Fast Gradient Method, Crafting Decision Tree, Carlini & Wagner, Zeroth Order Optimization) to carry out various malicious behaviors (e.g., data poisoning, misclassifying outputs, etc.). The author undertakes white-box and black-box attacks on an IHS based on an adversary's training and testing phase capabilities. The suggested adversarial approach can dramatically decrease the effectiveness of evaluating whether a sample was used to train the model. The proposed adversarial approach can significantly reduce the effectiveness of assessing whether a sample was used to train the model. The author used the (de-identified) medical research data set. The author used the National (Nationwide) Inpatient Sample (NIS) provided by Healthcare Cost and Utilization Project (HCUP) to test the attack. They showed that a simple access to the model prediction only (i.e., black box settings), access to the model itself (i.e., white box settings), or a leaked sample from the training data distribution can be used for this insight-based ML-based IHS in correctly identifying illnesses and expected behaviors of patients, ultimately leading to erroneous therapy.

Liu et al. [64] introduced SocInf, with an emphasis on membership inference as a basic issue. SocInf's main idea is to build a mimic model that behaves similarly to the public model in terms of prediction, and

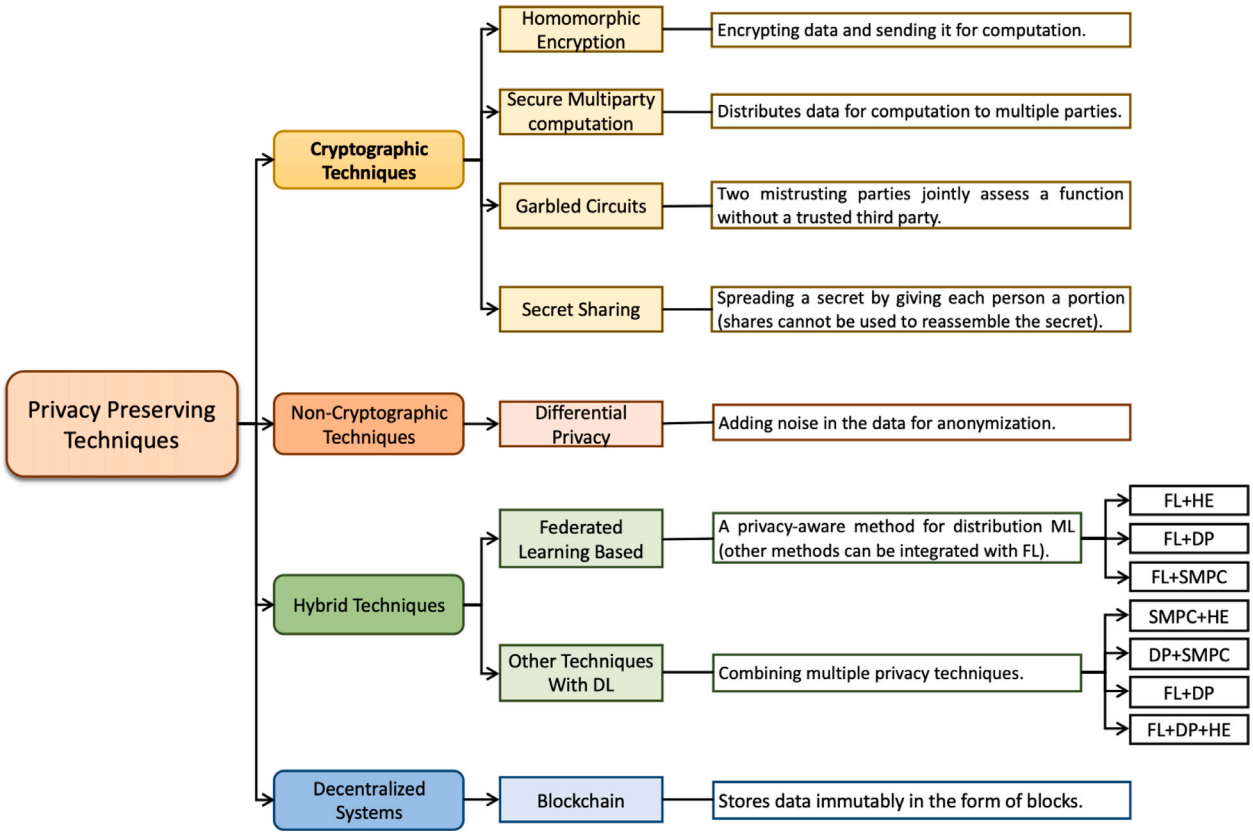

Fig. 4. Taxonomy of privacy-preserving techniques. (***Legends***: Homomorphic Encryption = HE, Secure Multiparty Computation = SMPC, Differential Privacy = DP, Federated Learning = FL, Deep Learning = DL).

then use the mimic model to reveal the discrepancies in prediction between the training and testing data sets. The targeted attack is on social media health data. SocInf can determine if a particular record is in the victim model's training set or not by using extensive analytics on the mimic model's predictions. SocInf's attack performance is tested against XGBoost-trained ML models, logistics, and an online cloud platform. The experiment findings reveal that SocInf can achieve inference accuracy and precision of 73% and 84%, on average, and 83% and 91%, at best, using genuine data. Clinical language models (CLMs) have been used to improve performance in biomedical natural language processing tasks using clinical data. Jagannatha et al. [63] through white-box or black-box access to CLMs, look at the dangers of training-data leakage. For DL models like Bidirectional Encoder Representations from Transformer (BERT) and Generative Pre-trained Transformer (GPT), they realized membership inference attacks to estimate empirical privacy leakage. Membership inference attacks on CLMs result in non-trivial privacy leakages of up to 7%, according to the author's findings. Three datasets were used for experimentation, namely MIMIC III, UMM (UMass Memorial Health Care), and VHA (Veterans Health Administration) hospitals.

Zhang et al. [61] approached the membership inference problem from the standpoint of the data owner, who wants to estimate the risk of disclosure before releasing any health data. Usynin et al. [65] present a new model inversion framework that builds on the fundamentals of gradient-based model inversion attacks but also depends on matching the reconstructed image's attributes and style to data controlled by an adversary. While keeping the same honest but curious threat paradigm, the author's strategy surpasses previous gradient-based techniques both qualitatively and statistically, allowing the attacker to gain enhanced reconstructions while staying undetected. MedAttacker is the first black-box adversarial attack to test the vulnerability of health risk prediction algorithms. Ye et al. [58] proposed a medAttacker that handles the issues posed by EHR data in two ways: hierarchical position selection, which uses a reinforcement learning (RL) framework to pick the attacked positions, and substitute selection, which uses a score-based method to identify substitutes. It initializes its RL position selection policy by exploiting the temporal context inside EHRs, in particular, each visit's contribution score and the saliency score of each code, which may be easily linked with the score-based deterministic substitution selection process changes. Real-world health insurance claim datasets, included heart failure, renal illness, and dementia. For HiTANet as victim model, which uses time information for health risk prediction, MedAttacker consistently achieves the best attack success rate, with success rates of 3.08%, 2.20%, and 1.74% higher than the second-best method patient wise weighted sampling (PWWS) in the heart failure, kidney disease, and dementia datasets, respectively.

Gupta et al. [62] demonstrated a realistic membership inference attack on DL models trained for 3D neuroimaging applications. The author predicted that brain age prediction models (deep learning models that predict a person's age based on a brain MRI scan) are vulnerable to attacks. Depending on model complexity and security assumptions, the author properly identified if an MRI scan was used in model training with a 60% to over 80% success rate. The experiment was done on T1 structural MRI scans of healthy subjects from the UK Biobank dataset. Rahman et al. [60] with appropriate adversarial scenarios, the author tested several COVID-19 diagnostic approaches that rely on DL algorithms. The results of the tests reveal that DL models that ignore protective mechanisms against adversarial perturbations are nevertheless vulnerable to adversarial attacks. Finally, the author went over the adversarial generation process, the attack model implementation, and the changes to the existing DL-based COVID-19 diagnostic applications in detail. Throughout the survey, we have observed that the attacks faced by healthcare are usually adversarial attacks, model inversion, membership inference attacks, and re-identification attacks. Table 2

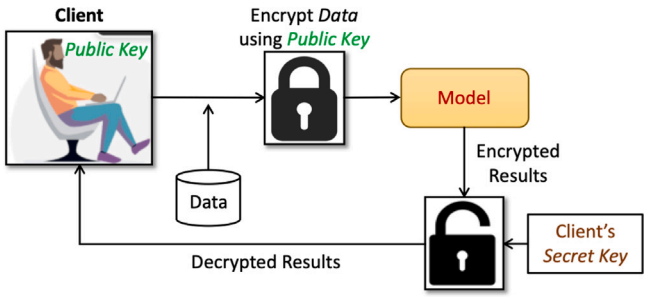

Fig. 5. Working of homomorphic encryption technique.

presents the summary of different privacy attacks in the healthcare domain.

4. Privacy preserving techniques

Developing reliable healthcare AI systems using ML for different clinical tasks requires large amounts of carefully curated data. A key barrier in the path of broader AI adoption is data access and its implementation at the commercial level. PPML ameliorates privacy techniques and infrastructure for improved security for data and ML, which is of utmost importance to overcome these barriers. This is important, especially in the case of the sensitive dataset, for the invention of better life-saving treatment, better diagnosis, and the selection of the right decision in an urgent situation. Numerous privacy-preserving methods allowed different information gatherings to cooperatively train ML models without delivering their private information in its unique structure. This was essentially performed by using cryptographic methodologies or differentially private information release [66]. Table 3 shows a comparison between privacy preservation techniques. In the training and testing of AI models, data privacy is most important when dealing with confidential or sensitive data. However, to achieve perfectly privacy-preserving AI, there are four pillars of PPML, i.e., training data privacy, input privacy, output privacy, and model privacy. The first three deal with data creator privacy, and the last protects the model creator's privacy. Fig. 4 describes the taxonomy of privacy preservation techniques.

4.1. Cryptographic techniques

The word cryptography comes from the Greek word kryptos, which means "hidden". Cryptography studies communication methods or strategies that assure secure data delivery from the sender to the recipient. The content is safely shared between the sender and the recipient while maintaining its confidentiality and integrity. The primary mechanism employed in this strategy for privacy preservation is encryption [67].

4.1.1. Homomorphic encryption

Homomorphic encryption (HE) is the technique in which the data owner encrypts its data and sends it for computation. The computational tasks are performed on the data without decrypting the data, and the output results of encrypted data are sent to the data owner. Fig. 5 explains the general operation of the HE that allows computation on encrypted data while preserving data attributes, allowing a third party to perform algorithms on the data without having inside knowledge. For example, if one wants to classify the data using the ML approach, he can operate on the data without knowing the actual data. Genomic data is increasingly used to train reliable ML models for precision medicine and stratified healthcare as DNA reveals the crucial relationship between biomarkers and diseases. But this data is susceptible and requires safeguarding.

Chowdhury et al. [68] proposed a new cipher, called DeCrypt, which was based on Triple Data Encryption Standard (3DES) that was resilient to a man-in-the-middle attack. Based on the results of the experiments, the DeCrypt cipher offers superior long-term security against sweet-32 attacks and is 61% faster than 3DES. When compared to 3DES, the proposed algorithm is more efficient because it requires less time to perform both encryption and decryption. Sarkar et al. [69] provide the algorithm that gives secure, fast, and private genome imputation using the ML approach. To ensure the privacy-preserving of the genome imputation, ML algorithms are used with homomorphic encryption techniques. The linear models are converted into encrypted models using homomorphic encryption techniques. The Paillier cryptosystem is used for encryption; this method selects a random number and raises its power to N. This eliminates the need for every N-th power for new encryption, making it a faster algorithm for encryption. The privacy-preserving method proposed performs 99% equivalent to the state-of-the-art plaintext solutions. Paul et al. [70] created the collective learning protocol, which is a secure system for exchanging classified time-series data inside an organization's entities to partly train binary classifier model parameters. Each data characteristic is encrypted by the protocol. The protocol is performed on the Medical Information Mart for Intensive Care (MIMIC-III) dataset [70] and the CKKS encryption scheme was used).

Lu et al. [78] focus on the privacy preservation of the patients sharing their data. The cryptographic techniques used for this purpose mainly aim to target the cloud environment for genomic data sharing. To ensure privacy, both genotype and phenotype are encrypted using homomorphic encryption. All the data statistics are shown using the frequency tables; the proposed method shows the statistics of encrypted data in the same format using frequency tables training the healthcare data, i.e., clinical data and genome data privacy-preserving. Carpov et al. [79] present a method to preserve the secrecy of data shared on the cloud for computation, i.e., many wireless gadgets or apps send data of patients for regular monitoring of their health using different algorithms. In this method, a mobile app is developed to share the client's data with the cloud while HE is used as an encrypting tool to secure the data from attacks or to evaluate the data privately.

4.1.2. Secure Multiparty Computation (SMPC)

SMPC represents a subfield of cryptography that performs data computation by distributing the data between different parties. Each party applying the algorithm to its secure data without knowing the rest of the data results in privacy preservation. For example, three employees want to calculate their average salary while maintaining privacy during the computation. They solve this issue by using the SMPC algorithm known as additive secret sharing [80]. Fig. 6 depicts the general process of the SMPC technique.

Each person breaks their salary into three parts to secure their data and keeps one part of the data for himself and shares the other two parts, one with each person. As a result, each gets three data parts. The data is useless for the other person when computation is done, as he does not know about the rest of the data, while it is useful for the person knowing the complete data, as he can get the result by adding the results of the distributed data. In this way, data privacy is ensured [81].

Akgun et al. [82] offer a way of querying genomic datasets in a privacy-protected manner using secure multi-party computing. The suggested technique allows genomic databases to be safely maintained in semi-honest cloud settings by privately outsourcing genomic data from an unlimited number of sources to the two non-colluding proxies. It uses XOR-based sharing to provide data privacy, query privacy, and output privacy, and unlike earlier systems, it allows searches to be performed effectively on hundreds of thousands of genomic data points. The framework functions as a virtual machine, abstracting safe computing protocols. The benefit of the implementation of the Garbled

Table 3
Comparison of privacy preserving techniques.

Privacy preservation techniques	Reference	Description	Advantages	Disadvantages
Homorphic Encryption	[71]	Encryption of data is done by the data owner and the data is decrypted after all the computation is done	Allows secure, efficient cloud use, collaboration with a third party. HE can be used to receive outsourced services for research and analysis without risking non-compliance.	Slow, Either require programmed or dedicated client–server app to work properly.
Secure Multiparty Computation	[72]	Data computation by distributing the data between different parties. Each party applies the algorithm to its secure data without knowing the rest of the data results in the privacy preservation	A significant number of the parties are malevolent and conspiratorial, the input data will stay private even if it is searched for an indefinite period of time and resources	Communication Overhead, In the computation, assumptions must be made regarding the proportions of malevolent coordinating parties.
Garbled Circuits	[73]	Allows two mistrusting parties to jointly evaluate a function using their own private inputs without the requirement for a trusted third party	Low round complexity, low latency, and, most critically, relative technical simplicity	Not reusable all the variants are used for one time. Privacy is compromised if more than one input is given for the particular circuit.
Secret Sharing	[74]	Technique of spreading the share among group of people	Ideal for highly sensitive and highly important information	Large size of shares
Differential Privacy	[75]	Adding the noise into data to introduce the data anonymity. All the commutation analysis is done on the anonymous data without any client's identity revelation	Due to its composability, tolerance to post-processing, and graceful deterioration in the presence of correlated data, provides strong and resilient assurances that permit modular design and study of differentially private methods	Greater noise infusion than previous methods. This is because standard approaches just need to prevent linking, whereas differential privacy prevents linkage by preventing reconstruction
Federated Learning	[76]	The data is distributed to different groups and companies, forming different datasets. The training using this kind of dataset results in local privacy	Data protection: Keeping the training dataset on the devices eliminates the need for a data pool for the model. Continual learning in real-time, data diversity	Attacks and failure in robustness, need to improve efficiency and effectiveness
Blockchain	[77]	This technique is private and shares personal data protect user data with private key encryption and zero-knowledge proofs.	Decentralization, Immutability, Transparency, and Access control.	Complexity, publicly available blockchains, scalability, and not much secure to data breach.

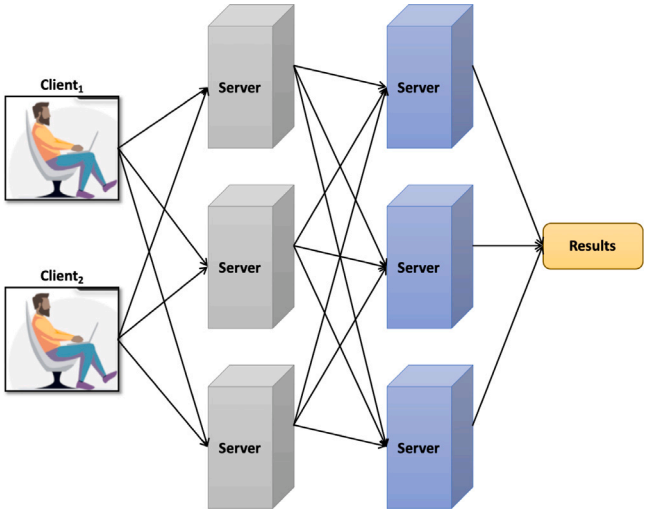

Fig. 6. Working of Secure Multiparty Computation technique.

Circuit (GC) protocol in the arithmetic and boolean circuits in the YASHAM (ABY) framework is that it has XOR-based sharing [83].

Li et al. [84] proposed a novel SMPC system model that consists of two servers. First, the patient encrypts their health record using HE and sends its data to the hospital server. The hospital processes the encrypted data and finds the traits of disease from its database that match the patient's health record and encrypts the data using HE and symmetric encryption algorithms. Kumar et al. [85] proposed a system to ensure privacy in the e-healthcare system. The confidential data is shared by the patient with the hospital or clinic server using an online mode of operation. The Paillier encryption algorithm is used for encryption purposes, shared with the hospital server and matched with disease records. Privacy preservation is made more secure in the model as it deals with many other threats. Marwan et al. [86] designed the framework using the Paillier homomorphic algorithm for data encryption. This technique is used for distributed databases for computation, which allows the hospitals to collectively calculate the number of people affected by the pandemic without revealing confidential information. Jangde et al. [87] highlighted that the threats of fraud in the healthcare domain can be overcome using SMPC. Due to its simplicity, the method is better than the previous ones.

4.1.3. Garbled circuits (GC)

A garbled circuit is a cryptographic approach that allows two distrusting parties to jointly evaluate a function over their private inputs without the requirement for a third party. Using the garbled circuit protocol, the function must be written as a Boolean circuit. Fig. 7 explains the generic working of the technique.

In [88], Yao first proposes a garbled circuit as a secure two-party computation solution. Sancho et al. [89] offers a distributed access control system that uses Garbled Circuits to execute an extensible Access Control Markup Language (XACML) like policy assessment across organizations, guaranteeing that no properties of one organization may be learned by others. Because garbled circuits are employed as the underlying protocol, an analysis of the suggested method's complexity in terms of non-XOR gates is presented. There are also some examples

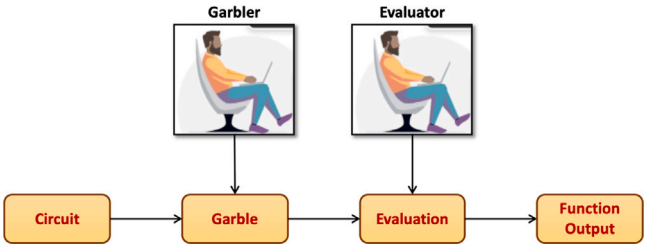

Fig. 7. Working of garbled circuits technique.

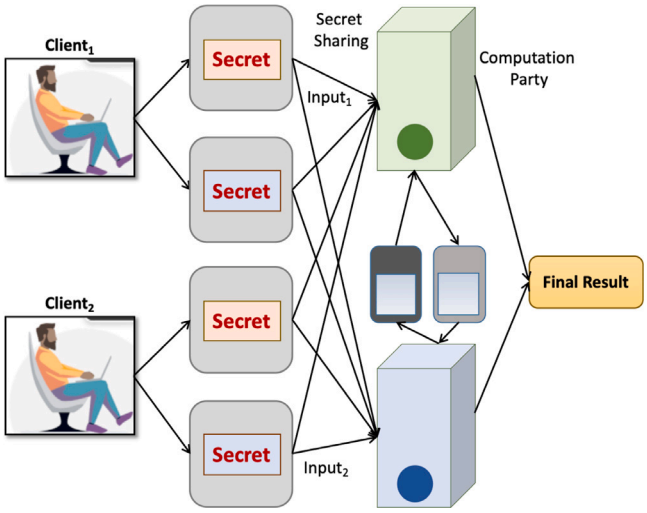

Fig. 8. Working of secret sharing technique.

of how much it would cost to use this method with different policy sizes, as well as a rough estimate of how many evaluations could be done per second.

Gong et al. [90] present a method for genome-aware health monitoring that is both private and secure. Users can only learn diagnostic outcomes based on their genetic and biological sensing data using the proposed method, while the healthcare practitioner learns nothing. Security analyses and performance assessments were carried out to demonstrate the efficacy and efficiency of the suggested technique. The author uses the GC to implement a subprotocol to compare two integers. Barni et al. [91] proposed two alternative electrocardiograms (ECG) classification methods are presented: the first is based on a quadratic discriminant function classifier and is implemented using a hybrid technique that combines homomorphic encryption and garbled circuit theories; the second is based on a neural network (NN) classifier and exclusively uses garbled circuit constructs. This method worked because NN could limit the size of the input, output, and internal values of the calculation in bits.

4.1.4. Secret sharing

Secret sharing (also known as secret splitting) is a method of disseminating information among a group of individuals, with each person receiving a piece of the information. Individual shares are useless on their own, thus the secret can only be reconstructed when a huge number of them, maybe of diverse kinds, are united together. Fig. 8 represents the general working principle of secret sharing.

Dey et al. [92] proposed a technique that aims to address the issue related to the electronic health system. The method is used to strengthen the electronic health system against tricksters. The algorithm uses a perceptron-based session key and a logistic map-based intermediate key was proposed. A lossless strict secret-sharing methodology is used to protect clinical data and patient privacy. Simple

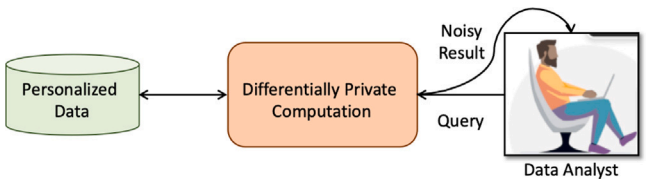

Fig. 9. Working of differential privacy technique.

mathematical operations are used to create the secret shares. The system is tested on the electronic health system of COVID-19.

Sarosh et al. [93] proposed a distributed security module for the protection of clinical images that contributes to 80% of medical data. The Rivest Cipher 6 (RC6) encryption algorithm is used with the computational secret-sharing scheme for the storage of the images in a distributed manner. Perfect secret sharing (PSS) is used to share the key. Using PSS, the n images and $n-1$ key shares can be made public because of this efficient algorithm. The remaining key shares can be made protected and secure using the Deoxyribonucleic Acid (DNA) substitution method. Analysis of the proposed scheme ensures the robustness of the scheme over the state of the art against attack. Anand et al. [94] proposed using Non-Subsampled Contourlet Transform (NSCT) and Multiresolution Singular Value Decomposition, to develop a robust X-ray image watermarking system. The maximum entropy component of the X-ray carrier image is first deconstructed using NSCT for watermark embedding. Multi-Scale Singular Value Decomposition (MSVD) is then used to retrieve low and high-frequency features of the carrier and mark the image. Moreover, the watermark detail is hidden by altering the carrier image's detail with the appropriate factor. Finally, to obtain a secure tagged carrier picture, Shamir's (k, n) secret-sharing procedure is used. Objective assessments of 200 X-ray images of COVID-19 patients demonstrate that the suggested method not only has outstanding invisibility but also has a high level of robustness against diverse attacks.

4.2. Non cryptographic techniques

4.2.1. Differential privacy (DP)

DP is a privacy-preserving technique that provides data anonymity by introducing noise into the data. All the commutation analysis is done on the anonymous data without any client's identity revelation. The dataset usually consists of an enormous amount of information about individuals, so privacy preservation is necessary for confidentiality [95]. For example, Apple uses DP for privacy preservation to collect data from devices like Macs, iPads, etc. Amazon applies the DP algorithm to get the customer's personal preference for shopping. Behavioral data was accumulated by Facebook using the DP to preserve the confidentiality law of the country. Fig. 9 explains the working of DP.

Sangeetha et al. [96] proposed the work in which instead of the data release, a DP-based model release with Support Vector Machine (SVM), Random Forest method, Logistic Regression, K-Nearest Neighbor, Decision Tree, and Naive Bayes are six ML classifiers suggested for a private model release. The model's accuracy is assessed by an experimental evaluation utilizing the benchmark heart disease dataset. The publicly available private model can be used to predict heart disease in patients. The heart disease dataset from the UCI ML repository was used in this study.

Ziller et al. [97] presents an open-source parallelized deep neural network (DNN) framework for the modification of per sample gradient. Privacy is preserved using the Gaussian DP system, and modification automation is ensured by using the shared memory of the neural network weights. Medical image segmentation, assess its application to the paediatric pneumonia dataset, an image classification task, and the Medical Segmentation Decathlon Liver dataset.

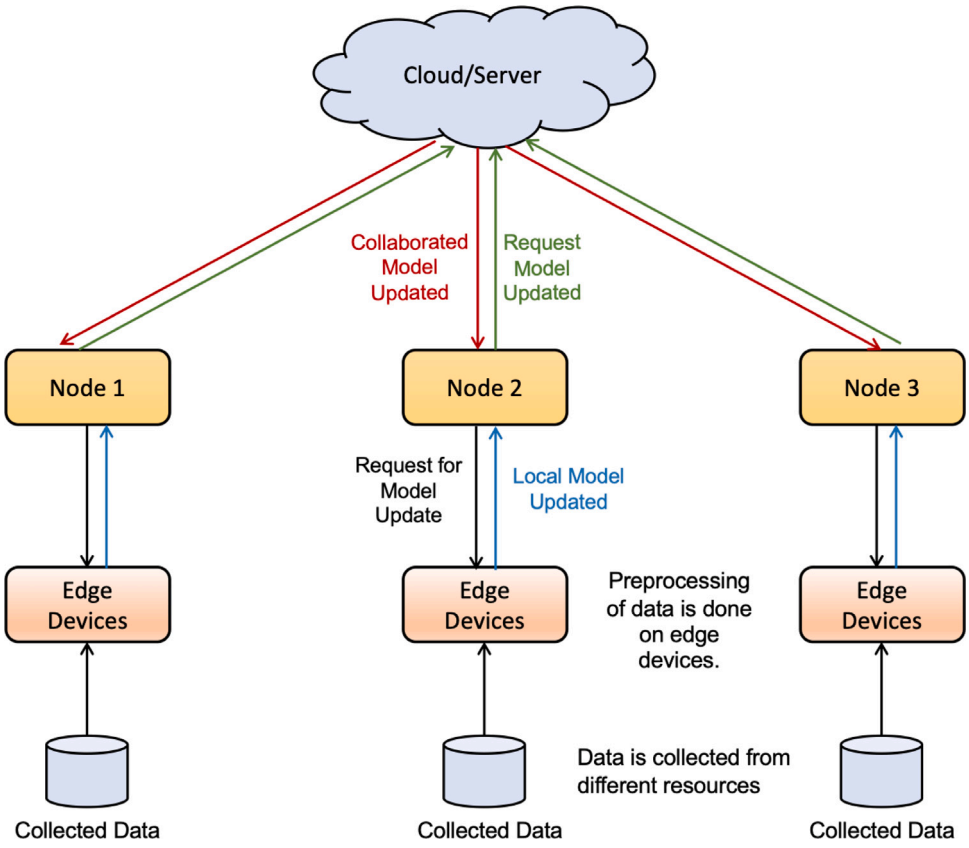

Fig. 10. Privacy preservation using federated learning.

Muftuoug et al. [98] proposed the method of ensuring privacy for the medical and location data collected from wearable devices or any other instrument. A total of 139 COVID-19-infected chest X-ray (CXR) images were collected, and a total of 373 public data sources were utilized to develop a diagnosis concept. It was trained using EfficientNet-B0, a powerful deep-learning network for making quick and effective decisions for medical providers during the PCR test. The EfficientNet-B0 model is used to diagnose COVID-19 from CXR images, and DP is implemented on it. In the Private Aggregation of Teacher Ensembles (PATE) methodology, the authors used the DP method to assure privacy. Their method achieved an accuracy of 71% while the original model's accuracy was 94.7%. Even though accuracy is compromised, promising results were obtained in preserving the privacy of the data.

Vadavalli et al. [99] proposed an algorithm that is not only used for breast cancer detection but also deals with the privacy concerns of the patient. The algorithm uses the DP approach. This algorithm modifies the important information about patients in the data collection. As a consequence, it not only increases the privacy of patient information when compared to other ways, but it also aids in obtaining accurate findings. The benefit is that it will have the least influence on the truth-finding approach's accuracy because the suitable characteristics for the cancer prediction data set will not be changed. This suggested system uses the Expectation–Maximization Algorithm to forecast the truth-finding. This is the first time that a differential privacy protection technique and a truth-finding strategy have been combined in this way. The proposed method has provided an accuracy of 96.14%.

4.2.2. Federated learning (FL)

FL seeks to create a combined ML model based on data from different locations. In FL, there are two processes: model training and model inference. Information, but not data, can be transferred between parties throughout the model training process. At each location, the exchange does not expose any protected private sections of the data. The trained model might be kept by one party or shared by several. The model is used to infer a new data instance during inference time. In a business-to-business (B2B) environment, for example, a federated medical-imaging system may receive a new patient with diagnoses from many institutions. The parties work together to make a prognosis in this scenario. Finally, a fair value-distribution system should be in place to share the benefits generated by the collaborative approach. Mechanisms should be designed in a way that ensures the federation's long-term viability. In general, FL is an algorithmic framework for developing ML models that has the following characteristics: a model is a function that maps a data instance at one party to an outcome [100]. Fig. 10 represents the general idea of FL.

- Two or more parties are interested in working together to create an ML model. Each side has data that it would want to contribute to the model's training.
- During the model-training process, each party's data remains with that party.
- The model can be partially transmitted from one party to another using an encryption strategy that prevents other parties from re-engineering the data at any particular party.
- The resulting model's performance is a close approximation of an ideal model developed with all data transferred to a single party.

FL comes with several built-in privacy features. The raw data stays on the device, while updates transmitted to the server are focused on a specific purpose, transitory, and aggregated as quickly as possible in the spirit of data reduction. In the case of EHR, FL algorithms can aid in finding patients with a similar medical history [101]. In [102], the authors elaborated on how FL techniques can be used in hospitals to help in taking records of how many patients are admitted to the intensive care unit (ICU) and even help in finding the approximate stay

of patients in the hospital. The wonders of FL are also prominent in the field of medical imaging or even segmentation techniques, i.e., lung image segmentation or breast tumor segmentation.

Dou et al. [103] proposed a system with external validation on patients from a worldwide study, demonstrating the efficacy of an FL system for detecting COVID-19-linked computed tomography (CT) anomalies. Investigate FL strategies to create a privacy-preserving AI model for COVID-19 medical image diagnostics that can generalize well to new datasets. During pandemics, FL might be a valuable technique for quickly developing therapeutically relevant AI across institutions and nations, alleviating the load of centralized aggregation of enormous volumes of sensitive data. The model was pre-trained using the DeepLesion dataset and then fine-tuned with COVID-19 internal training images. Qayyum et al. [104] for an automated diagnosis of COVID-19 used the developing idea of clustered FL (CFL). An automated system like this might help relieve the strain on healthcare institutions throughout the world, which have been under a lot of pressure since the COVID-19 pandemic broke out in late 2019. On two benchmark datasets, the author tests the proposed framework's performance under various experimental conditions. On both datasets, promising results were obtained, with improvements of 16% and 11% in overall F1-Scores over the multi-modal model trained in the conventional FL setup on the X-ray and ultrasound datasets, respectively.

Roth et al. [105] in a real-world collaborative scenario, they study the application of FL to create medical imaging classification models. This FL initiative brought together seven clinical institutions from across the world to train a model for breast density categorization based on the Breast Imaging, Reporting, and Data System (BIRADS). They successfully train AI models in federation despite significant variances in datasets from all locations (mammography system, class distribution, and data set size) and without centralizing data. The results reveal that models trained with FL outperform models trained only on an institute's local data by 6.3% on average. Furthermore, when the models' generalizability is examined using the testing data from the other participating sites, they find a 45.8% relative improvement. Sheller et al. [106] used the FL approach for multi-institutional collaborations for the privacy of each institute's data. FL among ten institutions produces models that are 99% as good as centralized data models, and generalizability is assessed using data from institutions outside the federation. The author explores the effects of data distribution among cooperating institutions on model quality and learning patterns, suggesting that improved data access through multi-institutional collaborations can improve model quality more than the faults produced by the collaborative technique. As a case study in evaluating FL against clinical decision support (CDS) on a medical imaging job, the challenge of differentiating healthy brain tissue from tissue infected by cancer cells was used. FL protects data on each device by sharing model changes, such as gradient information, rather than the original data. Model updates, on the other hand, can reveal sensitive information because they are based on original data. The adversaries can realize privacy attacks against FL, therefore, the need of making FL private is a need of the state of the art [107]. In [108], the authors presented a systematic review of incentivized FL along with describing various security implications. The authors argue that the adversarial threat rises with the use of incentives-driven FL. Therefore, an extra privacy layer is needed in FL architecture for privacy concerns. For instance, Qayyum et al. [109] presented a learning-based approach to make FL robust against label-flipping attacks.

4.3. Blockchain

Blockchain technology has the potential to revolutionize the healthcare sector by addressing issues related to data privacy and security. By using an immutable database and masking user identities through public key transactions, blockchain can improve the interoperability of current health records and provide a secure and reliable way to store and access medical information [110]. Additionally, the permanent storage of data on the blockchain can provide valuable historical information for research and analysis [111]. The incorporation of smart contracts and the integration of blockchain with IoT technology further enhances the potential for automation and efficiency in the healthcare industry [112]. Overall, the combination of blockchain and AI has the potential to provide verified historical data, improved privacy and security, increased interoperability, and ease of automation in the healthcare sector.

Zerka et al. [113] proposed a novel approach called Chained Distributed ML (CDML) which combines sequential distributed learning with a blockchain-based platform to address legal restrictions in multi-centric research. This approach can predict two-year lung cancer survival using open data from NSCLC-Radiomics and showed no statistically significant difference in performance compared to a centralized approach in six different scenarios. The combination of blockchain and distributed learning improves adaptability, trust, and the pace of AI adoption in multi-centric research. Additionally, Zhang et al. [114] proposed a blockchain-based system for protecting sensitive medical records using DP noise in FL as a privacy-preserving mechanism. The system also addresses the challenge of storage by keeping only the Interplanetary File System's hash value of the data in the blockchain and storing the original data locally.

Alzubi et al. [115] proposed a new method for protecting the privacy of electronic medical records using deep learning (DL) and blockchain technology. A CNN model was trained to detect normal and abnormal users, and access to the records was secured by integrating blockchain with cryptography-based federated learning (FL). Ngan et al. [116] proposed PriFL-Chain, which uses differential privacy (DP) in FL settings to train ML models without requiring users to disclose their raw data. The blockchain keeps a public record of user contributions. The system also uses Mobile Edge Computing (MEC) and InterPlanetary File Systems to reduce the load on the master server, save data communication costs, and increase adaptability. This combined strategy of FL, blockchain, InterPlanetary File System, and MEC effectively protects privacy, reduces the cost of training ML models, and enables the use of diverse community-sourced data. Edge computing can also be used to minimize the cost associated with the transmission and processing of data on cloud services [117].

4.4. Hybrid privacy-preserving techniques

With the increasing demand for privacy, researchers have started working on more efficient algorithms to present accurate methods. For this purpose, different preservation techniques are joined together to fulfill the task. Models that are more accurate and secure produce better outcomes while maintaining model privacy.

4.4.1. Hybrid privacy-preserving deep learning (HPPDL)

We divide the privacy-preserving techniques based on privacy techniques as follows: HE-based privacy preserving deep learning (PPDL), Secure MPC-based PPDL, and DP-based PPDL are the four types of PPDL methods.

Jain et al. [118] proposed research to investigate the use of Fully Homomorphic Encryption, a type of privacy-preserving ML technology, to enable Convolutional Neural Network (CNN) inference on encrypted real-world datasets. The computational depth of fully homomorphic encryption is limited. They are also high-resource operations. The datasets used for the experiment are MNIST and the melanoma dataset. TensorFlow ImageDataGenerator was used to enhance random pictures from a downloaded dataset to create image training data. The original photos were scaled down to 128 × 128 RGB images. The pictures might be classified as benign or malignant. The network is a modified version of the well-known LeNet model, in which the activation layer is put after the pooling layer, conventional rectified linear unit (ReLU) is substituted with approximate ReLU, and only the first completely

Table 4
Privacy preservation techniques (PPT) in healthcare literature review.

PPT	Type	Reference	Year	Method	Performance	Applications
HE	Cryptographic	[69]	2021	The linear models are converted into encrypted models using Pillar techniques.	Retaining an imputation efficacy of over 0.99 AUC score using multiple optimizations & approximations.	Gene imputation
		[70]	2021	Created the collective learning protocol, which is a system for exchanging classified time-series data inside an organization	The in-hospital mortality prediction model's the area under the precision–recall curve, the score increased.	Medical Healthcare
SMPC	Cryptographic	[82]	2022	Offers a way for querying genomic datasets in a privacy-protected manner using SMPC.	It is feasible to query a genomic database with 3000000 variations in under 400 ms using only five genomic query predicates.	Diagnosis and treatment
		[84]	2020	To ensure the privacy of Paillier encryption the algorithm is used for encryption purposes shared with the hospital server.	The technique can withstand a variety of known security threats. It is low cost for patients & can anticipate more disease.	EHR
		[85]	2020	The data is exchanged using the HE technique to secure patient data. The SMPC is used between the patient and the hospital.	It ensures the security and dependability of user data, the model is very sluggish as it is not operated on plain data.	Electronic heath system
GC	Cryptographic	[89]	2020	Offers a distributed access control system that guarantees no properties of one organization may be learned by others.	Although the results of the findings are encouraging, more study is required for the experiment on real-world data.	Polices of data sharing
SS	Cryptographic	[92]	2022	A perceptron-based session key and a logistic map-based intermediate keys were proposed. A lossless strict secret sharing.	On three separate session key groups, the computed average cryptographic times were 74.83, 62.1, and 43.1 ms.	EHR
		[93]	2021	Rivest Cipher 6 encryption algorithm is used with the computational secret sharing scheme for the storage of the images.	The number of Pixels Change Rate (NPCR) values is larger than 99.55 percent.	Healthcare
		[94]	2021	Using Non-Subsampled Contourlet Transform & Multiresolution Singular Value Decomposition of a robust system is proposed.	Test on 200 X-ray images of COVID-19 patients show that the suggested method not only has outstanding invisibility, a high level of robustness against diverse attacks.	COVID-19 detection
DP	Non cryptographic & Non hybrid	[96]	2022	Proposed the work in which instead of the data release, a DP-based on model release six ML classifiers suggested privacy.	Higher values may improve accuracy, according to experimental results on the benchmark dataset.	Heart disease prediction
		[97]	2021	Privacy is preserved using the Gaussian DP system & modification automation is ensured by using the sharing of a memory of the neural network weights.	The classification & recognition model had a mean receiver operator characteristic AUC of 0.848 in the private setting and 0.960 in the non-private setting.	Medical Segmentation, Classification.
		[98]	2020	The efficient net-B0 model is used to diagnose COVID-19 from CXR images and DP is implemented on it.	The accuracy obtained is 71% while the original model accuracy is 94.7%. Even though accuracy is compromised but promising results are obtained.	Diagnose COVID-19
		[99]	2019	This algorithm modifies the important information about patients in the data collection.	The proposed method is capable of providing 96.14% accuracy.	Breast cancer detection

(*continued on next page*)

connected layer contains an activation layer. The monotonic softmax function is not required for the inference process. The author used the Cheon-Kim-Kim-Song (CKKS) approach for encryption.

Rouhani et al. [119] introduced DeepSecure, a framework that allows DL to be used in a privacy-preserving context. The author used an approach with convolutional neural networks (CNN) to perform the learning process and used the GC protocol to make it private. DeepSecure allows client–server interaction to perform learning processes on a cloud server utilizing data from the user. They used a moderately genuine adversary model to prove their systems. The GC model has been widely demonstrated to keep client data confidential during the data transfer time. The disadvantage of this technique is that it places a limit on the number of instances that may be handled per round.

Yue et al. [121] proposed a new system for evaluating time-series medical pictures encrypted by a HE technique and presented the HE convolutional-LSTM network (HE-LSTM). To extract discriminative spatial features, many convolutional blocks are built, and LSTM-based sequence analysis layers are a technique for encoding temporal data from encrypted image sequences. A weighted unit and a sequence voting layer are also being developed to combine both spatial and temporal attributes with various weights to increase performance while reducing missed diagnoses. The author experimented on two datasets, the BreaKHis public dataset and a Cervigram dataset, whose results show that the proposed framework can encode both visual and sequential dynamics from the encrypted image sequence. The proposed method achieved 0.94 AUCs for both datasets. Vizit et al. [122] proposed a solution-based HE for the privacy of medical domain sensitive data. The considered encryption scheme, Matrix Operation for Randomization or Encryption (MORE), enables the computations within a neural network model to be directly performed on floating-point data with a relatively small computational overhead. The author first trains a model on encrypted data to estimate the outputs of a whole-body circulation (WBC) hemodynamic model and then provides a solution for classifying encrypted X-ray coronary angiography medical images. The findings highlight the potential of the proposed PPDL methods to outperform existing approaches by providing, within a reasonable amount of time, results equivalent to those achieved by unencrypted models. Zhang et al. [120] proposed a novel privacy-preserving approach for training deep neural networks. For every training iteration, we add decaying Gaussian noise to the gradients. This is in contrast to Google's TensorFlow Privacy's conventional method, which uses the same noise scale throughout the whole training process. The suggested solution, in contrast to existing methods, used a closed-form mathematical expression to calculate the privacy loss. It is simple to calculate and can be useful when users want to determine the best training time. To validate the efficiency of the suggested method, the author presents substantial experimental findings utilizing one real-world medical dataset, i.e., chest radiographs from the CheXpert collection, to see the effectiveness of

Table 4 (*continued*).

PPT	Type	Reference	Year	Method	Performance	Applications
FL	Non cryptographic & non hybrid	[103]	2021	An AI method with a worldwide validation the effort to construct privacy-preserving a CNN-based model for identifying CT abnormalities in COVID-19 patients.	The FL model had the best results on this set, with an AUC of 88.15% (86.38–89.91), the sensitivity of 73.31% (70.44–76.18), and accuracy of 91.93% (89.48–94.38).	COVID-19 detection
		[104]	2021	For an automated diagnosis of COVID-19 used the developing the idea of clustered FL.	On both datasets, promising results are obtained, with comparable outcomes overall F1-Score improvements of 16% and 11% have been reached.	COVID-19 detection
		[105]	2020	To create medical imaging classification models in a real-world collaborative scenario, the author used FL.	When the model is applied to a client's test data, it shows a 6.3% relative improvement. The models' generalizability increased by 45.8% on average.	Breast cancer classification
		[106]	2020	Demonstrates that data-private collaborative learning techniques, especially FL can attain the data maximum learning capacity.	The federated learning across ten institutions results in models that are 99% as good as centralized data models, and assess the generalizability using data from institutions outside the federation.	Brain Tumor
		[102]	2018	They have developed a decentralized iterative cluster Primal-Dual Splitting technique for the solving of the large-scale SVM problem.	Converges faster than the present centralized method. However at the cost of the communication between the agents.	EHR
HPPDL	Hybrid	[118]	2022	The use of Fully HE, a type of PPML technology, to enable CNN inference on encrypted real-world datasets. The CKKS encryption is used was used.	On melanoma, the test accuracy of this simple CNN model was found to be 80%, which is just slightly lower than the usual approach.	Skin cancer diagnosis
		[120]	2021	The author proposed privacy using DP. For each iteration Gaussian noise is added.	Achieved the AUC of 80% which is more than the DP stochastic gradient descent method.	Chest radiography
		[121]	2021	For evaluating time series medical images encrypted by a completely HE technique, the HE-CLSTM convolutional LSTM the network is presented.	Test accuracy increases to 93.71%, However, the time complexity increases. There is a trade-off between accuracy and efficiency.	Computer Aided diagnosis
		[122]	2020	The considered encryption scheme is Matrix Operation for Randomization or Encryption (MORE).	The results are not changed same as encrypted data is just a marginal increase in the computation time.	Medical Images
		[119]	2018	Used GC protocol with CNN to perform the learning process &make it private.	58 folds throughput increases, run time decreases	EHR
HPPFL	Hybrid	[123]	2022	A privacy-preserving FL strategy based on the cryptographic primitive of homomorphic re-encryption.	The scheme did model training while training model and data privacy.	Medical diagnosis
		[124]	2022	To private the local model Rényi DP training using a Gaussian noise mechanism.	In terms of private model training, the author found the DenseNet121 model is superior toResNet50 for all variables studied. The DenseNet with DP is more robust towards attack.	Medical images
		[125]	2020	Two domain adaption algorithms in this FL formulation, taking into account for the systemic heterogeneity in fMRI distributions from different sites.	The findings show that using multisite data without sharing, data can improve neuroimage analysis performance and lead to the discovery of credible disease-related biomarkers.	Medical Diagnosis
		[126]	2019	Proposed an FL environment, and explore the possibility of using DP methods to secure patient data.	Sharing fewer variables yields lower overall DP costs and consequently improved model performance by fixing the per-parameter DP costs.	Brain Tumor Segmentation

(*continued on next page*)

the proposed method. The proposed method achieved an AUC of 80%, which is more than the differential privacy stochastic gradient descent method, which is 60% so the result showed that the proposed model secures more privacy.

4.4.2. Hybrid privacy-preserving FL (HPPFL)

Despite having great potential for privacy preservation concerns, FL itself faces threats of privacy compromise. The FL can face attacks like membership inference attacks, and adversarial attacks that result in the need for a privacy-aware FL [128]. For example, based on local model updates from an IoMT device, an attacker can use membership inference attacks to predict the existence of a data sample in the local training dataset [129], such as blood type, disease name, gender information, and any other private information. This encourages researchers to innovate privacy-enhancing FL designs for applications, especially related to sensitive domains like healthcare.

Lee et al. [101] proposed a privacy-preserving framework for patient similarity learning across institutions in a federated context. The proposed method can find similar patients from one hospital to another without sharing any information about each patient. A federated patient hashing architecture was created, as well as a unique technique for learning context-specific hash codes to represent patients across institutions. The generated hash codes of corresponding patients may be used to effectively compute the similarities between patients. In k nearest neighbor with $k = 3$, attained a mean area under the curves of 0.9154 and 0.8012 with balanced and unbalanced data, respectively, while privacy was assured using HE. The dataset used is MIMIC-III. Li et al. [126] proposed an FL environment, and explored the possibility of using DP methods to secure patient data. On the BraTS dataset, develop and test FL algorithms for brain tumor segmentation. The findings of the experiments demonstrate that model performance and privacy protection costs are mutually exclusive.

Ku et al. [123] proposed a privacy-preserving FL strategy based on the cryptographic primitive of homomorphic re-encryption, which can both secure and train user data via homomorphic re-encryption. The IoT device encrypts and uploads user data, the fog node collects user data, and the server completes data aggregation and re-encryption in this approach. Furthermore, this technique can complete model

Table 4 (*continued*).

PPT	Type	Reference	Year	Method	Performance	Applications
		[127]	2019	Two layers of privacy protection in this framework are used. First, the model training process,it does not transmit or share raw data Second, it makes use of a DP strategy to protect the model from future privacy breaches.	Despite DP being widely used in federated settings, it can result in a considerable loss in model performance for healthcare applications.	EHR
		[101]	2018	A privacy-preserving framework for patient similarity learning across institutions in a federated context. The HE is used for encryption	In k nearest neighbor with k = 3, attained mean the area under the curves of 0.9154 and 0.8012 with balanced and unbalanced data, respectively	EHR
Bc	Decentralized	[115]	2022	Uses DL and Bc technology to propose a new method for protecting the privacy of patients' electronic medical records.	The method shows more promising results than other existing techniques.	Medical Data
		[116]	2022	PriFL-Chain which uses DP to be applied to FL in order to train ML models	Effectively protect privacy, reduce the cost of training ML models and make use of diverse community-sourced data.	Medical data
		[114]	2021	A blockchain-based system for protecting sensitive medical records (MPBC). Employs DP noise in FL as privacy-preserving in this framework.	Approach has been proven secure through extensive analysis, so medical data can be implemented with confidence.	Medical data
		[113]	2020	A novel distributed learning approach, Chained Distributed ML C-DistriM which combines sequential distributed learning with a Bc-based platform.	The combination of Bc and distributed learning helps to improve openness, trust, and the pace at which AI is adopted in multicentric research.	Medical Diagnosis

Homomorphic Encryption = HE, Secure Multiparty Computation = SMPC, Garbled Circuits = GC, Secret Sharing = SS, Differential Privacy = DP, Federated Learning = FL, Deep Learning = DL, Blockchain = Bc, Hybrid Privacy-Preserving Deep Learning = HPPDL, Hybrid Privacy-Preserving FL = HPPFL.

training while protecting user data and local models, according to the security analysis and experimental findings.

Li et al. [125] used a privacy-preserving method to solve the challenge of multi-site functional magnetic resonance imaging (fMRI) classification on ABIDE datasets. The author focuses on resolving the information privacy challenges faced by recovering private information from model gradients or weights. The author presented a privacy-preserving technique to handle the challenge of multi-site fMRI classification in this paper. The author presented an FL technique to handle the challenge, in which a decentralized iterative optimization algorithm is used and shared local model weights are changed via a randomization mechanism. Two domain adaption algorithms are in this FL formulation, taking into account the systemic heterogeneity in fMRI distributions from different sites. The author looks into a variety of practical elements of FL optimization and compares FL to other training methods. Ziegler et al. [124] by using image reconstruction attacks on local model updates from specific clients, The author showed that both model designs are susceptible to privacy violations. During the final rounds of training, the attack was extremely successful. The author combined Rényi differential privacy with a Gaussian noise mechanism into local model training to reduce the probability of privacy violation. The author analyzes model performance and attack susceptibility for privacy budgets.

Choudhury et al. [127] proposed an FL system that can learn a global model using distributed health data stored locally at several places. Two layers of privacy protection in this framework are used. First, throughout the model training process, it does not transmit or share raw data between sites or with a centralized server. Second, it makes use of a differential privacy strategy to better protect the model from future privacy breaches. One million patients' data is collected from the EHR for two healthcare applications. The Medical Information Mart for Intensive Care (MIMIC III) dataset is used by the author. Ali et al. [44] proposed to use HE in FL settings for textual misinformation detection related to the spread of messages related to COVID-19 on a social networking platform.

Table 4 shows different privacy preservation techniques in the healthcare domain.

4.5. Available tools for PPML

TenSEAL[2] is a library that combines HE with traditional ML frameworks. It takes care of all the difficulties that come with implementing tensor operations on encrypted data. Data encryption scrambles data into "ciphertext", making it difficult for those who do not have the necessary decryption key or password. TenSEAL is based on Microsoft SEAL's implementation of the CKKS. CKKS is a public key encryption technique that generates both a secret and a public key. While the public key can be shared for encryption, the private key must be kept private for decryption.

Clients can use one of the available frontend languages (C++ or Python) to deal with plain or encrypted tensors. The buffer protocol is used to exchange messages in a client–server situation. The context, plain tensors, and tensors are the three main components of the core application programming interface (API) [130].

PySyft[3] is an open-source multi-language toolkit for enabling secure and private ML by wrapping and extending popular DL frameworks like PyTorch in a transparent, lightweight, and user-friendly manner. Its goal is to make privacy-preserving ML techniques as accessible as possible to researchers and data scientists using Python bindings and common tools, as well as to be extensible so that new FL, Multi-Party Computation, or DP methods can be implemented and integrated flexibly and easily [131].

PyGrid[4] is a peer-to-peer platform for federated learning and data science based on the PySyft architecture. Gateways and nodes are the two parts of the architecture. The Gateway component acts as a DNS server, directing requests to the nodes that have the necessary datasets. The nodes are given by the data owners: they are private data clusters that their owners will maintain and monitor. The data does not leave the server of the data owner. The data scientists may then utilize PyGrid to do private statistical analysis on that dataset or even federated learning across several datasets from different institutions [132].

In the year 2020, OpenMined released PyDP,[5] a Python wrapper for Google's Differential Privacy project. The library includes a collection of differentially private algorithms for generating aggregated statistics

[2] https://github.com/OpenMined/TenSEAL
[3] https://github.com/OpenMined/PySyft
[4] https://github.com/OpenMined/PyGrid
[5] https://github.com/OpenMined/PyDP

Table 5
Comparison of privacy preserving tools.

Library	Language	Key Features	Applications	Privacy Preserving Techniques
TenSeal	Python API C++	An open-source library that can be readily incorporated into major ML frameworks for PPML using homomorphic encryption.	Tensors,Images	HE
PySyft	Python	An open source library provide the private and secure DL.	Images, Text	HE, SMPC, DP, FL
PyGrid	Python	A peer to peer network collects data to train AI models.	Text	Data-centric FL, SMPC
PyDP	Python	An open source library containing several differential privacy algorithms.	Text, Images	DP
SyferText	Python	A library use privacy-preserving natural language processing.	Text	SMPC, FL
TensorFlow Federated	Python	Able to computation on decentralized data	Text, Images	FL

over numeric datasets, including private or sensitive data. As a result, PyDP gives you complete control over the privacy and correctness of your Python model [133].

SyferText[6] is a python module that allows for privacy-preserving natural language processing. It uses PySyft to do FL and encrypted computations on text data using SMPC. SyferText is used in two primary scenarios: Secure plaintext preprocessing allows text to be preprocessed on a remote system without jeopardizing data privacy. Deploy a secure pipeline: SyferText will be able to package a whole pipeline made up of preprocessing components and trained PySyft models and safely deliver it to PyGrid [134]. TensorFlow Federated (TFF)[7] is an open-source framework for decentralized data ML and other computations. TFF was created to enable open study and experimentation through FL, an ML technique in which a shared global model is built across multiple clients that store their training data locally [135].

A comparison of these tools is summarized in Table Table 5 for reference.

5. Insights and pitfalls

The methods of preserving privacy are explained in detail, but there are some limitations to achieving privacy in AI-based healthcare systems, as outlined next.

- The applications using big and complicated homomorphic encrypted algorithms have certain limitations. Like today, all the HE-based encryption methods have computational overhead, defined as the ratio of the encrypted version's calculation time to the plain version's computational time.
  This cost significantly increases execution time and makes complicated functions of homomorphic computation unfeasible [136]. To ensure the computation's security, random numbers must be produced in SMPC. The creation of random numbers necessitates computing cost, which might slow down the execution time [13]. Secret sharing necessitates communication and connectivity among all parties, resulting in greater communication costs than plaintext computation. The DP computational cost is also very high, which limits the use of the technique [137].
- While using HE, the main problem faced is the lack of multi-user capability. If a large number of clients want to encrypt their data to ensure privacy and they all belong to the same system, it results in a problem as the algorithm will fail to support multiple parties. The solution to this problem can be that each user will be assigned a separate database, which becomes impossible if the dataset is huge and there are a lot of users [136].
- DP is best for low-sensitivity queries. The downside of maintaining differential privacy is that it usually needs more noise infusion than conventional methods. The major issue in DP is noise handling. As in low-sensitive data noise, the slight change in the record does not affect the result much, but in the case of highly sensitive data, noise handling limits the working of the algorithm [137].
- Currently, there are a handful of libraries available that support one or more encryption schemes. But even if libraries support the same scheme, they are not explainable. The encryption parameters are not in the standardized format. The explainability of the parameters is different in different libraries. This standardization effort contains a high-level overview of the security and secure parameter recommendations, as well as possible applications and design concerns. It is a good start, but developing solutions now forces one to use a certain library. It can be challenging to pick a particular library. The explainable ML model is needed to implement AI in healthcare applications.
- Existing work lacks implementation of hybrid techniques in the healthcare domain. Moreover, the implementation of privacy-preserving techniques results in accuracy compromise. Accuracy is a major concern for the implementation of AI in the healthcare domain.

6. Future research directions

Recent advances in AI have paved the way for the adoption of intelligent algorithms in healthcare systems. While motivations for using privacy-preserving techniques in the healthcare system are elaborated in-depth, this section presents several future research and development directions in this field.

6.1. Developing privacy-aware ML

In the current digital age, preserving the privacy of users is of utmost importance. Therefore, the development of privacy-preserving ML models is required. It can be particularly useful for developing such ML-based applications that require sensitive data, e.g., healthcare, biometrics, etc. Privacy-aware ML models can ensure the safe execution of the system and can eventually help in gaining the trust of end-users. Different techniques can be used to preserve privacy (as described above). However, the literature shows that these techniques are still vulnerable to different attacks. Therefore, the development of privacy-aware ML models is still an open issue that demands further development.

6.2. Developing adversarially robust ML

Ensuring the robustness of the ML/DL models has emerged as a major challenge in recent years. Despite the remarkable performance of these models in different complex tasks, these models have been shown to be vulnerable to carefully crafted adversarial samples. Although different defense strategies have been proposed. It is very important to ensure that these models perform dependably even in the presence of data corruption, distributional shifts, and malicious threats when we employ them in real-world security-critical applications. Despite significant attention from the research community, the development of adversarially robust ML models remains an open-ended problem.

[6] https://github.com/OpenMined/SyferText
[7] https://github.com/tensorflow/federated

6.3. Distributed ML

Multi-node ML methods and systems are referred to as distributed ML. They can enhance performance, boost accuracy, and scale to bigger input data volumes. Distributed ML and edge computing have progressed to the point where they can transform a company. Distributed devices, such as the Internet of Things (IoT) generate a vast amount of data, which might be used to identify hidden patterns and provide other insights about the user's private data. In general, distributed ML enables a cloud/server to collect a combined model(s) from multiple participants, where each participant trains their model locally using their private data rather than transmitting the actual data to the server. FL has appeared as a major advancement in the last few years. However, the literature shows that sensitive information can be inferred by exploiting parameter updates of individual participants [138].

6.4. Tiny ML in healthcare

TinyML offers a lot of promise in healthcare by improving different monitoring and personal health devices. Wearable technologies have a large power budget to continuously sense and transmit an individual's physiological and activity data, which necessitates constant connectivity and privacy protection. TinyML-based, small, pre-trained inference models for signal de-noising, temporal analysis, and classification on the device can thoroughly assess personal data in real-time, avoiding the need for data to stream continuously and reducing the risk of private data disclosure (in case data is shared over the servers for analytics). Low-power, highly accurate, real-time inference algorithms development for wearable devices will be especially important for increasingly complicated, data-rich physiological sensors like ECG, which continue to offer a considerable problem for today's wearable technologies. Several firms are experimenting with TinyML-like frameworks to improve personal health goods like hearing aids. Other personal health and well-being apps, such as vision enhancement or gait tracking, are likely to follow a similar trend. Continuous monitoring and assessment of an individual's well-being and mental health will be possible thanks to multi-sensor data fusion and deep neural network inference on an embedded device, enhancing the ways to treat people with various mental conditions like dementia, depression, and post-traumatic stress disorder (PTSD) [139]. For instance, in [140], the authors leveraged multi-sensor data (i.e., from smartphone sensors and a wearable device) to identify locomotor impairment. In the literature, privacy and security aspects of embedded ML have already been investigated from a human-centric perspective [141]. Specifically, the authors presented a pipeline for developing secure, private, and robust human-centric embedded ML applications such as healthcare. We envision that with development of tiny ML for deployment on such small-size devices will eventually lead to the protection of users' private data.

6.5. Addressing the trade-off of privacy vs. performance

The performance of ML models is tested using different metrics to identify which model is the most effective at discovering correlations and patterns between variables in a dataset based on the input, or training data. The greater a model's ability to generalize to 'unseen' data is, the better predictions and insights it can provide, and hence the more commercial value it can give, especially in healthcare, where the system's performance for diagnosis of the disease is a crucial and life-saving task. Therefore, it is quite challenging to preserve the privacy of the patient while also obtaining a more accurate system. To enhance the performance of AI systems, large volumes of data are required, but this may compromise the privacy rights of those involved [142]. Utilizing various privacy-preserving approaches has an impact on the model's performance. For example, when using the DP technique to protect the model's privacy, noise is added to the training dataset, and the model's accuracy suffers as a result. Consequently, to ensure the system's appropriate working, a proper trade-off between privacy and accuracy must be made. Privacy-preserving techniques have been widely developed in several FL frameworks to ensure data privacy. The privacy-preserving techniques used in such FL frameworks, on the other hand, tend to impair accuracy and efficiency. As a result, while implementing privacy-preserving features to the FL, it is important to strike a compromise between data utility and data privacy [143]. Privacy-preserving ML techniques approaches have an inherent compromise between the model's utility and the privacy provided by the applied technique.

6.6. Towards hybrid techniques

Hybrid methods use a combination of different privacy-preserving techniques and can provide better results in privacy preservation. They can also address privacy concerns in complex problems which might be difficult to solve using simple techniques. Different privacy-preserving strategies have different limits that must be overcome to enhance privacy solutions. Hybrid techniques are better since they combine techniques in several ways to create a more private model.

The literature suggests that the privacy preservation techniques in developing AI-based healthcare systems should be more on hybrid techniques as they are better for the complex scenario and to solve existing issues with the present techniques [144]. For example, FL, distributes computing across clients, however, the cost of communication among the clients and sender is the primary obstacle to FL scalability. This problem is inherited by hybrid techniques, and it is aggravated in FL-SMPC. Combining HE with FL (FL-HE) introduces a new scaling barrier to FL. There is a developing body of work on communication-efficient FL systems that can greatly increase FL scalability and make it suited for large-scale applications, such as those in biomedicine and healthcare [14].

7. Conclusions

Recent advances in AI has sought the attention of healthcare service providers to invest in AI-based solutions that have the potential to solve perennial healthcare problems related to workers' productivity, efficiency, and care outcomes. Healthcare is a highly regulated sector, and it is expected that these intelligent algorithms will take a while before being deployed in clinics to yield actual benefits. Since modern AI algorithms use data to learn to perform complex tasks, protecting privacy and confidentiality is the primary concern when sensitive datasets are shared for developing AI algorithms. This concern is addressed in other sectors by employing privacy preservation techniques which have shown promising results and are found crucial to promoting AI research. Many researchers have been trying to adapt these strategies for healthcare datasets handling. These concerns and strategies shall be assessed for their efficacy for broader AI adoption across all medical specialities. To this end, we present a comprehensive review of privacy-preserving techniques in the healthcare domain. We developed a taxonomy of privacy attacks and explained techniques that can be used to protect against such attacks involving healthcare datasets and AI models. Finally, we discussed various challenges and pitfalls of different privacy-preserving machine learning (PPML) techniques and highlighted numerous open research issues that require further development.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This publication was made possible by NPRP grant # 13S-0206-200273 from the Qatar National Research Fund (a member of the Qatar Foundation). Open Access funding provided by the Qatar National Library. The statements made herein are solely the responsibility of the authors.

References

[1] C. Milana, A. Ashta, Artificial intelligence techniques in finance and financial markets: A survey of the literature, Strateg. Chang. 30 (3) (2021) 189–209.
[2] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, et al., Highly accurate protein structure prediction with AlphaFold, Nature 596 (7873) (2021) 583–589.
[3] F. Urbina, F. Lentzos, C. Invernizzi, S. Ekins, Dual use of artificial-intelligence-powered drug discovery, Nat. Mach. Intell. 4 (3) (2022) 189–191.
[4] D. Lee, S.N. Yoon, Application of artificial intelligence-based technologies in the healthcare industry: Opportunities and challenges, Int. J. Environ. Res. Public Health 18 (1) (2021) 271.
[5] A. Qayyum, J. Qadir, M. Bilal, A. Al-Fuqaha, Secure and robust machine learning for healthcare: A survey, IEEE Rev. Biomed. Eng. 14 (2020) 156–180.
[6] J.L. Hall, D. McGraw, For telehealth to succeed, privacy and security risks must be identified and addressed, Health Aff. 33 (2) (2014) 216–221.
[7] E. Tom, P.A. Keane, M. Blazes, L.R. Pasquale, M.F. Chiang, A.Y. Lee, C.S. Lee, A.A.I.T. Force, Protecting data privacy in the age of AI-enabled ophthalmology, Transl. Vis. Sci. Technol. 9 (2) The Association for Research in Vision and Ophthalmology.
[8] M. Mamdouh, A.I. Awad, H.F. Hamed, A.A. Khalaf, Outlook on security and privacy in IoHT: Key challenges and future vision, in: The International Conference on Artificial Intelligence and Computer Vision, Springer, 2020, pp. 721–730.
[9] R.U. Rasool, H.F. Ahmad, W. Rafique, A. Qayyum, J. Qadir, Security and privacy of internet of medical things: A contemporary review in the age of surveillance, botnets, and adversarial ML, J. Netw. Comput. Appl. (2022) 103332.
[10] K. Abouelmehdi, A. Beni-Hessane, H. Khaloufi, Big healthcare data: Preserving security and privacy, J. Big Data 5 (1) (2018) 1–18.
[11] H.C. Tanuwidjaja, R. Choi, K. Kim, A survey on deep learning techniques for privacy-preserving, in: International Conference on Machine Learning for Cyber Security, Springer, 2019, pp. 29–46.
[12] P.P. Churi, A.V. Pawar, A systematic review on privacy preserving data publishing techniques., J. Eng. Sci. Technol. Rev. 12 (6) (2019).
[13] G.A. Kaissis, M.R. Makowski, D. Rückert, R.F. Braren, Secure, privacy-preserving and federated machine learning in medical imaging, Nat. Mach. Intell. 2 (6) (2020) 305–311.
[14] R. Torkzadehmahani, R. Nasirigerdeh, D.B. Blumenthal, T. Kacprowski, M. List, J. Matschinske, J. Späth, N.K. Wenke, J. Baumbach, Privacy-preserving artificial intelligence techniques in biomedicine, Methods Inf. Med. (2022).
[15] M.D. Hiller, L.F. Seidel, Patient care management systems, medical records, and privacy: A balancing act., Public Health Rep. 97 (4) (1982) 332.
[16] J.R. Maxeiner, Freedom of information and the EU data protection directive, Fed. Comm. LJ 48 (1995) 93.
[17] S.C. Bennett, The right to be forgotten: Reconciling EU and US perspectives, Berkeley J. Int'L L. 30 (2012) 161.
[18] E. Politou, E. Alepis, C. Patsakis, Forgetting personal data and revoking consent under the GDPR: Challenges and proposed solutions, J. Cybersecur. 4 (1) (2018) tyy001.
[19] K.P. Andriole, Security of electronic medical information and patient privacy: What you need to know, J. Am. Coll. Radiol. 11 (12) (2014) 1212–1216.
[20] P.F. Edemekong, P. Annamaraju, M.J. Haydel, Health Insurance Portability and Accountability Act, 2018.
[21] K.M. Manheim, L. Kaplan, Artificial Intelligence: Risks to Privacy and Democracy, 2018.
[22] M.C. Elish, D. Boyd, Situating methods in the magic of big data and AI, Commun. Monogr. 85 (1) (2018) 57–80.
[23] E. Shabunina, G. Pasi, A graph-based approach to ememes identification and tracking in social media streams, Knowl.-Based Syst. 139 (2018) 108–118.
[24] A. Amberkar, P. Awasarmol, G. Deshmukh, P. Dave, Speech recognition using recurrent neural networks, in: 2018 International Conference on Current Trends Towards Converging Technologies, ICCTCT, IEEE, 2018, pp. 1–4.
[25] J. Zeng, C. Li, L.-J. Zhang, A face recognition system based on cloud computing and AI edge for IOT, in: International Conference on Edge Computing, Springer, 2018, pp. 91–98.
[26] L.T. Car, D.A. Dhinagaran, B.M. Kyaw, T. Kowatsch, S. Joty, Y.-L. Theng, R. Atun, et al., Conversational agents in health care: Scoping review and conceptual analysis, J. Med. Internet Res. 22 (8) (2020) e17158.
[27] M. Montebello, AI injected e-learning, 19, 2018, p. 2018, Online VerfÜGbar Unter https://Link.Springer.Com/Content/Pdf/10.1007%2F978-3-319-67928-0.Pdf, Zuletzt GeprÜFt Am.
[28] E.W. Steyerberg, et al., Clinical Prediction Models, Springer, 2019.
[29] C.D. Sestili, W.S. Snavely, N.M. VanHoudnos, Towards security defect prediction with AI, 2018, arXiv preprint arXiv:1808.09897.
[30] A. Rathee, Data breaches in healthcare: A case study, Cybernomics 2 (2) (2020) 25–29.
[31] A.H. Seh, M. Zarour, M. Alenezi, A.K. Sarkar, A. Agrawal, R. Kumar, R. Ahmad Khan, Healthcare data breaches: Insights and implications, in: Healthcare, Vol. 8, (2) Multidisciplinary Digital Publishing Institute, 2020, p. 133.
[32] V. Sailakshmi, Analysis of Cloud Security Controls in AWS, Azure, and Google Cloud, 2021.
[33] S. Nigam, Telehealth and telemedicine: Clinical and regulatory issues, Telehealth Med. Today 1 (1) (2016).
[34] D.D. Koch, et al., Is the HIPAA security rule enough to protect electronic personal health information (PHI) in the cyber age? J. Healthc. Finance 43 (3) (2016).
[35] G. Hempel, D.B. Janosek, D.B. Raziano, Hacking humans: A case study and analysis of vulnerabilities in the advancing medical device landscape, Cyber Secur. Peer Rev. J. 3 (4) (2020) 351–362.
[36] N.C. Shachmurove, W. McCulloch, Health care companies face financial strain from data breaches, Am. Bankruptcy Inst. J. 40 (8) (2021) 20–52.
[37] E.M. Matos, Cybersecurity Readiness: Smaller Healthcare Organizations, US National Capitol Region (Ph.D. thesis), Marymount University, 2021.
[38] H. Journal, June 2022 healthcare data breach report, HIPAA J. (2022) https://www.hipaajournal.com/june-2022-healthcare-data-breach-report/.
[39] G. Bansal, D. Gefen, et al., The impact of personal dispositions on information sensitivity, privacy concern and trust in disclosing health information online, Decis. Support Syst. 49 (2) (2010) 138–150.
[40] C.S. Kruse, B. Smith, H. Vanderlinden, A. Nealand, Security techniques for the electronic health records, J. Med. Syst. 41 (8) (2017) 1–9.
[41] C. Butpheng, K.-H. Yeh, H. Xiong, Security and privacy in IoT-cloud-based e-health systems—A comprehensive review, Symmetry 12 (7) (2020) 1191.
[42] N. Papernot, M. Abadi, U. Erlingsson, I. Goodfellow, K. Talwar, Semi-supervised knowledge transfer for deep learning from private training data, 2016, arXiv preprint arXiv:1610.05755.
[43] M. Yang, L. Lyu, J. Zhao, T. Zhu, K.-Y. Lam, Local differential privacy and its applications: A comprehensive survey, 2020, arXiv preprint arXiv:2008.03686.
[44] H. Ali, R.T. Javed, A. Qayyum, A. AlGhadhban, M. Alazmi, A. Alzamil, K. AlUtaibi, J. Qadir, SPAM-das: Secure and privacy-aware misinformation detection as a service, 2022.
[45] M. Al-Rubaie, J.M. Chang, Privacy-preserving machine learning: Threats and solutions, IEEE Secur. Priv. 17 (2) (2019) 49–58.
[46] K. Rasheed, A. Qayyum, M. Ghaly, A. Al-Fuqaha, A. Razi, J. Qadir, Explainable, trustworthy, and ethical machine learning for healthcare: A survey, Comput. Biol. Med. (2022) 106043.
[47] B.K. Rai, A.K. Srivastava, Security and privacy issues in healthcare information system, Int. J. Emerg. Trends Technol. Comput. Sci. 3 (6) (2014) (ISSN: 2278-6858).
[48] M. Vucetic, A. Uzelac, N. Gligoric, E-health transformation model in Serbia: Design, architecture and developing, in: 2011 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery, IEEE, 2011, pp. 566–573.
[49] M.D.N. Huda, N. Sonehara, S. Yamada, A privacy management architecture for patient-controlled personal health record system, J. Eng. Sci. Technol. 4 (2) (2009) 154–170.
[50] A. Chester, Y.S. Koh, J. Wicker, Q. Sun, J. Lee, Balancing utility and fairness against privacy in medical data, in: 2020 IEEE Symposium Series on Computational Intelligence, SSCI, IEEE, 2020, pp. 1226–1233.
[51] G. Mai, K. Cao, P.C. Yuen, A.K. Jain, On the reconstruction of face images from deep face templates, IEEE Trans. Pattern Anal. Mach. Intell. 41 (5) (2018) 1188–1202.
[52] N. Papernot, P. McDaniel, A. Sinha, M.P. Wellman, Sok: Security and privacy in machine learning, in: 2018 IEEE European Symposium on Security and Privacy (EuroS&P), IEEE, 2018, pp. 399–414.
[53] A. Qayyum, A. Ijaz, M. Usama, W. Iqbal, J. Qadir, Y. Elkhatib, A. Al-Fuqaha, Securing machine learning in the cloud: A systematic review of cloud machine learning security, Front. Big Data 3 (2020) 587139.
[54] R. Shokri, M. Stronati, C. Song, V. Shmatikov, Membership inference attacks against machine learning models, in: 2017 IEEE Symposium on Security and Privacy, SP, IEEE, 2017, pp. 3–18.
[55] C. Song, T. Ristenpart, V. Shmatikov, Machine learning models that remember too much, 2018, arXiv:170907886.
[56] M.A.U. Alam, Person re-identification attack on wearable sensing, 2021, arXiv preprint arXiv:2106.11900.
[57] S.K. Karmaker Santu, V. Bindschadler, C. Zhai, C.A. Gunter, NRF: A naive re-identification framework, in: Proceedings of the 2018 Workshop on Privacy in the Electronic Society, 2018, pp. 121–132.

[58] M. Ye, J. Luo, G. Zheng, C. Xiao, T. Wang, F. Ma, MedAttacker: Exploring black-box adversarial attacks on risk prediction models in healthcare, 2021, arXiv preprint arXiv:2112.06063.

[59] A.I. Newaz, N.I. Haque, A.K. Sikder, M.A. Rahman, A.S. Uluagac, Adversarial attacks to machine learning-based smart healthcare systems, in: GLOBECOM 2020-2020 IEEE Global Communications Conference, IEEE, 2020, pp. 1–6.

[60] A. Rahman, M.S. Hossain, N.A. Alrajeh, F. Alsolami, Adversarial examples—Security threats to COVID-19 deep learning systems in medical IoT devices, IEEE Internet Things J. 8 (12) (2020) 9603–9610.

[61] Z. Zhang, C. Yan, B.A. Malin, Membership inference attacks against synthetic health data, J. Biomed. Inform. 125 (2022) 103977.

[62] U. Gupta, D. Stripelis, P.K. Lam, P. Thompson, J.L. Ambite, G. Ver Steeg, Membership inference attacks on deep regression models for neuroimaging, in: Medical Imaging with Deep Learning, PMLR, 2021, pp. 228–251.

[63] A. Jagannatha, B.P.S. Rawat, H. Yu, Membership inference attack susceptibility of clinical language models, 2021, arXiv preprint arXiv:2104.08305.

[64] G. Liu, C. Wang, K. Peng, H. Huang, Y. Li, W. Cheng, Socinf: Membership inference attacks on social media health data with machine learning, IEEE Trans. Comput. Soc. Syst. 6 (5) (2019) 907–921.

[65] D. Usynin, D. Rueckert, G. Kaissis, Beyond gradients: Exploiting adversarial priors in model inversion attacks, 2022, arXiv preprint arXiv:2203.00481.

[66] S. Al-Kuwari, Privacy-preserving AI in healthcare, in: Multiple Perspectives on Artificial Intelligence in Healthcare, Springer, 2021, pp. 65–77.

[67] R.L. Rivest, Cryptography and machine learning, in: International Conference on the Theory and Application of Cryptology, Springer, 1991, pp. 427–439.

[68] D. Chowdhury, A. Dey, R. Garai, S. Adhikary, A.D. Dwivedi, U. Ghosh, W.S. Alnumay, Decrypt: A 3DES inspired optimised cryptographic algorithm, J. Ambient Intell. Humaniz. Comput. (2022) 1–11.

[69] E. Sarkar, E. Chielle, G. Gürsoy, O. Mazonka, M. Gerstein, M. Maniatakos, Fast and scalable private genotype imputation using machine learning and partially homomorphic encryption, IEEE Access 9 (2021) 93097–93110.

[70] J. Paul, M.S.M.S. Annamalai, W. Ming, A. Al Badawi, B. Veeravalli, K.M.M. Aung, Privacy-preserving collective learning with homomorphic encryption, IEEE Access 9 (2021) 132084–132096.

[71] M.L. Gaid, S.A. Salloum, The International Conference on Artificial Intelligence and Computer Vision, Springer, 2021, pp. 634–642.

[72] M.C. Hastings, Secure Multi-Party Computation in Practice (Ph.D. thesis), University of Pennsylvania, 2021.

[73] E. Lee, S. Minner, An information sharing framework for supply chain networks: What, when, and how to share, in: IFIP International Conference on Advances in Production Management Systems, Springer, 2021, pp. 159–168.

[74] P. Sarosh, S.A. Parah, G.M. Bhat, Utilization of secret sharing technology for secure communication: A state-of-the-art review, Multimedia Tools Appl. 80 (1) (2021) 517–541.

[75] B. Jiang, J. Li, G. Yue, H. Song, Differential privacy for industrial internet of things: Opportunities, applications and challenges, IEEE Internet Things J. (2021).

[76] L.U. Khan, W. Saad, Z. Han, E. Hossain, C.S. Hong, Federated learning for internet of things: Recent advances, taxonomy, and open challenges, IEEE Commun. Surv. Tutor. (2021).

[77] S. Nakamoto, Bitcoin: A peer-to-peer electronic cash system, Decentralized Bus. Rev. (2008) 21260.

[78] W.-J. Lu, Y. Yamada, J. Sakuma, Privacy-preserving genome-wide association studies on cloud environment using fully homomorphic encryption, in: BMC Medical Informatics and Decision Making, vol. 15, (5) Springer, 2015, pp. 1–8.

[79] S. Carpov, T.H. Nguyen, R. Sirdey, G. Constantino, F. Martinelli, Practical privacy-preserving medical diagnosis using homomorphic encryption, in: 2016 IEEE 9th International Conference on Cloud Computing, CLOUD, IEEE, 2016, pp. 593–599.

[80] T. Dugan, X. Zou, A survey of secure multiparty computation protocols for privacy preserving genetic tests, in: 2016 IEEE First International Conference on Connected Health: Applications, Systems and Engineering Technologies, CHASE, IEEE, 2016, pp. 173–182.

[81] R. Cramer, I.B. Damgård, et al., Secure Multiparty Computation, Cambridge University Press, 2015.

[82] M. Akgün, N. Pfeifer, O. Kohlbacher, Efficient privacy-preserving whole genome variant queries, Bioinformatics (2022).

[83] O. Goldreich, S. Micali, A. Wigderson, How to play any mental game, in: Annual ACM Symposium on Theory of Computing.

[84] D. Li, X. Liao, T. Xiang, J. Wu, J. Le, Privacy-preserving self-serviced medical diagnosis scheme based on secure multi-party computation, Comput. Secur. 90 (2020) 101701.

[85] A.V. Kumar, M.S. Sujith, K.T. Sai, G. Rajesh, D.J.S. Yashwanth, Secure multi-party computation enabled E-healthcare system with homomorphic encryption, in: IOP Conference Series: Materials Science and Engineering, vol. 981, (2) IOP Publishing, 2020, 022079.

[86] M. Marwan, A. Kartit, H. Ouahmane, Applying secure multi-party computation to improve collaboration in healthcare cloud, in: 2016 Third International Conference on Systems of Collaboration, SysCo, IEEE, 2016, pp. 1–6.

[87] P. Jangde, D.K. Mishra, A secure multiparty computation solution to healthcare frauds and abuses, in: 2011 Second International Conference on Intelligent Systems, Modelling and Simulation, IEEE, 2011, pp. 139–142.

[88] A.C.-C. Yao, How to generate and exchange secrets, Sfcs 1986, in: 27th Annual Symposium on Foundations of Computer Science, IEEE, 1986, pp. 162–167.

[89] J. Sancho, J. García, Á. Alesanco, Distributed access control for cross-organizational healthcare data sharing scenarios, in: European Medical and Biological Engineering Conference, Springer, 2020, pp. 407–413.

[90] Y. Gong, C. Zhang, Y. Hu, Y. Fang, Privacy-preserving genome-aware remote health monitoring, in: 2016 IEEE Global Communications Conference, GLOBECOM, IEEE, 2016, pp. 1–6.

[91] M. Barni, P. Failla, R. Lazzeretti, A.-R. Sadeghi, T. Schneider, Privacy-preserving ECG classification with branching programs and neural networks, IEEE Trans. Inf. Forensics Secur. 6 (2) (2011) 452–468.

[92] J. Dey, A. Bhowmik, S. Karforma, Neural perceptron & strict lossless secret sharing oriented cryptographic science: Fostering patients' security in the "new normal" COVID-19 E-health, Multimedia Tools Appl. (2022) 1–32.

[93] P. Sarosh, S.A. Parah, G.M. Bhat, A.A. Heidari, K. Muhammad, Secret sharing-based personal health records management for the internet of health things, Sustainable Cities Soc. 74 (2021) 103129.

[94] A. Anand, A.K. Singh, Secret sharing based watermarking for copy-protection and ownership control of medical image, in: 2021 12th International Conference on Computing Communication and Networking Technologies, ICCCNT, IEEE, 2021, pp. 01–07.

[95] C. Dwork, Differential privacy: A survey of results, in: International Conference on Theory and Applications of Models of Computation, Springer, 2008, pp. 1–19.

[96] S. Sangeetha, G. Sudha Sadasivam, A. Srikanth, Differentially private model release for healthcare applications, Int. J. Comput. Appl. (2022) 1–6.

[97] A. Ziller, D. Usynin, R. Braren, M. Makowski, D. Rueckert, G. Kaissis, Medical imaging deep learning with differential privacy, Sci. Rep. 11 (1) (2021) 1–8.

[98] Z. Müftüoğlu, M.A. Kizrak, T. Yildlnm, Differential privacy practice on diagnosis of COVID-19 radiology imaging using EfficientNet, in: 2020 International Conference on INnovations in Intelligent SysTems and Applications, INISTA, IEEE, 2020, pp. 1–6.

[99] A. Vadavalli, R. Subhashini, An improved differential privacy-preserving truth discovery approach in healthcare, in: 2019 IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference, IEMCON, IEEE, 2019, pp. 1031–1037.

[100] Q. Yang, Y. Liu, Y. Cheng, Y. Kang, T. Chen, H. Yu, Federated learning, Synth. Lect. Artif. Intell. Mach. Learn. 13 (3) (2019) 1–207.

[101] J. Lee, J. Sun, F. Wang, S. Wang, C.-H. Jun, X. Jiang, Privacy-preserving patient similarity learning in a federated environment: Development and analysis, JMIR Med. Inform. 6 (2) (2018) e7744.

[102] T.S. Brisimi, R. Chen, T. Mela, A. Olshevsky, I.C. Paschalidis, W. Shi, Federated learning of predictive models from federated electronic health records, Int. J. Med. Inf. 112 (2018) 59–67.

[103] Q. Dou, T.Y. So, M. Jiang, Q. Liu, V. Vardhanabhuti, G. Kaissis, Z. Li, W. Si, H.H. Lee, K. Yu, et al., Federated deep learning for detecting COVID-19 lung abnormalities in CT: A privacy-preserving multinational validation study, NPJ Digit. Med. 4 (1) (2021) 1–11.

[104] A. Qayyum, K. Ahmad, M.A. Ahsan, A. Al-Fuqaha, J. Qadir, Collaborative federated learning for healthcare: Multi-modal COVID-19 diagnosis at the edge, IEEE Open J. Comput. Soc. 3 (2022) 172–184.

[105] H.R. Roth, K. Chang, P. Singh, N. Neumark, W. Li, V. Gupta, S. Gupta, L. Qu, A. Ihsani, B.C. Bizzo, et al., Federated learning for breast density classification: A real-world implementation, in: Domain Adaptation and Representation Transfer, and Distributed and Collaborative Learning, Springer, 2020, pp. 181–191.

[106] M.J. Sheller, B. Edwards, G.A. Reina, J. Martin, S. Pati, A. Kotrotsou, M. Milchenko, W. Xu, D. Marcus, R.R. Colen, et al., Federated learning in medicine: Facilitating multi-institutional collaborations without sharing patient data, Sci. Rep. 10 (1) (2020) 1–12.

[107] A. Blanco-Justicia, J. Domingo-Ferrer, S. Martínez, D. Sánchez, A. Flanagan, K.E. Tan, Achieving security and privacy in federated learning systems: Survey, research challenges and future directions, Eng. Appl. Artif. Intell. 106 (2021) 104468.

[108] A. Ali, I. Ilahi, A. Qayyum, I. Mohammed, A. Al-Fuqaha, J. Qadir, Incentive-driven federated learning and associated security challenges: A systematic review, 2021.

[109] A. Qayyum, M.U. Janjua, J. Qadir, Making federated learning robust to adversarial attacks by learning data and model association, Comput. Secur. 121 (2022) 102827.

[110] A.L. Duca, C. Bacciu, A. Marchetti, How distributed ledgers can transform healthcare applications, Blockchain Eng. (2016) 25.

[111] T.-T. Kuo, H.-E. Kim, L. Ohno-Machado, Blockchain distributed ledger technologies for biomedical and health care applications, J. Am. Med. Inform. Assoc. 24 (6) (2017) 1211–1220.

[112] A. Panarello, N. Tapas, G. Merlino, F. Longo, A. Puliafito, Blockchain and IoT integration: A systematic survey, Sensors 18 (8) (2018) 2575.

[113] F. Zerka, V. Urovi, A. Vaidyanathan, S. Barakat, R.T. Leijenaar, S. Walsh, H. Gabrani-Juma, B. Miraglio, H.C. Woodruff, M. Dumontier, et al., Blockchain for privacy preserving and trustworthy distributed machine learning in multicentric medical imaging (C-DistriM), IEEE Access 8 (2020) 183939–183951.

[114] H. Zhang, G. Li, Y. Zhang, K. Gai, M. Qiu, Blockchain-based privacy-preserving medical data sharing scheme using federated learning, in: International Conference on Knowledge Science, Engineering and Management, Springer, 2021, pp. 634–646.

[115] J.A. Alzubi, O.A. Alzubi, A. Singh, M. Ramachandran, Cloud-IIoT-based electronic health record privacy-preserving by CNN and blockchain-enabled federated learning, IEEE Trans. Ind. Inform. 19 (1) (2022) 1080–1087.

[116] L. Ngan Van, A. Hoang Tuan, D. Phan The, T.-K. Vo, V.-H. Pham, A privacy-preserving approach for building learning models in smart healthcare using blockchain and federated learning, in: Proceedings of the 11th International Symposium on Information and Communication Technology, 2022, pp. 435–441.

[117] S. Adhikary, A. Ghosh, E-BMI: A gait based smart remote BMI monitoring framework implementing edge computing and incremental machine learning, Smart Health 24 (2022) 100277.

[118] N. Jain, K. Nandakumar, N. Ratha, S. Pankanti, U. Kumar, PPDL-privacy preserving deep learning using homomorphic encryption, in: 5th Joint International Conference on Data Science & Management of Data, 9th ACM IKDD CODS and 27th COMAD, 2022, pp. 318–319.

[119] B.D. Rouhani, M.S. Riazi, F. Koushanfar, Deepsecure: Scalable provably-secure deep learning, in: Proceedings of the 55th Annual Design Automation Conference, 2018, pp. 1–6.

[120] X. Zhang, J. Ding, M. Wu, S.T. Wong, H. Van Nguyen, M. Pan, Adaptive privacy preserving deep learning algorithms for medical data, in: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2021, pp. 1169–1178.

[121] Z. Yue, S. Ding, L. Zhao, Y. Zhang, Z. Cao, M. Tanveer, A. Jolfaei, X. Zheng, Privacy-preserving time-series medical images analysis using a hybrid deep learning framework, ACM Trans. Internet Technol. 21 (3) (2021) 1–21.

[122] A. Vizitiu, C.I. Nită, A. Puiu, C. Suciu, L.M. Itu, Applying deep neural networks over homomorphic encrypted medical data, Comput. Math. Methods Med. 2020 (2020).

[123] H. Ku, W. Susilo, Y. Zhang, W. Liu, M. Zhang, Privacy-preserving federated learning in medical diagnosis with homomorphic re-encryption, Comput. Stand. Interfaces 80 (2022) 103583.

[124] J. Ziegler, B. Pfitzner, H. Schulz, A. Saalbach, B. Arnrich, Defending against reconstruction attacks through differentially private federated learning for classification of heterogeneous chest X-ray data, 2022, arXiv preprint arXiv:2205.03168.

[125] X. Li, Y. Gu, N. Dvornek, L.H. Staib, P. Ventola, J.S. Duncan, Multi-site fMRI analysis using privacy-preserving federated learning and domain adaptation: ABIDE results, Med. Image Anal. 65 (2020) 101765.

[126] W. Li, F. Milletarì, D. Xu, N. Rieke, J. Hancox, W. Zhu, M. Baust, Y. Cheng, S. Ourselin, M.J. Cardoso, et al., Privacy-preserving federated brain tumour segmentation, in: International Workshop on Machine Learning in Medical Imaging, Springer, 2019, pp. 133–141.

[127] O. Choudhury, A. Gkoulalas-Divanis, T. Salonidis, I. Sylla, Y. Park, G. Hsu, A. Das, Differential privacy-enabled federated learning for sensitive health data, 2019, arXiv preprint arXiv:1910.02578.

[128] V. Mothukuri, R.M. Parizi, S. Pouriyeh, Y. Huang, A. Dehghantanha, G. Srivastava, A survey on security and privacy of federated learning, Future Gener. Comput. Syst. 115 (2021) 619–640.

[129] K.Y. He, D. Ge, M.M. He, Big data analytics for genomic medicine, Int. J. Mol. Sci. 18 (2) (2017) 412.

[130] A. Benaissa, B. Retiat, B. Cebere, A.E. Belfedhal, Tenseal: A library for encrypted tensor operations using homomorphic encryption, 2021, arXiv preprint arXiv:2104.03152.

[131] A. Ziller, A. Trask, A. Lopardo, B. Szymkow, B. Wagner, E. Bluemke, J.-M. Nounahon, J. Passerat-Palmbach, K. Prakash, N. Rose, et al., Pysyft: A library for easy federated learning, in: Federated Learning Systems, Springer, 2021, pp. 111–139.

[132] V. Turina, Z. Zhang, F. Esposito, I. Matta, Combining split and federated architectures for efficiency and privacy in deep learning, in: Proceedings of the 16th International Conference on Emerging Networking EXperiments and Technologies, 2020, pp. 562–563.

[133] G.M. Garrido, J. Near, A. Muhammad, W. He, R. Matzutt, F. Matthes, Do I get the privacy I need? Benchmarking utility in differential privacy libraries, 2021, arXiv preprint arXiv:2109.10789.

[134] M. Samir, M. Azab, M.R. Rizk, N. Sadek, PYGRID: A software development and assessment framework for grid-aware software defined networking, Int. J. Netw. Manage. 28 (5) (2018) e2033.

[135] Z. Sun, P. Kairouz, A.T. Suresh, H.B. McMahan, Can you really backdoor federated learning? 2019, arXiv preprint arXiv:1911.07963.

[136] M. Ogburn, C. Turner, P. Dahal, Homomorphic encryption, Procedia Comput. Sci. 20 (2013) 502–509.

[137] M. Saifuzzaman, T.N. Ananna, M.J.M. Chowdhury, M.S. Ferdous, F. Chowdhury, A systematic literature review on wearable health data publishing under differential privacy, 2021, arXiv preprint arXiv:2109.07334.

[138] F. Boenisch, A. Dziedzic, R. Schuster, A.S. Shamsabadi, I. Shumailov, N. Papernot, When the curious abandon honesty: Federated learning is not private, 2021, arXiv preprint arXiv:2112.02918.

[139] I. Fedorov, M. Stamenovic, C. Jensen, L.-C. Yang, A. Mandell, Y. Gan, M. Mattina, P.N. Whatmough, Tinylstms: Efficient neural speech enhancement for hearing aids, 2020, arXiv preprint arXiv:2005.11138.

[140] S. Adhikary, A. Ghosh, Dynamic time warping approach for optimized locomotor impairment detection using biomedical signal processing, Biomed. Signal Process. Control 72 (2022) 103321.

[141] M.A. Butt, A. Qayyum, H. Ali, A. Al-Fuqaha, J. Qadir, Towards secure private and trustworthy human-centric embedded machine learning: An emotion-aware facial recognition case study, Comput. Secur. 125 (2023) 103058.

[142] L. Song, P. Mittal, Systematic evaluation of privacy risks of machine learning models, $\{USENIX\}$ Security 21, in: 30th $\{USENIX\}$ Security Symposium, 2021.

[143] L. Sankar, S.R. Rajagopalan, H.V. Poor, Utility-privacy tradeoffs in databases: An information-theoretic approach, IEEE Trans. Inf. Forensics Secur. 8 (6) (2013) 838–852.

[144] N. Truong, K. Sun, S. Wang, F. Guitton, Y. Guo, Privacy preservation in federated learning: An insightful survey from the GDPR perspective, Comput. Secur. 110 (2021) 102402.