IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received June 28, 2020, accepted July 12, 2020, date of publication July 22, 2020, date of current version August 5, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3011099

Secondary Use of Electronic Health Record: Opportunities and Challenges

SHAHID MUNIR SHAH AND RIZWAN AHMED KHAN
Faculty of Information Technology, Barrett Hodgson University, Karachi 74900, Pakistan
Corresponding author: Rizwan Ahmed Khan (rizwan.khan@bhu.edu.pk)

ABSTRACT In the present technological era, healthcare providers generate huge amounts of clinical data on a daily basis. Generated clinical data is stored digitally in the form of Electronic Health Record (EHR) as a central data repository of hospitals. Data contained in EHR is not only used for the patients' primary care but also for various secondary purposes such as clinical research, automated disease surveillance and clinical audits for quality enhancement. Using EHR data for secondary purposes without consent or in some cases even with consent creates privacy issues. Secondly, EHR data is also made accessible to various stakeholders including different government agencies at various geographical sites through wired or wireless networks. Sharing of EHR across multiple agencies makes it vulnerable to cyber attacks and also makes it difficult to implement strict privacy laws as in some cases data is shared with organization that is governed by specific regional law. Privacy of individuals could be severely affected when their sensitive private information contained in EHR is leaked or exposed to the public. Data leaks can cause financial losses or an individual may encounter social boycott if his / her medical condition is exposed in public. To protect patients personal data from such threats, there exists different privacy regulations such as General Data Protection Regulation (GDPR), Health Insurance Portability and Accountability Act (HIPAA) and My Health Record (MHR). However, continually evolving state-of-the-art techniques in Machine Learning (ML), Data Analytics (DA) and hacking are making it even more difficult to completely protect an individual's/patient's privacy. In this article, we have systematically examined various secondary uses of EHR with the aim to highlight how these secondary uses affect patients' privacy. Secondly, we have critically analyzed GDPR & HIPAA regulations and highlighted their possible areas of improvement, considering escalating use of technology and different secondary uses of EHR.

INDEX TERMS Electronic health records (EHR), ethical concerns, general data protection regulation (GDPR), privacy, secondary uses of EHR.

I. INTRODUCTION

Clinical data is generated in the form of ongoing patient diagnostic services. These services usually take place in hospitals, clinics or laboratories through different clinical trials (via medical imaging or doctors' prescriptions) or through wireless body area network using wearable sensors [1]. All of these sources produce a huge amount of clinical data world wide and its volume is experiencing an exponential growth. It is estimated that clinical data will swell up to 2314 Exabytes by 2020 from a figure of 153 Exabytes in 2013 with an annual growth rate of 48% [2].

The associate editor coordinating the review of this manuscript and approving it for publication was Yu-Huei Cheng.

In most of the countries (especially developing countries), data generated during routine clinical practices is stored manually in the form of paper based medical records. This procedure is adopted by the medical practitioners because of ease of handling, lack of understanding or for the purpose of treating more patients in less time. However, this method of storing patient's medical information is not useful for the patients and does not guarantee accurate and timely deliverance of healthcare services. Some other problems associated with manual recording of clinical/medical data are:

1) Paper based medical records can easily be altered or can be lost and may cause severe consequences.
2) Physicians/clinicians can prescribe wrong medications (due to alteration of paper based medical records) or cannot advise right medications during

follow up visits without properly knowing past medical records of patients.
3) It is not practical for a person to carry a huge bunch of paper based past medical records during follow up visits or to describe complete medical history to a physician/clinician in case of change of physician or hospital.
4) Reviewing and analyzing paper based records poses cumbersome task for new physicians or medical staff when patients change their physicians or hospital.

To avoid all of the above described difficulties, an automated/electronic online patient information system through which patients' complete medical record is made available to healthcare professionals is required. Such electronic system also serves the purpose of storing patients data for longer time without any alterations and makes it accessible through different locations to support quick decision making processes [3].

Healthcare organizations are now adopting techniques to digitize medical records to overcome challenges (described above) faced by them or by patients while using paper based medical records [4]. With the new technique, patients' clinical data is now stored as Electronic Health Records (EHR). EHR are the patients' computerized health records that contain patients' complete information along with their medical history in a format (refer Figure 1) that can be easily shared among different health care providers or can be accessed by them through different linked locations when required [5].

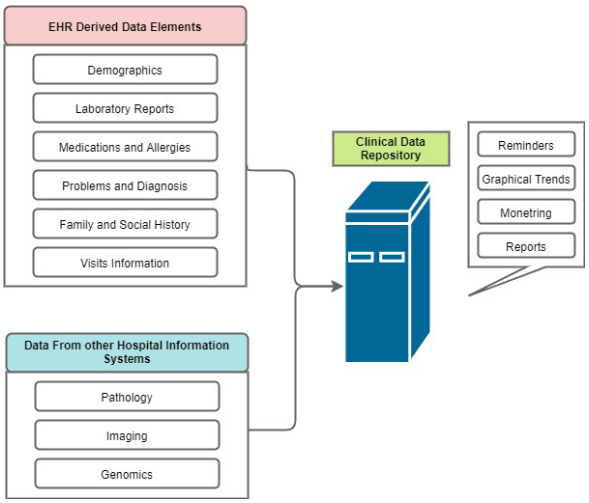

FIGURE 1. EHR as a clinical data repository.

Adoption of EHR provide range of benefits over the traditional paper based medical record systems. For Example:

1) EHR are capable of storing structured, coded and electronic patient data all together to form a complete history of patient's health [6].
2) Electronic data saved as EHR makes a Decision Support System (DSS) for monitoring health outputs to improve health care quality [7], where DSS is a tool, usually software based tool, that supports decision making by providing automated analysis of data [8].
3) EHR system acts as a central database of information for patient documentation and billing, maintaining quality, and supporting patient related sensitive decisions [9].
4) Data saved in EHRs can be accessed through multiple locations simultaneously and also can be shared with different partner organizations conveniently. Thus, making data accessible to the concerned physicians across multiple sites to better provide healthcare services.
5) EHR reduces the probability of errors related to medical data analysis as it stores complete medical records and thus lowers overall healthcare cost [10].

With all the above mentioned benefits of using EHR, certain risk factors are also associated with it. The most important issue is the data security and patient's privacy. In case, if EHR data is leaked/theft or stolen from the database it can be misused (by altering dosage of drugs or treatment procedure etc.) and may cause severe complications or even lead to the patient's death [11]. It is therefore utmost important to protect patient's information in the central database from unauthorized wrong hands. Patient's information may also be stolen while it is in transmission to the other linked services over the network or when it is stored on distributed servers of cloud.

Information contained in EHR is also used for different secondary purposes (other than patient personal care) such as clinical research, health promotions, clinical audit and clinical governance, national screening and preventive campaigns, audits against national standards, national statistics, planning future services, and resource allocations etc., [12] (refer Section V for discussion on secondary uses of EHR). For all such uses, patients may not be willing to share their information as often patients share their private health data for their personal care and not for the other/secondary uses. Using patient sensitive information for different secondary purposes without their consent can seriously affect their privacy.

To safeguard patient privacy or personal data, there exist privacy standards in different regions of the world such as General Data Protection Regulation (GDPR) in Europe [13], [14], Health Insurance Portability and Accountability Act (HIPAA) in the United States (US) [15] and My Health Record (MHR) in Australia [16], [17]. These standards provide legislation to protect personal data but with fast paced advancement in Data Analytics (DA) and Artificial Intelligence (AI) [18], [19] poses new challenges for such standards.

Our contributions in this article are following:

1) In this study we have described various secondary uses of EHR with the aim to highlight how these secondary uses affect patients' privacy, refer Section V for discussion on secondary uses of EHR.

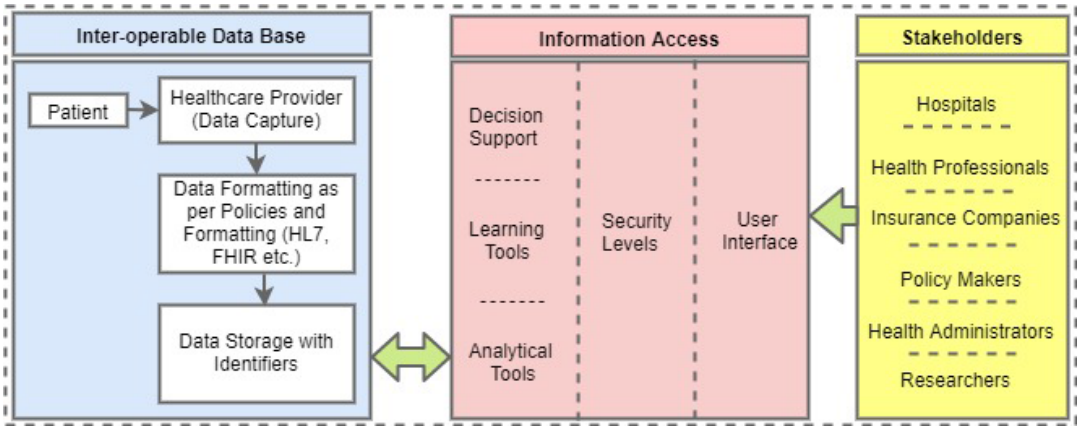

FIGURE 2. Conceptual overview of EHR system.

2) In this article we have discussed various issues associated with secondary use of EHR (refer Section VI). Referred section also elaborates on security and privacy issues of using EHR data (Section VI-D).
3) This article has systematically analyzed GDPR and HIPAA regulation and enlisted their challenges for ensuring data privacy in this era where usage of EHR data (that contains sensitive personal information) has grown exponentially, refer Section VII for discussion on this issue.

Our contributions in this article are oriented toward understanding ethical concerns when dealing with personal data in the era of AI. Research domain of our contributions (described above) needs more collaborative efforts by the research community working in the domain of medicine, computing and law to achieve better insight. Ethical issues arising due to fast proliferation of AI-assisted technologies [20] will raise various serious concerns, specially related to privacy of individuals. Due to the complex nature of this interdisciplinary research domain, it is hard to find literature on the topic and thus, our article is novel as it systematically analyzes uses of sensitive EHR data which, if violated, creates many privacy and ethical concerns.

The rest of the paper is structured as follows: Section II describes EHR along with their different standards. Section III describes information sources of EHR. Section IV presents an overview of various Deep Learning (DL) approaches for EHR data analysis. Section V describes use of EHR in various secondary purposes. Section VI presents challenges of using EHR for secondary purposes. Section VII describes systematic analysis of popular data protection regulations i.e. GDPR and HIPAA in the context of patients privacy and data security with respect to secondary uses of EHR. Finally, in the Section VIII conclusions are presented.

II. ELECTRONIC HEALTH RECORDS (EHR): DATA SHARING

EHR is a clinical data repository containing basic patient information such as a patient's personal profile, their complete family history, laboratory reports, physicians and other medical staff notes etc. Along with this primary information, EHR also contains data form the other hospital information systems such as Imaging data from Radiology department, patients genomics data from Genetic department or Endoscopic and Colonoscopic data from Gastroenterology department etc. Figure 1 illustrates the most important data elements included in the EHR.

EHR also provides functionality of generating reminders for routine screenings and disease reporting, generating graphical trends against various parameters such as blood pressure monitoring, heart beat monitoring, blood glucose level monitoring etc. The same is also shown in Figure 1. Such reporting is highly beneficial for patients health and safety especially when patients are in critical condition and their strict monitoring is required.

Conceptually EHR system can be divided into two basic parts [21]. Creation part and the access part (refer Figure 2). Creation part is based on the interaction of patients with the healthcare providers. This part explains, how the data from the patient is captured, how it is formatted according to the policies and standard and finally, how the formatted data is stored in an interoperable database. Access part is based on the access of the data stored in EHR by the different authorized users or organization. This part explains how confidential information from EHR can be securely accessed by the authorized users via user friendly interfaces.

A. EHR STANDARDS

For the effective use of data contained in EHR, it must be shared through different linked locations such as clinics, hospitals, radiology departments, pharmacies, laboratories and patient homes [22] (refer Figure 3).

Shared data at multiple locations ensures patients solitary care by identifying their basic needs in terms of care, safety, timeliness, and effective monitoring. It also helps medical staff (physician, nurses etc.) to take the right actions based on patient conditions. The data usefulness can further be increased if the data contained in EHR is linked with different

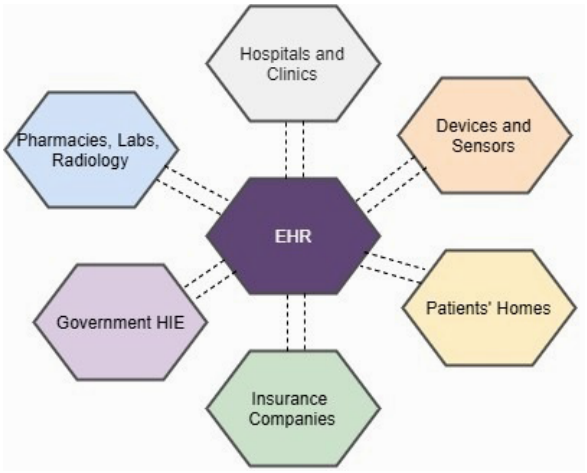

FIGURE 3. EHR data sharing.

clinical decision support systems (CDSS). CDSS refers to an automated medical data analysis tool that suggests next steps for treatment and generates alerts by predicting future conditions/trends by analyzing provided data [23]. By this way, the physicians can take sensitive decisions quickly and effectively [24].

However, without any industry standard for information exchange, it is usually difficult to share and exchange EHR data across multiple sites. The same difficulty was faced by the healthcare organizations to communicate EHR data with each other and with different DSSs when there was no industry standard available for health information exchange. It was the main reason behind the slow adoption of the EHR system in healthcare organizations even if their adoption was highly beneficial for them [25].

1) HEALTH LEVEL SEVEN (HL7) STANDARD

The Health Level Seven (HL7) organization was established in the US in March 1987 to develop consistent common standards for Hospital Information System (HIS) [26]. Afterwards this organization defined HL7-Clinical Document Architecture (HL7 CDA) as EHR messaging standard for easy integration, interchange, sharing and retrieval of information across different clinical information systems. The HL7 standard allows different healthcare organizations to share and exchange patient information via encoded data exchange. It provides a common syntax of information for different clinical information systems to share information (contained in EHR) conveniently [9].

The HL7 CDA Framework 1.0 release, became an American National Standards Institute (ANSI) approved HL7 standard in November 2000 [27]. After the release of the first version, version 2 and version 3 releases were also made available with some new standards and modifications [28]. HL7 CDA is a markup for specifying composition and semantics of data ingredients of EHR such as a discharge report, admission summary, progress and procedure reports and to exchange them with various stakeholders. It is an absolute object document that may hold clinical data in various formats such as text, image, sound, or other multimedia content. Extensible Markup Language (XML) is used to encode the HL7 CDA clinical documents, which then can be exchanged in form of HL7 messages or using other transport solutions.

An HL7 CDA message consists of a header and a body. Header contains information regarding patient, source (provider) and the authentication of the message. On the other hand, the body of the message includes organized clinical reports i.e. lab, radiology, Magnetic Resonance Imaging (MRI), Computed Tomography (CT) scan, ultrasound etc.

2) FAST HEALTHCARE INTEROPERABILITY RESOURCES (FHIR)

In order to improve inter interoperability and exchange of information, HL7 released different versions from time to time. In 1988, HL7 version 2 was released to enhance and streamline information exchange mechanisms/procedures, that can be used by different departments across hospitals [29]. However, different limitations were exposed in this version such as a difficult implementation process, having a number of optional segments and above all lack of proper representation that is capable enough to identify techniques for exchanging messages and interfaces [30]. To overcome the shortcomings of version 2, version 3 was developed in the year 1995. Although HL7 version 3 resolved much of the problems of previous versions, it could not resolve the incompatibility issue raised because of a variety of sub versions [31]. In order to further improve HL7 standards, another novel interoperability standard i.e. Fast Healthcare Interoperability Resources (FHIR) was initiated in the year 2011 [32] by HL7 organization. FHIR standards are very simple to adapt, possess scalability and are robust in nature. These standards have potential capabilities of supporting work flows in small devices like mobile phones [33].

III. ELECTRONIC HEALTH RECORD INFORMATION SOURCES

Adoption of EHR is beneficial both for patients, physicians and healthcare providers. It improves overall healthcare quality, omits paperwork, reduces medical errors and increases work efficiency as well as reduces overall healthcare cost [34].

Beside patients personal care, EHR data is also used for different secondary purposes (refer sectionV for secondary uses of EHR). However, EHR data has not been utilized to its full potential for secondary uses and one of the reasons for under utilization of EHR data is non uniformity in its data components. Non uniformity in data elements exists because of the fact that during daily clinical practices, EHR data is often recorded in free text and unstructured format. Therefore, EHR contains structured and unstructured sets of information. Figure 4 elaborates more on the structured and unstructured data components of EHR. As shown in

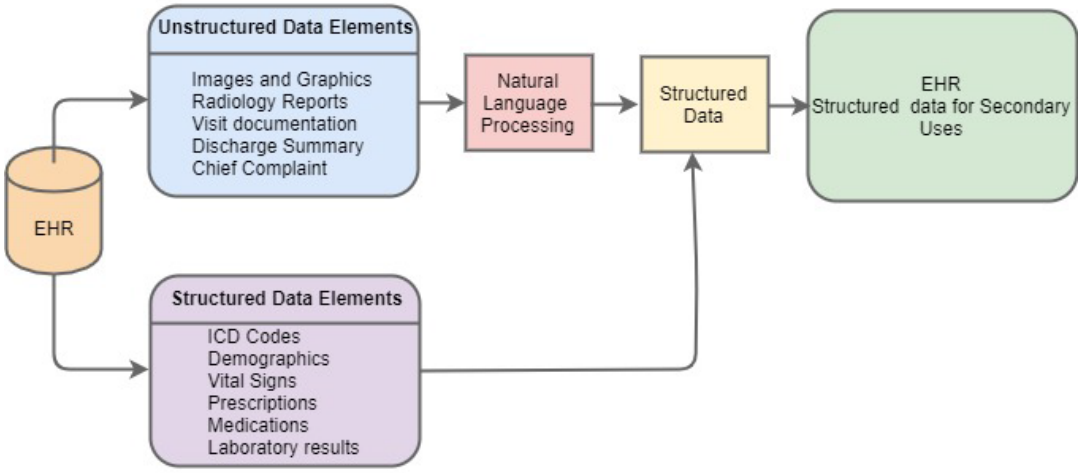

FIGURE 4. Unstructured and structured data elements of EHR.

Figure 4, structured data includes laboratory results, vital signs, prescriptions, medications and International Classification of Diseases (ICD) codes whereas, unstructured data includes narrative information (free text) such as images and graphics, radiology reports, visit notes, discharge summary, chief complaint etc.

Figure 4 depicts that the major portion of EHR data consisted of unstructured elements. Such data elements are not represented in any standard coding scheme such as ICD codes, therefore, their retrieval, reporting and aggregation is not easy like structured data using commonly available database tools [34]. It is therefore required to convert unstructured data elements into structured data in order to make it equally useful for secondary uses.

One of the methods to convert unstructured data into structured data is manually reviewing EHR by the experts using text charts or data abstraction methods [35]. However, these methods are time consuming and not reliable to capture all the structured information. Furthermore, it is beyond the capacity of human to review clinical data of EHR in large volume. Natural language processing (NLP) methods have shown their usefulness for extraction of structured clinical data from unstructured data elements [36], [37].

Mostly, NLP uses statistical (probabilistic) Machine Learning (ML) models to derive language data from large volumes of free text data. These models use text data to identify common patterns and associations in the data. NLP based ML models give meanings to words and phrases in text and convert unstructured data elements of EHR into structured codes. In short, NLP captures unstructured data elements of EHR, analyzes the data elements with respect to their grammatical structures, obtains meanings from grammatical structures, and finally summarizes information to make it useful for further analysis.

IV. DEEP LEARNING FOR EHR DATA ANALYSIS

As discussed above, techniques under the umbrella of NLP are used to extract information from unstructured clinical notes. Such techniques utilize different sequence labeling algorithms such as Conditional Random Fields (CRFs), Hidden Markov Models (HMMs), Artificial Neural Networks (ANN) etc., [38] to label relevant information to be extracted. Sequence labeling is a robust technique that has been used for automatic recognition of various tasks e.g. speech recognition [39], network intrusion detection [40], mental illness detection using social media content [41]. Data contained in EHR (in the form of clinical notes) is often noisy, incomplete and inconsistent as well as it contains grammatical errors, misspelled words etc. (refer Section VI for more detail), thus effectiveness of sequence labeling is limited for extraction of relevant information from EHR data [42]. To sum up, It is usually a difficult task for the traditional sequence labeling algorithms to extract relevant information from EHR data [43] and traditional algorithms also face challenges in analyzing EHR data due as such algorithms are not suitable for dealing with huge columns of data [44].

To overcome challenges faced by traditional NLP approaches in extracting information from EHR, recently, Deep Learning (DL approaches have gained popularity [45]. Popularity of DL models is due to:

1) their capability of analyzing multiple data types,
2) their ability to extract optimal features and learn representation from data automatically [46].
3) their robustness against high complex functions, and
4) their performance gets better on large datasets and it further improves as data grows in size [47].

Various DL architectures and frameworks have been used for EHR analysis. For example, deep patient [48] a unsupervised deep feature learning architecture, deepr [49]

a convolution Neural Network (CNN) based architecture, Med2Vec [50] a Multi Layer Perceptron (MLP) based architecture, Doctor AI [51] a Recurrent Neural Networks (RNN) based architecture etc. are few of the popular Deep Neural Networks (DNN) architectures employed for EHR data analysis and processing. Even though, various DL architectures have been used, the most popular DNN architecture used for EHR data analysis is RNN [52], [53]. This is due to the fact that RNN are capable of dealing with the temporal nature of EHR data in an effective manner [54]. In short, with the technology advancements, at one side EHR have become more informative sources than ever before but on the other hand, DL approaches have opened new avenues to extract and use information contained in EHR for different real world applications.

DL models have shown promising results in EHR data analysis and its processing [55], [56] but with new advancements in medical testing, nature of EHR data (its unstructured elements) is getting complex. With time EHR will contain such pieces of data/information that was never used before. Therefore, for the effective management of advanced and vast data contained in EHR, more efficient tools are required. Researchers need to develop such tools that can efficiently manage EHR data and can convert it into knowledge that benefits society.

V. SECONDARY USES OF EHR

One of the contributions of this study, as described above, is systematic analysis of various secondary uses of EHR data with the aim to highlight how these secondary uses affect patient's privacy. This section discusses different popular secondary uses of EHR data. Section VI elaborates on challenges associated with the secondary use of EHR data while Section VI-D focuses on privacy and security challenges of EHR data.

A. CLINICAL RESEARCH

The basic purpose of clinical research is to use EHR for design and execution of clinical trials for new medicines [57].

Health related issues are directly proportional to the population. Since the population of the world (especially third world countries) has increased at a fast pace in the past few years. This abrupt increase in the population of the world has posed many challenges for healthcare professionals. For example, healthcare organizations, hospitals, laboratories are facing shortage of trained medical staff to tackle healthcare needs of large population and because of insufficient healthcare facilities, new types of diseases are grown in people or existing diseases exhibit more complicated behaviors. Therefore, there exists a need to discover new drugs with better results as well as new techniques and robust strategies to fight against new grown diseases or existing complicated diseases structures i.e. Covid-19 [58]. All such activities require clinical research to be conducted. Thus, clinical research holds a pivotal role in tackling some of the hard pressed medical issues.

Some of the other areas where clinical research is required are:

1) Prediction of diseases based on patients present data [59].
2) Study of drug behaviors with different diseases or different patients i.e. study on antibiotics [60].
3) Developing vaccines for the prevention of diseases before their attack [61].

Other than the areas mentioned above, there exists multiple domains (refer Table 1) where clinical research is essential to overcome the existing problems of the medical world and to ensure high quality of healthcare delivery to the patients.

TABLE 1. Different domains of clinical research.

Domains of clinical research	Examples
Hypothesis Generation	To understand the people response on introducing new drugs.
Epidemiology	To find causes of disease in community of people.
Drug utilization	To figure out the use of medicines and to determine the frequency.
Patient Recruitment	To raise the awareness of clinical trials.
Health Technology Assessments	Evaluating large number of research publications on a topic of interest and generating highly consolidated information for policy makers and health care providers.
Comparative Effectiveness	Obtaining real world evidences from the analysis of real world data generated through routine clinical practices to help decision makers for making effective decisions.
Pragmatic Trails	Observing patients' treatment and their outcomes in real world situations to provide on ground real information to the decision makers so that they can make effective decisions for the enhancement of quality of care.

In the domain of clinical research, EHR is an essential part because it is a basic information source and a possible way of exchanging clinical information with different stakeholders. Based on this exchange of information, various health statistics are developed and decisions are made. For example, based on data collected word wide, the World Health Organization (WHO) publishes various reports time to time for public awareness and for the authorities knowledge to understand the current trends and future needs related to particular diseases [62], [63].

Table 2 lists possible information sources available in EHR that can help in successfully carrying out clinical research in different domains.

Table 2 shows that EHR contains enough information to carry out clinical research in different domains. Successful utilization of this information for research purposes requires development of new and emerging research infrastructures capable of exchanging information based on latest published standards. However, when data is shared across different healthcare organizations, it raises different security and privacy concerns. These concerns are discussed in Section VI-D.

TABLE 2. Different information sources available in EHR that can help in carrying out clinical Research.

Data Sources	Explanation	Possible areas of Clinical Research
Demographics	It includes patient's basic information such as name, age, gender, date of birth, address, contact, allergies, past medical history and diagnosis	Data analysis, community based research, age related research, disease surveillance, and all other epidemiological human population studies.
Daily Habits (Risky Behaviors)	Using tobacco, alcohol, and other sedative drugs.	Cancer research, chronic drug usage implications on health, mental illness, psychological disorders.
Facts and Monitoring	Weight, height, blood pressure, blood sugar, heart beat.	Hypertension research, Body Mass Index (BMI) based research, diabetic research, early childhood growth studies and cognitive outcomes.
Laboratory Data	Complete Blood Count, Prothrombin Time, Basic Metabolic Panel, Comprehensive Metabolic Panel, Lipid Profile, Liver Functioning Test, Thyroid Stimulating Hormone, Hemoglobin A1C, Urinalysis, Microbiological Culture with antibiotic resistance tests and others	To investigate the origin of disease, study of communicable and non communicable diseases as well as blood disorders.
Various Encounters Data	Human population or, hyperlipidemia, diabetes, anxiety and obesity, allergies, reflux esophagitis, respiratory problems, depressive disorder, asthma, nail fungus, urinary tract kidney failure, migraine	Research of all non communicable diseases.
Special tests & Procedures	Appendectomy, Electrocardiogram, Biopsies, Angiographies, Therapies	Special investigative tests for the advance research on non communicable diseases identification and control.
Imaging	Magnetic Resonance Imaging, Ultra Sound, Computerized Axial Tomography, Positron Emission Tomography etc.	Cancer Research, identification and monitoring of Congenital anomalies diseases in unborn babies, bone fractures and tumors, can be used to monitor response of tumors to chemotherapy or radiations.

B. PUBLIC HEALTH SURVEILLANCE

Another secondary use of EHR is Public Health Surveillance (PHS). PHS is a process of collecting, analyzing and interpreting data related to a specific disease for administering and assessing public health on the whole [64]. PHS particularly investigates those diseases, which harm or may tend to harm a large population and grow in communities like epidemic or pandemic diseases. Its main functions include collection of facts about a particular disease, risk factors of its spread and interpretation and analysis of the collected facts for controlling the disease to prevent the public from its severe effects.

One of the examples of PHS is the surveillance of the Dengue outbreak in Pakistan that has been reported in [65]. Dengue is a viral disease which causes high fever in patients and spreads in people because of the bite of a particular Aedes aegypti mosquito. Recently, it has affected around 40% of world's population. Pakistan is one of the most affected country from it. There are several other examples of PHS world wide like reported in [66], [67].

As discussed above, PHS is a common practice world wide but mostly in third world countries it is performed using manual procedures [68], [69]. In these methods, data about disease for surveillance purposes is collected using traditional methods. For example, physicians prescription records are gathered either from patients or from hospitals and clinics or through public surveys [70], similarly data from other departments of the hospitals (laboratories, radiology departments, emergency departments etc) is also collected manually by visiting the logs of these departments' databases. The collected data is then cross communicated between public health staff and health protecting agencies via telephonic and fax communication networks. Collected data is stored on papers manually and manual procedures are used to analyze the stored data [71]. This method of PHS is time consuming, requires large manpower and needs huge efforts to record, store and analyze the data. It is also not a reliable method as there are chances of errors due to manual handling of data. Inaccurate and uncertain outcomes are possible based on the collection and inspection of manually collected and stored data [72]. Such traditional methods are not suitable for the confirmation of certain diseases, understanding its severity, its transmission risks, and the spread of other linked diseases.

With more effective ways, surveillance of diseases can be performed by actively monitoring patients' EHR. As EHR is rich in variety of data, the summary generated by analyzed data is provided to public health agencies for prevention and control of diseases. Health surveillance by EHR provides the glance of the health status of the community, which promotes the quality of healthcare. It tracks the key diseases, with more effective ways than manual procedures. Use of EHR provides the opportunity to automate the PHS. It is an effective way of preventing outbreaks by discovering utmost danger cases irrespective of merely reacting to outbreaks [73]. Figure 5 elaborates the effectiveness of EHR based automated surveillance against the traditional manual surveillance systems.

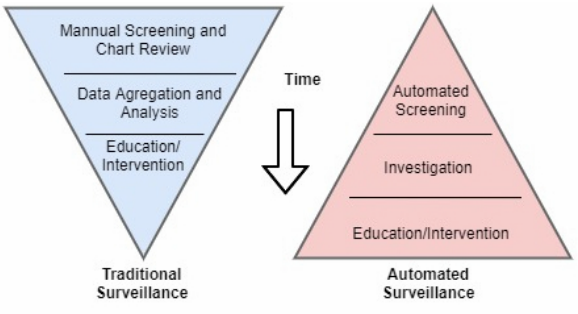

FIGURE 5. Traditional v/s automated surveillance.

During traditional surveillance, most of the time is utilized on manual screenings and reviewing charts and less time is

saved for the actual intervention. On the other hand automated PHS assisted by EHR data is time efficient in analyzing data. The same is shown in Figure 5.

Use of EHR for public health surveillance has proved to be effective in developed countries such as the United Kingdom (UK), United States (US), France, Norway, Canada and Australia [74]. In these countries, local health departments have diverted their manual surveillance system towards EHR based electronic surveillance system. This practice has advanced the functionality of PHS [75]. Developing nations have also initiated adoption of EHRs for PHS to robustly analyze data and take actions, if required [76]. Thus, it is an important need of the present day automated surveillance systems to use data from EHR. For example, Integrated Disease Surveillance and Response System (IDSRS) requires data to be obtained from patient medical records [77], [78].

During the ongoing COVID-19 outbreak, the importance of digital data recording systems (like EHR) has been clearly revealed and demonstrated to the whole world [79]. EHR has been proved to be an efficient tool for detecting, monitoring and managing needs of a health systems [80]. Leveraging on artificial intelligence based tracking systems different countries were able to track movement of Corona positive individuals, thus tracking down any further local transmissions of the virus [80], [81]. It would not be possible, without extensive utilization of EHR, to place in practice a system of effective public health surveillance specially to predict, monitor and manage pandemic like COVID-19 [58], [82].

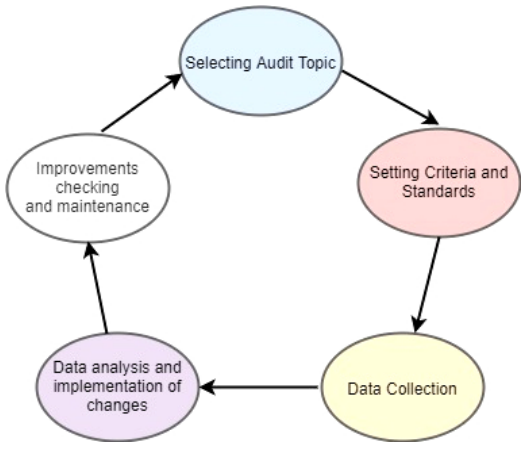

FIGURE 6. Clinical audit as a cyclic process.

C. CLINICAL AUDIT AND QUALITY ASSURANCE

The aim of clinical audit is to enhance patient care via rigorous analysis of care provided against benchmark standards [83]. Clinical audit is a systematic way of settling standards, analyzing data based on standards, performing actions to meet settled standards and executing proper monitoring to sustain the standards. Clinical audit is a cyclic process (refer Figure 6) that contains different stages to be followed for the achievement of best practices in clinical practices.

Standards settled for the clinical audits require not only to be obeyed by the medical staff (doctors, nurses, midwives, therapists etc.) but also by the healthcare organizations like hospitals, clinics, nursing homes, ambulatory surgical centers, autonomous laboratories, radiology units, collection units etc. Clinical audit focuses on broadly accepted methods to improve overall healthcare quality. For example, organizational development, information management and statistics evaluation are the key functions of clinical audits.

The role of EHR is important in clinical audits as it provides detailed and accurate information to the auditors. Using EHR for clinical audits give convenience to the auditors to perform the clinical audits as compared to use of traditional clinical data for audit [84]. Figure 6 shows that the data collection and data analysis are the important parts of the clinical audit. In order to perform quality clinical audits, the clinical data must be easily available as well as the available data must be reliable to perform clinical audit. EHR conveniently provides data to the auditors from multiple access points to perform clinical audits to better provide the quality of the care to the patients.

VI. CHALLENGES ASSOCIATED WITH SECONDARY USES OF EHR

Primarily, EHR data is collected for patient's individual care and administrative billing purposes. Using this data for different secondary purposes (as elaborated in Section V) is always challenging [6], [85]. It is because priorities and settings of primary and secondary uses are different. The quality of data collected for the primary purpose cannot be the same as the quality of data collected for secondary uses. For example, data collected for clinical research needs much more care and attention during collection than the data gathered during routine clinical practices in the form of EHR. The quality of collected clinical data is a serious concern of the researchers. Due to this reason, with respect to reuse of clinical data, the authors of [86] suggested that the data must be used for its primary purpose only.

Following are different factors of concern that affect the quality of clinical data collected through EHR.

A. CORRECTNESS

Correctness refers to the accuracy of the collected data that is directly linked with its initial documentation (how the data was collected, recorded and stored). EHR data is collected through routine clinical practices during which the clinicians priority is to collect the patients data according to their own point of interest and according to different administrative needs but not according to their various secondary uses (refer Section V). The chances of errors are obvious in this case. According to the study presented in [87], data accuracy collected through EHR ranges between 44% to 100%.

Errors in EHR may lead to different outcomes, if their data is used for various secondary purposes. Errors include:

1) Inaccurate predictions by clinical researchers
2) Degradation in health standards and statistics as data analyzed was error prone.
3) False health surveillance results that may lead to unforeseen medical emergency.

Improvement in accuracy of EHR is essential to make it equally beneficial for primary as well as secondary uses.

B. INCOMPLETENESS

Another factor that affects quality of clinical data is related to the completeness of the EHR. Usually, EHR do not contain complete patient history. It is because patients do not always trust a single healthcare organization and may visit several such organizations to get a sense of satisfaction. Study conducted in [88] showed that out of 1.1 million adult patients, 31% visited two or more hospitals, whereas, one percent patients visited five or more hospitals for acute care during a five year period of their study.

Patients also miss follow up visits suggested by the physicians or sometimes due to the perfunctory of the concerned medical staff (who records patient related data), incomplete records are stored in EHR. A Study was conducted at Columbia University on 3068 pancreatic cancer patients out of which only 48% patients had complete pathology records, while the rest had incomplete records about the disease [89].

EHR data is also considered incomplete for secondary uses because of the data ''locked-up'' condition. Locked-up condition means records have details regarding patients but it is not present in the coded portion of the record or in other words data present in EHR is structured and unstructured (already described in Section III). Structured data is in the format that can be easily processed by the computers. On the other hand unstructured data mostly requires NLP (for hand written prescriptions) technique to be applied to make it structured (detail is provided in Section III) and processable by computers [90], [91].

C. INCONSISTENCY

EHR data is handled by various individuals and at different locations. Multiple persons are involved in entering, storing and processing the data, therefore, data contains several definitions. Most of the data is present without mentioning proper units as units are often remembered by the medical staff and they can understand language written by each other. On the other hand for the non concerned person (who wants to use the data for secondary purposes), it may be highly difficult to interpret the data without specified units. Involvement of different individuals in preparation and processing of EHR leads to an inconsistent form of data. It means, data present in EHR is not uniform. In such non uniform data, it is often difficult to relate assessments of different practitioners (because the assessment of different clinicians is often different). Secondly, data inconsistency also arises due to the fact that the data is collected with different tools at different locations, which may be time varying (data coding regulations and system abilities may change with time) [85]. Inconsistent data may lead to erroneous data analysis and wrong results. Therefore, such inconsistent data is not useful for secondary use.

D. SECURITY AND PRIVACY CHALLENGES

EHR based clinical data provides many advantages over manual paper based medical records. It is cost effective, improves overall healthcare quality and above all can be easily accessed through different linked locations. All such advantages motivate health providing agencies and medical practitioners to adopt an EHR based system. However, adoption of EHR and its data processing introduces several privacy and security issues. Especially, when this data is used for secondary purposes (refer Section V for discussion on secondary uses of EHR). In the next subsections, security and privacy challenges related to secondary uses of EHR have been separately discussed.

1) SECURITY CHALLENGES

EHR data is the most vulnerable data to the Cyber Threats. The prime reason for criminals to target healthcare data is to get financial gain. Criminals sell valuable data taken from EHR to the ''darkweb'' [92] (darkweb refers to the content on the web that is not indexed by search engines and thus remains hidden from the general public) and achieve high financial gain. For the criminals, EHR data is more informative than credit cards because it contains various fixed identifiers and important financial information that is extremely valuable in black markets. Fixed identifiers of EHR data can not be reset like the ones in credit cards. Such identifiers in EHR are the best information sources for the criminals to get easy access to the patient's bank accounts for getting loans or to capture their passports and other important documents (property, insurance etc.) [93]. There are several cases, which shows that highly sensitive information of patients was easily stolen by simply stealing EHR data. For example, a recent article published a story about theft of EHR data (20,000 records) from North Carolina-based Catawba Valley Medical Center. Stolen data contained patients' names, dates of birth, medical data, health insurance information and social security numbers [94].

Since, the data in EHR contains more detailed information than the other sources, therefore, in case of cyber attacks (Ransomware, Distributed Denial of Service (DDoS) etc.), a big population can be affected at once that can lead much beyond the financial losses [95]. For example, in the United States (US), 4.5 millions patients' affected by losing their data form a popular group of hospitals through hackers attack [96]. Similarly, 80 millions people were affected because their healthcare data was lost from a health insurance company in US [97].

Other than personal and financial information, EHR data also contains patient's highly confidential data in the form of physicians' personal notes, neuroimaging data [98], [99], X-rays, ultrasounds as well as lab reports. This data may

include lab results of HIV and other sexually transmitted diseases [100], mental disorders [101], personality disorders [102], contagious diseases as well as doctors sensitive comments about patient mental illness or personality disorders etc. All such data is stored in the hospital's local database (each hospital may have its own local electronic database), which is connected across other hospitals' or health providers' databases via wired or wireless connections for sharing purposes. Transfer of such confidential data over the internet creates several security risks. It provides a chance to hackers and other harmful attackers to access the data and use it for their own purpose [103]. In case of patients monitored at home, the data from patients is collected through a distributed network of sensors. Securing such data is another big challenge because there are greater chances of spying and skimming [104].

With the passage of time healthcare technologies are extending and new technologies are being introduced to provide instant help to the patients and to enhance healthcare quality. For example, different smart devices monitor health (with the general purpose devices or wearable sensors) and prescribe medications as well as provide telemedicine technology for delivering remote care [105]. Patients now can easily access healthcare facilities by integrating their mobile phones with telemedicine and telehealth services using simple mobile applications [106]. As the technology in healthcare is continually evolving, its inter connectivity is also evolving. With the help of interconnected networks, patients' information is made broadly available to the relevant organizations and staff to provide quality healthcare. Exchange of patient information over the large inter connected network is beneficial in many ways but has increased existing security risks.

With the increased used of smart healthcare services, e health solutions [107], [108], and digital record systems, EHR data generates sheer volume of data (Alone in US 48% growing annually [109]), therefore, most of the data of EHR database is stored on cloud services [110].

Cloud storage of EHR is beneficial in many ways like it provides cost effective storage, easy access as well as processing and updating of information is achieved with improved effectiveness and efficiency but on the other hand opens new doors for the threats and breaches [111]. It is because highly sensitive and confidential patients' information contained in EHR is stored on a third party server where the owner does not have any direct access [112]. Vulnerabilities also increase because of the fact that during cloud services, a large amount of EHR data runs on a wide network of integrated remote servers and is accessed by multiple authorized users as a single echo system from different distributed locations [113].

Cyber security is a technology that safeguards computer networks and information contained in them from different cyber attacks [114]. In case of healthcare data, cyber security technology needs to be robust and strong as the healthcare sector presents a lucrative avenue to cyber criminals to attack and get hold of very sensitive data to gain large financial benefits.

There have been many efforts reported in literature to protect information contained in EHR while it is accessible to different stakeholders through network. Different cryptographic, non cryptographic and hybrid access control models have been developed to securely access EHR data [115], [116], but with the advancements in technology, securing data in EHR is more challenging as it was ever before. Therefore, there is a high need of securing data in EHR over the network and on cloud servers [117].

Especially securing EHR data on the cloud needs more attention. According to the recently published research in literature [111] the existing privacy and security-protecting mechanisms are not enough to ensure foolproof security in the e-health cloud. Even though, researchers have introduced a few very advanced encryption methods such as Attribute Based Encryption (ABE) [118], Key Policy Attribute Based Encryption (KP-ABE) [119] and Cipher text policy Attribute Based Encryption (CP-ABE) [120] but these are not of much help because the data hosted on clouds is not only vulnerable to external hackers' attacks but also to the internal attacks from the authorized people (database administrators and key managers). The above mentioned advanced access control method cannot provide support when the key managers are attackers.

In order to overcome such deficiencies of cloud based storage, recently blockchain chain technology has been introduced in the healthcare sector [121], [122]. Although blockchain technology provides a range of benefits over the cloud computing technology, it does not provide fine-grained control of access over EHR. In blockchain technology, only the patient's private key can decrypt the encrypted EHR [123].

It is worth mentioning here to explain that the security of healthcare data is not only today's concern rather it was the concern before the emergence of the EHR [104]. Data security was well studied before the EHR came into existence (paper based patients records were needed to be safeguarded within the premises of hospitals and not on large scales) but with the adoption of EHR multiple gateways opened for accessing patients' information remotely. Furthermore, The patient's EHR contains more detailed information all together in a single source as compared to the previous paper based medical records, which were distributed among different departments of hospitals. With the adoption of EHR it is now easy for the criminals to attack millions of people at a time and to steal their valuable information (because EHR are interconnected with numerous networks. In the case of paper based records it was not possible to steal millions of patients records at a time).

In short, adoption of EHR has not only provided the range of benefits but also introduced potential risks of cyber attacks. Healthcare organizations spend more on increasing their integration but do not spend much on their system protection.

In order to gain patients' trust and to give them satisfaction regarding their data safety, the healthcare providers have to think about developing robust practical standards and solutions with particular healthcare/ EHR needs.

2) PRIVACY CHALLENGES

Privacy is defined as ''right to be left alone'' or to keep away from public domain [124]. The United Nations General Assembly (UNGA) declared privacy as a fundamental human right in its universal declaration of human rights. However, in this digital era the term privacy has become subjective and is interpreted and implemented differently by each state or country [125]. Such ambiguities are sometimes exploited for different reasons, for example EHR data is used to gain financial benefits [126] or for different secondary purposes, refer Section V for discussion on secondary uses of EHR.

As mentioned above, EHR data contains several security risks especially when the information contained in them is shared with different stakeholders over the interconnected networks. Other than security issues, there are certain privacy concerns linked with exchange and sharing of EHR data. These privacy concerns are usually raised due to the fact that when the patients data (which was recorded for the purpose of patient individual care) is being shared or linked without consent or knowledge of a particular individual. Usually consent of an individual is necessary for sharing of data but ambiguity arises when different healthcare organizations have different perspectives on the question of ''who owns the data?''. Does data belong to the patient, his/her physician, health insurance organization, healthcare organization, social security agency or is it jointly owned by all [104], [127]?

Breach of data can happen due to various reasons, refer Section VI-D1, which has many ethical repercussions. For example, disclosing a patient's sensitive private information such as sexually transmitted diseases or mental illness in the public domain can negatively impact an individual's reputation. In extreme cases such individuals can face social boycott as people start avoiding an individual if they know that he/she has sexually transmitted diseases like HIV, chlamydia etc., [128]. Secondly, a person's status in the society is seriously affected if his/her mental illness is disclosed to the public [129]. Another dimension to this issue is financial impact on an individual's life as medical insurance companies usually calculate premium/cost of insurance based on medical history and life events. In such cases insurance companies can increase their premium [130], [131].

The privacy of clinical data has been subject to a lot of research and it has been difficult to determine how much of the data belongs to the patient and how much of it may belong to healthcare organizations and whether the consent of the owner of data is needed, in case the data is to be used for the research purpose [127], [132], [133]. Privacy of patients can be affected when his/her data is used for clinical research or secondary use, refer Section V for discussion on secondary uses of EHR. For example, a blood sample given by a patient is stored in a laboratory and after carrying out requested analysis the same sample is analyzed again for the purpose of clinical research. Even though the sample is returned back to the laboratory without any damage, still it violated data privacy because by this way the patient's control over his/her data was lost [127], [132].

In research conducted by Bovenberg and Almeida [134] referred to a case of patients versus Myriad Genetics, a molecular diagnostic company. The case was about four US cancer patients who wanted to have full access to their genomic data. Myriad claimed that patients were provided with all the information that was necessary to be included in their reports and additional data was not part of the medical record set. Patients, however, claimed that the additional data was acquired from their lab samples, hence they have the right over data and only they should decide what happens to their data.

In order to protect sensitive data many patients try to conceal their sensitive information. It is because of the lack of confidence in the system's security retaining their data. It also shows mistrust of patients' on medical staff (doctors, nurses and the others) because patients think that they might disclose their confidential information to the public that may create embarrassment for them in society [135]. Some events have happened in the past because of which patients have become more sensitive in disclosing their private information. For example, in 2013, one of the medical technicians of a US hospital was found guilty in selling patients medical information [136]. Similarly, a hospital in the US informed its 34000 patients that their medical information has been lost from their agent [137]. Due to all such incidents, patients don't feel confident in disclosing their information even to the physicians. Hiding facts and information from the physicians and the medical staff can lead to treatment failure. Thus, such challenges may have severe consequences for patients, healthcare providers and even for the governments.

It is highly recommended from policy makers, leaders and related authorities to discuss privacy and security concerns of EHR data (database storage policies or its sharing policies and paradigms) and formulate policies to address these concerns. There are some existing policies, which need to be revised or reformulated according to the present day era, an era of data analytics, big data and artificial intelligence.

VII. POPULAR DATA PROTECTION REGULATIONS AND THEIR CHALLENGES

A. GENERAL DATA PROTECTION REGULATION (GDPR)

In order to protect patients' personal sensitive data from different security threats and privacy violations, in some regions of the world, data protection regulations have been enforced by the authorities. The most popular data protection regulations are General Data Protection Regulation (GDPR) [13], Health Insurance Portability and Accountability Act (HIPAA) [138] In this study we have critically

analyzed these regulations in terms of how these protect patient privacy and enforce data security.

After years of discussions, drafting, negotiations and efforts, in April 2016 GDPR was passed by European Union. On 25 May 2018, the European Parliament and Council of the European Union both with their combined efforts enforced the GDPR 2016/679 [139]. Since then, professionals, citizens and authorities across Europe and beyond are strictly bound to the legal regimes imposed by GDPR. It is an exhaustive document of legislation that addresses challenges of data protection of personal data. The aim of GDPR is to control and improve handling and processing of personal data particularly of European citizens. It oversees every aspect of citizens personal data handling and has recommended to impose heavy penalties for non compliance that may include prosecution of any organization in the world that is found guilty of privacy breach or misusing European citizens data [140].

GDPR is not only beneficial for the citizens but also for the organizations as it gives citizens confidence to share their data with the organizations when required. It also boosts organizations business and helps them in their smooth running without any hurdle of acquiring citizens data (without trust citizens usually do not share their data when required by the organizations, refer Section VI-D2 for discussion on mistrust between data provider and data handler). Even with all these obvious advantages, organizations in the past were rigid to adapt (at present they are forced to adapt) privacy regulations imposed by GDPR [141]. This is due to the fact that enterprises and organizations were facing challenges in implementing these regulations [142]. The organizations were already complying with the regulations imposed by the European Data Protection Directive (EDPD) of 1995 [143] and were not prepared for the new changes or possibly there was a lack of awareness of the new requirements raised by the GDPR. Another issue with the implementation of GDPR was financial needs, human resource requirements as well as proper training of the employees to understand the GDPR regulations [144].

GDPR defines six main data protection principles (other data protection principles further clarify them or further enhance them) that organizations (healthcare organizations) have to comply with when processing European citizens personal data [145].

Each of these principles is briefly explained below with implications on EHR data.

1) **Lawfulness, fairness and transparency** (Article 5(1)(a)): This article states that citizens personal data must be processed lawfully, fairly and transparently. Lawful processing of data is further defined in Article 6, which states that in order to process personal data lawfully, it is necessary for the data controllers to set out/obey one of the following conditions (In this section the term ''data controller'' is used multiple times and in the context of this study this term refers to healthcare organizations which records and stores/hold personal data):
   - ''The data subject must be given consent (Article 6(1)(a))''.
   - ''Processing is necessary for the performance of a contract to which the data subject is party (Article 6(1)(b))''.
   - ''Processing is necessary for compliance with the law (Article 6(1)(c))''.
   - ''Processing is necessary to protect vital interest of the data subject (Article 6(1)(d))''.
   - ''Processing is necessary for the performance of a task carried out in the public interest (Article 6(1)(e)''.
   - ''Processing is necessary for a legitimate interest of the controller or third party (Article 6(1)(f)''.

   In order to process personal data lawfully, all the clauses of the Article 6 (mentioned above) are important to be followed by the data controllers but the most pertinent clause of the article 6 in the context of EHR data is 6(1)(a) that relates to the processing of personal data with the consent of the person whose data is being used. However, based on the employer-employee or physician-patient relationships, where one party (physician in our case) is in power and processes other party's personal data, consent is not a proper legal basis to be relayed upon [146]. This is due to the fact that data protection regulation requires consent should be genuinely free without any pressure/intimidation. It can only be possible if the patients have freedom in giving their consent or not and have a choice to withdraw their consent at any point of time without any detriment as easy as they gave it.
2) **Purpose limitations (Article 5(1)(b))**: Purpose limitations bounds organizations (healthcare organizations) and individuals to collect personal data only for a specific, explicit and legitimate purpose and the data must be used for achieving that purpose only. Data purpose must be clearly defined before its collection and it should not be further processed in a way that is incompatible with the original defined purpose(s).
3) **Data minimization (Article 5(1)(c))**: In order to use personal data, it must be limited to its primary purpose only. It must not be collected more than its need.
4) **Accuracy(Article 5(1)(d))**: In dealing with the citizens personal data it must be responsibly dealt for example, if the data needs updation and inaccurate or incomplete data elements need to be removed, all must be done with high accuracy.
5) **Storage limitations (Article 5(1)(e))**: Storage limitations refer to the fact that personal data must be deleted after it has been used and no longer further needed. It means data should be collected with a proper predefined time-line and it must be removed after the time-line is reached.
6) **Integrity and Confidentiality (Article 5(1)(f))**: It is the entire responsibility of the individuals or organizations (who want to process citizens personal data) to

ensure the safe processing of data and to protect it from unauthorized use. During processing, data must be safe from any accidental loss, damage or demolition and it must be protected against any unlawful use.

If analyzed critically, clauses (b-f) of Article 5 have contradictory nature in the context of EHR data concepts. The regulations mentioned in these clauses (such as data minimization, purpose limitation) limits the quantity of data collection and enforces its deletion soon after the purpose has been achieved. On the other hand, healthcare organizations encourage collecting more and more data and to save it for longer periods of time for the purpose of detailed analysis, mining and predictions [147], as discussed in Section V.

Article 25 further enhances the ideas presented in Article 5 by defining privacy by design i.e. ''The controller must implement appropriate technical and organizational measures for ensuring that, by default, only personal data which are necessary for each specific purpose of the processing are processed''. Although, this Article enhances protection of personal data by demanding privacy by design form the controllers but it is difficult to implement because of its broader definition and due to the requirement of additional implementation cost and resources. Furthermore, privacy by design can show rigid behavior with time (like the other embedded technical solutions) because of not updating its measures frequently [148].

It has already been described in this study (refer Section VI) that the healthcare data is one of the most vulnerable data in terms of security threats, therefore needs special attention for protection during processing. Article 9 of GDPR defines the processing of such special categories of data, which requires additional protections in processing such as genetic data, biometric data, healthcare data etc. Article 9 imposes additional obligations and provides more restrictive legal basis for processing health related sensitive data. The recommendation of this article is to obtain explicit consent of collecting and processing sensitive personal data. Although, explicit consent of data processing is required in processing any type of personal data (Refer article 6(1)(a) mentioned above) but in case of processing healthcare data, obtaining consent is usually difficult, specially for secondary purpose, refer Section V for secondary uses of EHR data. Obtaining explicit consent for every secondary use is a time consuming, costly as well as an exhausting process [149]. There has been a great debate on obtaining specific consent in literature. The conclusive outcome of all such debates is to shift specific consent into a broader consent of data processing that covers the range of its future uses (such as secondary uses of EHR) [150].

At present, most of the patients are not aware (or do not want to be aware) about what happens to their data once it has been taken from them and also they do not know about the data processing procedures undertaken by the healthcare providers. According to Spiekermann *et al.* [151], if individuals knew about today's healthcare business model and how third parties use personal private data, they would be surprised and feel betrayed. Obviously, under such circumstances, obtaining broad consent is not logical.

Article 32 of GDPR defines security of processing of personal data. According to it, to process and maintain security of personal data pseudonymisation should be performed [13]. Pseudonymisation is a technique to ensure that an individual won't be identified through personal data (personal data includes direct and indirect identifiers that can identify a person for example, name, ID number, location, contact information (Article 4)) [13]. The process is to replace the main characteristic of an individual with randomly generated indicators. The information regarding identification must be stored separately [152]. Even if pseudonymisation technique is applied, it is possible to re-identify individuals by combining different data sets [153]. Re-identification pulls down the illusion of privacy policies, which are promised by technologists. Lawmakers should re-evaluate law and consider the weakness of pseudonymisation [154].

Other than the regulations described above, one of the most controversial regulation is the ''Right to be Forgotten'' (Article 17). This article imposes an obligation of erasure of one's personal data on the controllers. It gives the right to the users to erase their data any time from all the available places from where they want as per their request. According to concept of healthcare data where decision support and predictive systems are being made by archiving the patients' personal data (consider case of public health surveillance or clinical research, refer Section V), this article creates huge controversy because logically no more backups or archives of data would be applicable by the organizations.

B. HEALTH INSURANCE PORTABILITY AND ACCOUNTABILITY ACT (HIPAA)

In the year 1996 US congress made and passed HIPAA act to protect the patients' confidential health information contained in their medical records. HIPAA act brought forth multiple standards for addressing the privacy and security related issues with the patients Protected Health Information (PHI) [155]. In terms of HIPAA, PHI refers to the ''Individually Identifiable Information'' that is transmitted through electronic or any other media with the exception of educational or employment records [156]. The basic purpose of HIPAA was to secure healthcare information that is subjected to the health transactions. The HIPAA regulations were slowly implemented and its additional rules and regulations were released from time to time. During the period from 2002 to 2007, the following six additional regulations were released [157].

- In October 2002, standards for electronic claims and transactions were established.
- In April 2003, guidelines for the disclosure of patient health information were established.

- In June 2004, National Employer Identifier Rule (NERI) was established according to which the federal tax identification number was considered as the employer's national identifier.
- In April 2005, some technical as well as administrative protocols were established for the security and integrity of the patient's health information.
- In February 2006, HIPAA enforcement rule was established to guide how the government will enforce the organizations to implement HIPAA regulations.
- In May 2007, National Provider Identifier (NPI) rule was established based on which a national identifier for each provider was made and the procedures for spreading, storing, and updating the identifier was set up.

In 2009, Health Information Technology for Economic and Clinical Health (HITECH) Act was passed. The prime objective of this act was to promote across the USA, implementation of EHR by providing different incentive programs for adoption of EHR and penalties for not implementing it [158], [159]. Alongside the EHRs implementation, HITEC also expanded the existing HIPAA privacy and security regulations and enhanced the monetary penalties of HIPAA violations [160].

In 2013, HIPAA was further enhanced by adding new rules known as Omnibus rules 2013 [158]. After the Omnibus rules, HIPAA's coverage was expanded from healthcare providers (physicians, hospitals, insurance companies etc.) to third party administrators (pharmacy benefit managers, hospitals consultants etc.) who after the implementation of these rules, get punishments in case of any privacy breach [161].

In the year 2016, 21st Century Cures Act was passed [162], which further enhanced existing HIPAA regulations by solving several interoperability issues. The Cures Act defines specific policies to promote patient access to their data. Particularly, its purpose was to establish strong linkage and partnership between healthcare organizations and health information exchange organizations to ease patient access to their information contained in EHR, in a format that is easy to understand & handle, secure while accessing, and can be updated automatically [163].

HIPAA provides comprehensive guidelines to understand the use of technology for the collecting, storing and transmitting PHI. It focuses on streamlining procedures for implementation of different security measures but on the other hand it does not elaborate on how to practically implement such measures. After its legislation in 1966, HIPAA has not undergone any major iteration, therefore, its rules have been outdated and cannot provide much help in safeguarding against the vague threats of the present digital age [164]. Due to this reason, alternative frameworks are being adopted by the healthcare providers in order to give their systems full support against threats. For example, in complement to HIPAA, National Institute of Standards and Technology (NIST) provided a framework (published in 2014 and augmented in 2018) that specifically focuses the areas where HIPAA lacks (like, helping organizations by educating and training their employees) [165].

Another major issue with HIPAA regulations is that its protection applies only on covered entities i.e. health care providers, clinicians, pharmacies, health care facilities and healthcare clearinghouses but not on the non covered entities i.e. different types of information sharing platforms like social media posts [166]. Therefore, it can cause personal information leak [167]. In case of releasing patient information from uncovered entities like online e-commerce platforms, social media posts, fitness trackers data on the internet etc. HIPAA does not provide any protection.

In short, HIPAA with HITEC and Cure regulations have not advanced itself with fast growing technology. It does not completely fulfill patients' expectations of immediate availability of their health data electronically when needed [168] and robustly securing personal data. Therefore, HIPAA rules need further refinement [169] keeping in view of fast pace proliferation of technology and AI assisted gadgets.

In light of above discussion, HIPAA rules need major amendments in order to robustly protect patients personal and sensitive data and to make it accessible instantly where and when needed. There is also a need to define and include different non covered entities into HIPAA's scope to extend its protection to different information sources critically associated with the EHR data in the present digital era.

VIII. CONCLUSION

The objective of this research article is to provide overview of EHR and its various secondary uses, how such uses affect individuals privacy and whether the existing important privacy regulations i.e. GDPR and HIPPA overcome these privacy challenges. Article began with an overview of EHR, its data sources that contribute to making EHR. Then, different standards for sharing EHR data i.e. HL7 and FHIR are discussed. Subsequently, thorough analysis of various secondary uses of EHR with the aim to highlight how these secondary uses affect patients' privacy is presented. In the last, the article critically examined GDPR and HIPAA regulations and highlighted possible areas of improvement in these regulations, considering escalating use of technology and different secondary uses of EHR.

Presented article outlined various secondary uses of EHR to give readers an idea that how effectively EHR data can be used in different domains such as clinical research, public health surveillance and clinical audits to provide effective, timely and quality healthcare facilities to the patients, refer Section V. In order to use EHR data for secondary purposes more effectively, challenges associated with the secondary uses of EHR have also been described to make readers well aware of the EHR data challenges when using it for secondary purposes.

In the present technological era, adoption of EHR has positively impacted healthcare services. With the help of seamless

data sharing an individual can avail instant healthcare services at their location of preference. However, with evolving technology, risks of data security and compromise of privacy have also been significantly increased. EHR data contains highly personal and sensitive information i.e. ID/social security number, bank details, family information and medical history. Unauthorized access to EHR information can have devastating financial and social impact on individuals if such sensitive information is leaked in the public sphere. In this article different ethical and privacy issues arising from EHR data leak are discussed in detail in Section VI-D. In the referred section, data security and patients' privacy risks related to the secondary uses of EHR especially when EHR data is stored on cloud, transmitted through network and shared & exchanged with multiple stakeholders are critically studied.

There exists different privacy regulations to protect patients privacy and data security when EHR data is used for secondary purposes and transferred & exchanged with multiple concerned stakeholders through different linked locations. However, there is a need to critically examine such regulations to analyze them for calculating their effectiveness in terms of safeguarding personal data as per present era needs. There is also a need to highlight the challenges of such regulations to further improve their effectiveness in safeguarding personal data from the potential cyber attacks and to cope with the technological advancements of cyber attacks. In this study, important privacy regulations i.e. GDPR and HIPPA are studied in perspective of secondary use of EHR, refer Section VII. Our purpose is to highlight possible improvements areas in these regulations to make them more effective in protecting privacy and data security and to make them robust against escalating AI-assisted techniques in data analytics and cyber attacks.

REFERENCES

[1] S. Al-Janabi, I. Al-Shourbaji, M. Shojafar, and S. Shamshirband, "Survey of main challenges (security and privacy) in wireless body area networks for healthcare applications," *Egyptian Inform. J.*, vol. 18, no. 2, pp. 113–122, Jul. 2017.

[2] P. K. D. Pramanik, S. Pal, and M. Mukhopadhyay, "Healthcare big data: A comprehensive overview," in *Intelligent Systems for Healthcare Management and Delivery*. Hershey, PA, USA: IGI Global, 2019, pp. 72–100.

[3] M. Maghazil, "A comparative analysis of data security in computer-based and paper-based patient record systems from the perceptions of healthcare providers in major hospitals in Saudi Arabia," Ph.D. dissertation, School Eng. Appl. Sci., George Washington Univ., Washington, DC, USA, 2004.

[4] O. Ben-Assuli, "Electronic health records, adoption, quality of care, legal and privacy issues and their implementation in emergency departments," *Health Policy*, vol. 119, no. 3, pp. 287–297, Mar. 2015.

[5] C. Spiranovic, A. Matthews, J. Scanlan, and K. C. Kirkby, "Increasing knowledge of mental illness through secondary research of electronic health records: Opportunities and challenges," *Adv. Mental Health*, vol. 14, no. 1, pp. 14–25, Jan. 2016.

[6] D. F. Lobach and D. E. Detmer, "Research challenges for electronic health records," *Amer. J. Preventive Med.*, vol. 32, pp. S104–S111, May 2007.

[7] P. J. O'Connor, J. M. Sperl-Hillen, W. A. Rush, P. E. Johnson, G. H. Amundson, S. E. Asche, H. L. Ekstrom, and T. P. Gilmer, "Impact of electronic health record clinical decision support on diabetes care: A randomized trial," *Ann. Family Med.*, vol. 9, no. 1, pp. 12–21, Jan. 2011.

[8] A. Temko, W. Marnane, G. Boylan, and G. Lightbody, "Clinical implementation of a neonatal seizure detection algorithm," *Decis. Support Syst.*, vol. 70, pp. 86–96, Feb. 2015.

[9] T. Seymour, D. Frantsvog, and T. Graeber, "Electronic health records (EHR)," *Amer. J. Health Sci.*, vol. 3, no. 3, pp. 201–210, Jul. 2012.

[10] N. Menachemi and T. C. Collum, "Benefits and drawbacks of electronic health record systems," *Risk Manage. Healthcare Policy*, vol. 4, p. 47, May 2011.

[11] J. Wang, Z. Zhang, K. Xu, Y. Yin, and P. Guo, "A research on security and privacy issues for patient related data in medical organization system," *Int. J. Secur. Appl.*, vol. 7, no. 4, pp. 287–298, 2013.

[12] S. Teasdale, D. Bates, K. Kmetik, J. Suzewits, and M. Bainbridge, "Secondary uses of clinical data in primary care," *J. Innov. Health Inform.*, vol. 15, no. 3, pp. 157–166, Sep. 2007.

[13] *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons With Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46*, General Data Protection Regulation, 2016.

[14] J. P. Albrecht, "How the GDPR will change the world," *Eur. Data Protection Law Rev.*, vol. 2, no. 3, pp. 287–289, 2016.

[15] *Health Insurance Portability and Accountability Act of 1996*, Accountability Act, 1996.

[16] B. Hemsley, S. McCarthy, N. Adams, A. Georgiou, S. Hill, and A. S. Balandin, "Legal, ethical, and rights issues in the adoption and use of the 'my health record' by people with communication disability in Australia," *J. Intellectual Develop. Disab.*, vol. 43, no. 4, pp. 506–514, 2018.

[17] P. C.-I. Pang and S. Chang, "The Twitter adventure of #MyHealthRecord: An analysis of different user groups during the opt-out period," *Stud. Health Technol. Inform.*, vol. 266, pp. 142–148, Aug. 2019.

[18] R. Munir and R. A. Khan, "An extensive review on spectral imaging in biometric systems: Challenges & advancements," *J. Vis. Commun. Image Represent.*, vol. 65, Dec. 2019, Art. no. 102660.

[19] R. A. Khan, A. Crenn, A. Meyer, and S. Bouakaz, "A novel database of children's spontaneous facial expressions (LIRIS-CSE)," *Image Vis. Comput.*, vols. 83–84, pp. 61–69, Mar./Apr. 2019.

[20] M. S. Jaliaawala and R. A. Khan, "Can autism be catered with artificial intelligence-assisted intervention technology? A comprehensive survey," *Artif. Intell. Rev.*, vol. 53, no. 2, pp. 1039–1069, Feb. 2020.

[21] N. A. Latha, B. R. Murthy, and U. Sunitha, "Electronic health record," *Int. J. Eng.*, vol. 1, no. 10, pp. 25–27, 2012.

[22] K. Häyrinen, K. Saranto, and P. Nykäsnen, "Definition, structure, content, use and impacts of electronic health records: A review of the research literature," *Int. J. Med. Inform.*, vol. 77, no. 5, pp. 291–304, May 2008.

[23] K. Kawamoto, C. A. Houlihan, E. A. Balas, and D. F. Lobach, "Improving clinical practice using clinical decision support systems: A systematic review of trials to identify features critical to success," *BMJ*, vol. 330, no. 7494, p. 765, Apr. 2005.

[24] C. Castaneda, K. Nalley, C. Mannion, P. Bhattacharyya, P. Blake, A. Pecora, A. Goy, and K. S. Suh, "Clinical decision support systems for improving diagnostic accuracy and achieving precision medicine," *J. Clin. Bioinf.*, vol. 5, no. 1, p. 4, Dec. 2015.

[25] A. Boonstra and M. Broekhuis, "Barriers to the acceptance of electronic medical records by physicians from systematic review to taxonomy and interventions," *BMC Health Services Res.*, vol. 10, no. 1, p. 231, Dec. 2010.

[26] D. Kalra and D. Ingram, "Electronic health records," in *Information Technology Solutions for Healthcare*. London, U.K.: Springer, 2006, pp. 135–181.

[27] R. H. Dolin, L. Alschuler, C. Beebe, P. V. Biron, S. L. Boyer, D. Essin, E. Kimber, T. Lincoln, and J. E. Mattison, "The HL7 clinical document architecture," *J. Amer. Med. Inform. Assoc.*, vol. 8, no. 6, pp. 552–569, Nov. 2001.

[28] R. H. Dolin, L. Alschuler, S. Boyer, C. Beebe, F. M. Behlen, P. V. Biron, and A. Shabo, "HL7 clinical document architecture, release 2," *J. Amer. Med. Inform. Assoc.*, vol. 13, no. 1, pp. 30–39, Jan. 2006.

[29] T. Benson and G. Grieve, "HL7 version 2," in *Principles of Health Interoperability*. Cham, Switzerland: Springer, 2016, pp. 223–242.

[30] G. W. Beeler, "HL7 version 3—An object-oriented methodology for collaborative standards development," *Int. J. Med. Inform.*, vol. 48, pp. 151–161, Feb. 1998.

[31] T. Al-Enazi and S. El-Masri, "HL7 engine module for healthcare information systems," *J. Med. Syst.*, vol. 37, no. 6, p. 9986, Dec. 2013.
[32] D. Bender and K. Sartipi, "HL7 FHIR: An agile and RESTful approach to healthcare information exchange," in *Proc. 26th IEEE Int. Symp. Comput.-Based Med. Syst.*, Jun. 2013, pp. 326–331.
[33] M. Sharma and H. Aggarwal, "HL-7 based middleware standard for healthcare information system: FHIR," in *Proc. 2nd Int. Conf. Commun., Comput. Netw.* Singapore: Springer, 2019, pp. 889–899.
[34] A. Atreja, J.-P. Achkar, A. K. Jain, C. M. Harris, and B. A. Lashner, "Using technology to promote gastrointestinal outcomes research: A case for electronic health records," *Amer. J. Gastroenterol.*, vol. 103, no. 9, pp. 2171–2178, Sep. 2008.
[35] J. Lin, T. Jiao, J. E. Biskupiak, and C. McAdam-Marx, "Application of electronic medical record data for health outcomes research: A review of recent literature," *Expert Rev. Pharmacoeconomics Outcomes Res.*, vol. 13, no. 2, pp. 191–200, Apr. 2013.
[36] K. P. Liao, T. Cai, G. K. Savova, S. N. Murphy, E. W. Karlson, A. N. Ananthakrishnan, V. S. Gainer, S. Y. Shaw, Z. Xia, P. Szolovits, S. Churchill, and I. Kohane, "Development of phenotype algorithms using electronic medical records and incorporating natural language processing," *BMJ*, vol. 350, no. 11, p. h1885, Apr. 2015.
[37] K. Kreimeyer, M. Foster, A. Pandey, N. Arya, G. Halford, S. F. Jones, R. Forshee, M. Walderhaug, and T. Botsis, "Natural language processing systems for capturing and standardizing unstructured clinical information: A systematic review," *J. Biomed. Inform.*, vol. 73, pp. 14–29, Sep. 2017.
[38] F. Liu, C. Weng, and H. Yu, "Advancing clinical research through natural language processing on electronic health records: Traditional machine learning meets deep learning," in *Clinical Research Informatics*. Cham, Switzerland: Springer, 2019, pp. 357–378.
[39] D. Palaz, M. Magimai-Doss, and R. Collobert, "End-to-end acoustic modeling using convolutional neural networks for HMM-based automatic speech recognition," *Speech Commun.*, vol. 108, pp. 15–32, Apr. 2019.
[40] A. Mahendiran and R. Appusamy, "An intrusion detection system for network security situational awareness using conditional random fields," *Int. J. Intell. Eng. Syst.*, vol. 11, no. 3, pp. 196–204, Jun. 2018.
[41] A. Choudhury and C. M Greene, "Prognosticating autism spectrum disorder using artificial neural network: Levenb erg-marquardt algorithm," *Arch. Clin. Biomed. Res.*, vol. 2, no. 6, pp. 188–197, 2018.
[42] B. Shickel, P. J. Tighe, A. Bihorac, and P. Rashidi, "Deep EHR: A survey of recent advances in deep learning techniques for electronic health record (EHR) analysis," *IEEE J. Biomed. Health Informat.*, vol. 22, no. 5, pp. 1589–1604, Sep. 2018.
[43] P. Yadav, M. Steinbach, V. Kumar, and G. Simon, "Mining electronic health records (EHRs): A survey," *ACM Comput. Surv.*, vol. 50, no. 6, pp. 1–40, Jan. 2018.
[44] O. Sofrygin, Z. Zhu, J. A. Schmittdiel, A. S. Adams, R. W. Grant, M. J. Laan, and R. Neugebauer, "Targeted learning with daily EHR data," *Statist. Med.*, vol. 38, no. 16, pp. 3073–3090, Jul. 2019.
[45] V. Osmani, L. Li, M. Danieletto, B. Glicksberg, J. Dudley, and O. Mayora, "Processing of electronic health records using deep learning: A review," 2018, *arXiv:1804.01758*. [Online]. Available: http://arxiv.org/abs/1804.01758
[46] C. Xiao, E. Choi, and J. Sun, "Opportunities and challenges in developing deep learning models using electronic health records data: A systematic review," *J. Amer. Med. Inform. Assoc.*, vol. 25, no. 10, pp. 1419–1428, Oct. 2018.
[47] A. Esteva, A. Robicquet, B. Ramsundar, V. Kuleshov, M. DePristo, K. Chou, C. Cui, G. Corrado, S. Thrun, and J. Dean, "A guide to deep learning in healthcare," *Nature Med.*, vol. 25, no. 1, pp. 24–29, Jan. 2019.
[48] R. Miotto, L. Li, and J. T. Dudley, "Deep learning to predict patient future diseases from the electronic health records," in *Proc. Eur. Conf. Inf. Retr.* Cham, Switzerland: Springer, 2016, pp. 768–774.
[49] N. Wickramasinghe, "Deepr: A convolutional net for medical records," Tech. Rep., 2017.
[50] E. Choi, M. T. Bahadori, E. Searles, C. Coffey, M. Thompson, J. Bost, J. Tejedor-Sojo, and J. Sun, "Multi-layer representation learning for medical concepts," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2016, pp. 1495–1504.
[51] E. Choi, M. T. Bahadori, A. Schuetz, W. F. Stewart, and J. Sun, "Doctor AI: Predicting clinical events via recurrent neural networks," in *Proc. Mach. Learn. Healthcare Conf.*, 2016, pp. 301–318.
[52] A. N. Jagannatha and H. Yu, "Bidirectional RNN for medical event detection in electronic health records," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Hum. Lang. Technol.*, 2016, p. 473.
[53] A. Jagannatha and H. Yu, "Structured prediction models for RNN based sequence labeling in clinical text," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2016, p. 856.
[54] J. R. A. Solares, F. E. D. Raimondi, Y. Zhu, F. Rahimian, D. Canoy, J. Tran, A. C. P. Gomes, A. H. Payberah, M. Zottoli, M. Nazarzadeh, N. Conrad, K. Rahimi, and G. Salimi-Khorshidi, "Deep learning for electronic health records: A comparative review of multiple deep neural architectures," *J. Biomed. Inform.*, vol. 101, Jan. 2020, Art. no. 103337.
[55] V. Yadav and S. Bethard, "A survey on recent advances in named entity recognition from deep learning models," 2019, *arXiv:1910.11470*. [Online]. Available: http://arxiv.org/abs/1910.11470
[56] S. Silvestri, A. Esposito, F. Gargiulo, M. Sicuranza, M. Ciampi, and G. De Pietro, "A big data architecture for the extraction and analysis of EHR data," in *Proc. IEEE World Congr. Services (SERVICES)*, vol. 2642, Jul. 2019, pp. 283–288.
[57] P. Coorevits, M. Sundgren, G. O. Klein, A. Bahr, B. Claerhout, C. Daniel, M. Dugas, D. Dupont, A. Schmidt, and P. Singleton, "Electronic health records: New opportunities for clinical research," *J. Internal Med.*, vol. 274, no. 6, pp. 547–560, 2013.
[58] S. Mahmood, K. Hasan, M. C. Carras, and A. Labrique, "Global preparedness against COVID-19: We must leverage the power of digital health (preprint)," *JMIR Public Health Surveill.*, vol. 6, no. 2, Mar. 2020, Art. no. e18980.
[59] Y. Xiao, J. Wu, Z. Lin, and X. Zhao, "A deep learning-based multi-model ensemble method for cancer prediction," *Comput. Methods Programs Biomed.*, vol. 153, pp. 1–9, Jan. 2018.
[60] C. Willyard, "The drug-resistant bacteria that pose the greatest health threats," *Nature*, vol. 543, no. 7643, p. 15, Mar. 2017.
[61] I. H. Spicknall, K. J. Looker, S. L. Gottlieb, H. W. Chesson, J. T. Schiffer, J. Elmes, and M.-C. Boily, "Review of mathematical models of HSV-2 vaccination: Implications for vaccine development," *Vaccine*, vol. 37, no. 50, pp. 7396–7407, Nov. 2019.
[62] *Global Status Report on Alcohol and Health 2018*, World Health Org., Geneva, Switzerland, 2019.
[63] *Global Hepatitis Report 2017*, World Health Org., Geneva, Switzerland, 2017.
[64] S. M. Teutsch and R. E. Churchill, *Principles and Practice of Public Health Surveillance*. New York, NY, USA: Oxford Univ. Press, 2000.
[65] S. Ahmad, M. Asif, R. Talib, M. Adeel, M. Yasir, and M. H. Chaudary, "Surveillance of intensity level and geographical spreading of dengue outbreak among males and females in Punjab, Pakistan: A case study of 2011," *J. Infection Public Health*, vol. 11, no. 4, pp. 472–485, Jul. 2018.
[66] A. M. Schwartz, A. F. Hinckley, P. S. Mead, S. A. Hook, and K. J. Kugeler, "Surveillance for lyme disease—United States, 2008–2015," *MMWR Surveill. Summaries*, vol. 66, no. 22, pp. 1–12, Nov. 2017.
[67] K. E. Mace, P. M. Arguin, and K. R. Tan, "Malaria surveillance—United States, 2015," *MMWR Surveill. Summaries*, vol. 67, no. 7, pp. 1–28, 2018.
[68] E. Khan, J. Siddiqui, S. Shakoor, V. Mehraj, B. Jamil, and R. Hasan, "Dengue outbreak in Karachi, Pakistan, 2006: Experience at a tertiary care center," *Trans. Roy. Soc. Tropical Med. Hygiene*, vol. 101, no. 11, pp. 1114–1119, Nov. 2007.
[69] A. R. Anvikar, N. Shah, A. C. Dhariwal, G. S. Sonal, M. M. Pradhan, S. K. Ghosh, and N. Valecha, "Epidemiology of Plasmodium vivax malaria in India," *Amer. J. Tropical Med. Hygiene*, vol. 95, pp. 108–120, Dec. 2016.
[70] B. Shah and P. Mathur, "Surveillance of cardiovascular disease risk factors in India: The need & scope," *Indian J. Med. Res.*, vol. 132, no. 5, pp. 634–642, 2010.
[71] H. Siswoyo, M. Permana, R. P. Larasati, J. Farid, A. Suryadi, and E. R. Sedyaningsih, "EWORS: Using a syndromic-based surveillance tool for disease outbreak detection in Indonesia," *BMC*, vol. 2, no. S3, pp. 1–5, Dec. 2008.
[72] M. E. Sips, M. J. M. Bonten, and M. S. M. van Mourik, "Automated surveillance of healthcare-associated infections: State of the art," *Current Opinion Infectious Diseases*, vol. 30, no. 4, pp. 425–431, Aug. 2017.

[73] A. Atreja, S. M. Gordon, D. A. Pollock, R. N. Olmsted, and P. J. Brennan, "Opportunities and challenges in utilizing electronic health records for infection surveillance, prevention, and control," *Amer. J. Infection Control*, vol. 36, no. 3, pp. S37–S46, Apr. 2008.

[74] J. W. Keck, J. T. Redd, J. E. Cheek, L. J. Layne, A. V. Groom, S. Kitka, M. G. Bruce, A. Suryaprasad, N. L. Amerson, T. Cullen, R. T. Bryan, and T. W. Hennessy, "Influenza surveillance using electronic health records in the American Indian and Alaska native population," *J. Amer. Med. Inform. Assoc.*, vol. 21, no. 1, pp. 132–138, Jan. 2014.

[75] G. S. Birkhead, M. Klompas, and N. R. Shah, "Uses of electronic health records for public health surveillance to advance public health," *Annu. Rev. Public Health*, vol. 36, no. 1, pp. 345–359, Mar. 2015.

[76] W. Odero, J. Rotich, C. T. Yiannoutsos, T. Ouna, and W. M. Tierney, "Innovative approaches to application of information technology in disease surveillance and prevention in western kenya," *J. Biomed. Inform.*, vol. 40, no. 4, pp. 390–397, Aug. 2007.

[77] S. A. Ali, M. Salman, M. Din, K. Khan, M. Ahmad, F. H. Khan, and M. Arif, "Dengue outbreaks in Khyber Pakhtunkhwa (KPK), Pakistan in 2017: An integrated disease surveillance and response system (IDSRS)-based report," *Polish J. Microbiol.*, vol. 68, no. 1, pp. 115–119, 2019.

[78] P. C. Onyebujoh, A. K. Thirumala, and J.-B. Ndihokubwayo, "Integrating laboratory networks, surveillance systems and public health institutes in africa," *Afr. J. Lab. Med.*, vol. 5, no. 3, pp. 1–4, Oct. 2016.

[79] A. Kapoor, S. Guha, M. K. Das, K. C. Goswami, and R. Yadav, "Digital healthcare: The only solution for better healthcare during COVID-19 pandemic?" *Indian Heart J.*, vol. 72, no. 2, pp. 61–64, 2020.

[80] J. J. Reeves, H. M. Hollandsworth, F. J. Torriani, R. Taplitz, S. Abeles, M. Tai-Seale, M. Millen, B. J. Clay, and C. A. Longhurst, "Rapid response to COVID-19: Health informatics support for outbreak management in an academic health system," *J. Amer. Med. Inform. Assoc.*, vol. 27, no. 6, pp. 853–859, Jun. 2020.

[81] N. L. Bragazzi, H. Dai, G. Damiani, M. Behzadifar, M. Martini, and J. Wu, "How big data and artificial intelligence can help better manage the COVID-19 pandemic," *Int. J. Environ. Res. Public Health*, vol. 17, no. 9, p. 3176, May 2020.

[82] C. N. Shappell and C. Rhee, "Leveraging electronic health record data to improve sepsis surveillance," Tech. Rep., 2020.

[83] R. Burgess and J. Moorhead, *New Principles of Best Practice in Clinical Audit*. Abingdon, U.K.: Radcliffe Publishing, 2011.

[84] P. Esposito, "Clinical audit, a valuable tool to improve quality of care: General methodology and applications in nephrology," *World J. Nephrol.*, vol. 3, no. 4, p. 249, 2014.

[85] K. B. Bayley, T. Belnap, L. Savitz, A. L. Masica, N. Shah, and N. S. Fleming, "Challenges in using electronic health record data for CER: Experience of 4 learning organizations and solutions applied," *Med. Care*, vol. 51, pp. S80–S86, Aug. 2013.

[86] J. van der Lei, "Use and abuse of computer-stored medical records," *Methods Inf. Med.*, vol. 30, no. 2, pp. 79–80, 1991.

[87] W. R. Hogan and M. M. Wagner, "Accuracy of data in computer-based patient records," *J. Amer. Med. Inform. Assoc.*, vol. 4, no. 5, pp. 342–355, Sep. 1997.

[88] F. C. Bourgeois, K. L. Olson, and K. D. Mandl, "Patients treated at multiple acute health care facilities: Quantifying information fragmentation," *Arch. Internal Med.*, vol. 170, no. 22, pp. 1989–1995, 2010.

[89] T. Botsis, G. Hartvigsen, F. Chen, and C. Weng, "Secondary use of EHR: Data quality issues and informatics opportunities," *Summit Transl. Bioinf.*, vol. 2010, pp. 1–5, Mar. 2010.

[90] A. N. Kho, L. V. Rasmussen, J. J. Connolly, P. L. Peissig, J. Starren, H. Hakonarson, and M. G. Hayes, "Practical challenges in integrating genomic data into the electronic health record," *Genet. Med.*, vol. 15, no. 10, pp. 772–778, Oct. 2013.

[91] N. Ramakrishnan, D. Hanauer, and B. Keller, "Mining electronic health records," *Computer*, vol. 43, no. 10, pp. 77–81, Oct. 2010.

[92] G. Weimann, "Going dark: Terrorism on the dark Web," *Stud. Conflict Terrorism*, vol. 39, no. 3, pp. 195–206, Mar. 2016.

[93] C. S. Kruse, B. Frederick, T. Jacobson, and D. K. Monticone, "Cybersecurity in healthcare: A systematic review of modern threats and trends," *Technol. Health Care*, vol. 25, no. 1, pp. 1–10, Feb. 2017.

[94] J. Davis. (Oct. 2018). *3 Phishing Hacks Breach 20,000 Catawba Valley Patient Records*. [Online]. Available: https://www.healthcareitnews.com/news/3-phishing-hacks-breach-20000-catawba-valley-patient-records

[95] M. Ahmed and A. S. S. M. B. Ullah, "False data injection attacks in healthcare," in *Proc. Australas. Conf. Data Mining*. Singapore: Springer, 2017, pp. 192–202.

[96] M. R. Fuentes, "Cybercrime and other threats faced by the healthcare industry," Trend Micro, Tokyo, Japan, Tech. Rep., 2017.

[97] B. Edwards, S. Hofmeyr, and S. Forrest, "Hype and heavy tails: A closer look at data breaches," *J. Cybersecur.*, vol. 2, no. 1, pp. 3–14, Dec. 2016.

[98] H. Sharif and R. A. Khan, "A novel machine learning based framework for detection of autism spectrum disorder (ASD)," 2019, *arXiv:1903.11323*. [Online]. Available: http://arxiv.org/abs/1903.11323

[99] J. N. Giedd, "Structural magnetic resonance imaging of the adolescent brain," *Ann. New York Acad. Sci.*, vol. 1021, no. 1, pp. 77–85, Jun. 2004.

[100] H. Julien and I. Fourie, "Reflections of affect in studies of information behavior in HIV/AIDS contexts: An exploratory quantitative content analysis," *Library Inf. Sci. Res.*, vol. 37, no. 1, pp. 3–9, Jan. 2015.

[101] L. Bellak, "The schizophrenic syndrome and attention deficit disorder: Thesis, antithesis, and synthesis?" *Amer. Psychol.*, vol. 49, no. 1, pp. 25–29, 1994.

[102] A. W. Bateman, J. Gunderson, and R. Mulder, "Treatment of personality disorder," *Lancet*, vol. 385, pp. 735–743, Feb. 2015.

[103] J. G. Ronquillo, J. E. Winterholler, K. Cwikla, R. Szymanski, and A. Levy, "Health IT, hacking, and cybersecurity: National trends in data breaches of protected health information," *JAMIA Open*, vol. 1, no. 1, pp. 15–19, 2018.

[104] M. Meingast, T. Roosta, and S. Sastry, "Security and privacy issues with health care information technology," in *Proc. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2006, pp. 5453–5458.

[105] D. V. Dimitrov, "Medical Internet of Things and big data in healthcare," *Healthcare Inform. Res.*, vol. 22, no. 3, pp. 156–163, 2016.

[106] R. S. Weinstein, A. M. Lopez, B. A. Joseph, K. A. Erps, M. Holcomb, G. P. Barker, and E. A. Krupinski, "Telemedicine, telehealth, and mobile health applications that work: Opportunities and barriers," *Amer. J. Med.*, vol. 127, no. 3, pp. 183–187, Mar. 2014.

[107] F. Leu, C. Ko, I. You, K.-K.-R. Choo, and C.-L. Ho, "A smartphone-based wearable sensors for monitoring real-time physiological data," *Comput. Electr. Eng.*, vol. 65, pp. 376–392, Jan. 2018.

[108] C. D. Grood, A. Raissi, Y. Kwon, and M. J. Santana, "Adoption of e-Health technology by physicians: A scoping review," *J. Multidisciplinary Healthcare*, vol. 9, p. 335, Aug. 2016.

[109] L. Minor, "Harnessing the power of data in health," Stanford Med. Heal. Trends Rep., 2017.

[110] C. Esposito, A. De Santis, G. Tortora, H. Chang, and K.-K.-R. Choo, "Blockchain: A panacea for healthcare cloud-based data security and privacy?" *IEEE Cloud Comput.*, vol. 5, no. 1, pp. 31–37, Jan. 2018.

[111] S. Chenthara, K. Ahmed, H. Wang, and F. Whittaker, "Security and privacy-preserving challenges of e-Health solutions in cloud computing," *IEEE Access*, vol. 7, pp. 74361–74382, 2019.

[112] A. Abbas and S. U. Khan, "A review on the state-of-the-art privacy-preserving approaches in the e-Health clouds," *IEEE J. Biomed. Health Informat.*, vol. 18, no. 4, pp. 1431–1441, Jul. 2014.

[113] R. Zhang and L. Liu, "Security models and requirements for healthcare application clouds," in *Proc. IEEE 3rd Int. Conf. Cloud Comput.*, Jul. 2010, pp. 268–275.

[114] A. Nazir and R. A. Khan, "Combinatorial optimization based feature selection method: A study on network intrusion detection," 2019, *arXiv:1906.04494*. [Online]. Available: http://arxiv.org/abs/1906.04494

[115] P. Vimalachandran, H. Wang, Y. Zhang, G. Zhuo, and H. Kuang, "Cryptographic access control in electronic health record systems: A security implication," in *Proc. Int. Conf. Web Inf. Syst. Eng.* Cham, Switzerland: Springer, 2017, pp. 540–549.

[116] U. Premarathne, A. Abuadbba, A. Alabdulatif, I. Khalil, Z. Tari, A. Zomaya, and R. Buyya, "Hybrid cryptographic access control for cloud-based EHR systems," *IEEE Cloud Comput.*, vol. 3, no. 4, pp. 58–64, Jul. 2016.

[117] S. Chenthara, H. Wang, and K. Ahmed, "Security and privacy in big data environment," Tech. Rep., 2018.

[118] C. Guo, R. Zhuang, Y. Jie, Y. Ren, T. Wu, and K.-K.-R. Choo, "Fine-grained database field search using attribute-based encryption for E-healthcare clouds," *J. Med. Syst.*, vol. 40, no. 11, p. 235, Nov. 2016.

[119] K. Liu, "Secure electronic health record system based on online/offline KP-ABE in the cloud," in *Proc. 2nd Int. Conf. Internet Things, Big Data Secur.*, 2017, pp. 110–116.

[120] H. Cui, R. H. Deng, G. Wu, and J. Lai, "An efficient and expressive ciphertext-policy attribute-based encryption scheme with partially hidden access structures," in *Proc. Int. Conf. Provable Secur.* Cham, Switzerland: Springer, 2016, pp. 19–38.
[121] W. J. Gordon and C. Catalini, "Blockchain technology for healthcare: Facilitating the transition to patient-driven interoperability," *Comput. Struct. Biotechnol. J.*, vol. 16, pp. 224–230, Jan. 2018.
[122] L. Chen, W.-K. Lee, C.-C. Chang, K.-K.-R. Choo, and N. Zhang, "Blockchain based searchable encryption for electronic health record sharing," *Future Gener. Comput. Syst.*, vol. 95, pp. 420–429, Jun. 2019.
[123] C. E. Exceline and J. Norman, "Existing enabling technologies and solutions to maintain privacy and security in healthcare records," in *Security and Privacy of Electronic Healthcare Records: Concepts, Paradigms and Solutions*. 2019, p. 155.
[124] *Health Data in the Information Age: Use, Disclosure, and Privacy*, Committee Regional Health Data Netw. Inst. Med., Nat. Academies Press, Washington, DC, USA, 1994.
[125] M. Kayaalp, "Patient privacy in the era of big data," *Balkan Med. J.*, vol. 35, no. 1, pp. 8–17, 2018.
[126] L. J. Camp and M. E. Johnson, *The Economics of Financial and Medical Identity Theft*. Springer, 2012.
[127] N. R. Council, *Networking Health: Prescriptions for the Internet*. Washington, DC, USA: National Academies Press, 2000.
[128] SNS. (2015). *HIV Positive Couple Face Social Boycott*. [Online]. Available: https://www.thestatesman.com/world/hiv-positive-couple-face-social-boycott-83187.html
[129] W. P. Corrigan, C. A. Watson, P. Byrne, and K. E. Davis, "Mental illness stigma: Problem of public health or social justice?" *Social Work*, vol. 50, no. 4, pp. 363–368, 2005.
[130] D. J. Knutson, "Risk adjustment of insurance premiums in the United States and implication for people with disabilities," in *The Future of Disability in America*. Washington, DC, USA: The National Academies Press, 2007.
[131] A. Abbas, K. Bilal, L. Zhang, and S. U. Khan, "A cloud based health insurance plan recommendation system: A user centered approach," *Future Gener. Comput. Syst.*, vol. 43–44, pp. 99–109, Feb. 2015.
[132] G. Richter, C. Borzikowsky, W. Lieb, S. Schreiber, M. Krawczak, and A. Buyx, "Patient views on research use of clinical data without consent: Legal, but also acceptable?" *Eur. J. Hum. Genet.*, vol. 27, pp. 841–847, Jan. 2019.
[133] D. I. Shalowitz and F. G. Miller, "Disclosing individual results of clinical research: Implications of respect for participants," *J. Amer. Med. Assoc.*, vol. 294, no. 6, pp. 737–740, 2005.
[134] J. A. Bovenberg and M. Almeida, "Patients v. Myriad or the GDPR access right v. The EU database right," *Eur. J. Hum. Genet.*, vol. 27, no. 2, pp. 211–215, Feb. 2019.
[135] B. Sadan, "Patient data confidentiality and patient rights," *Int. J. Med. Inform.*, vol. 62, no. 1, pp. 41–49, Jun. 2001.
[136] District of Columbia U.S. Attorney's Office. (2012). *Former Howard University Hospital Employee Pleads Guilty to Selling Personal Information About Patients*. [Online]. Available: https://archives.fbi.gov/archives/washingtondc/press-releases/2012/former-howard-university-hospital-employee-pleads-guilty-to-selling-personal-information-about-patients
[137] N. Jamshed, F. Ozair, A. Sharma, and P. Aggarwal, "Ethical issues in electronic health records: A general overview," *Perspect. Clin. Res.*, vol. 6, no. 2, p. 73, 2015.
[138] *Health Insurance Portability and Accountability Act of 1996 (HIPAA)*, 1996.
[139] E. Politou, A. Michota, E. Alepis, M. Pocs, and C. Patsakis, "Backups and the right to be forgotten in the GDPR: An uneasy relationship," *Comput. Law Secur. Rev.*, vol. 34, no. 6, pp. 1247–1257, Dec. 2018.
[140] S. Sirur, J. R. C. Nurse, and H. Webb, "Are we there yet?: Understanding the challenges faced in complying with the general data protection regulation (GDPR)," in *Proc. 2nd Int. Workshop Multimedia Privacy Secur. (MPS)*, 2018, pp. 88–95.
[141] N. Gruschka, V. Mavroeidis, K. Vishi, and M. Jensen, "Privacy issues and data protection in big data: A case study analysis under GDPR," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2018, pp. 5027–5033.
[142] C. Tankard, "What the GDPR means for businesses," *Netw. Secur.*, vol. 2016, no. 6, pp. 5–8, Jun. 2016.
[143] *Directive 95/46/EC on the Protection of Individuals With Regard to the Processing of Personal Data and on the Free Movement of Such Data*, Directive, 1995.
[144] C. Tikkinen-Piri, A. Rohunen, and J. Markkula, "EU general data protection regulation: Changes and implications for personal data collecting companies," *Comput. Law Secur. Rev.*, vol. 34, no. 1, pp. 134–153, Feb. 2018.
[145] M. Goddard, "The EU general data protection regulation (GDPR): European regulation that has a global impact," *Int. J. Market Res.*, vol. 59, no. 6, pp. 703–705, Nov. 2017.
[146] M. J. Taylor and M. Prictor, "Insight or intrusion? Correlating routinely collected employee data with health risk," *Social Sci.*, vol. 8, no. 10, p. 291, Oct. 2019.
[147] O. Tene and J. Polonetsky, "Big data for all: Privacy and user control in the age of analytics," *Northwestern J. Technol. Intellectual Property*, vol. 11, no. 5, pp. 1–38, 2012.
[148] G. Bincoletto, "A data protection by design model for privacy management in electronic health records," in *Annual Privacy Forum*. 2019.
[149] M. Mostert, A. L. Bredenoord, M. C. I. H. Biesaart, and J. J. M. van Delden, "Big data in medical research and EU data protection law: Challenges to the consent or anonymise approach," *Eur. J. Hum. Genet.*, vol. 24, no. 7, pp. 956–960, Jul. 2016.
[150] C. A. Harle, E. H. Golembiewski, K. P. Rahmanian, B. Brumback, J. L. Krieger, K. W. Goodman, A. G. Mainous, and R. E. Moseley, "Does an interactive trust-enhanced electronic consent improve patient experiences when asked to share their health records for research? A randomized trial," *J. Amer. Med. Inform. Assoc.*, vol. 26, no. 7, pp. 620–629, Jul. 2019.
[151] S. Spiekermann, A. Acquisti, R. Böhme, and K.-L. Hui, "The challenges of personal data markets and privacy," *Electron. Markets*, vol. 25, no. 2, pp. 161–167, Jun. 2015.
[152] P. Voigt and A. von dem Bussche, "Scope of application of the GDPR," in *The EU General Data Protection Regulation*. Cham, Switzerland: Springer, 2017, pp. 9–30.
[153] T. Z. Zarsky, "Incompatible: The GDPR in the age of big data," *Seton Hall Law Rev.*, vol. 47, no. 4, p. 26, 2016.
[154] P. Ohm, "Broken promises of privacy: Responding to the surprising failure of anonymization," *UCLA Law Rev.*, vol. 57, p. 1701, Aug. 2009.
[155] A. M. Wheeler and B. Bertram, *The Counselor and the Law: A Guide to Legal and Ethical Practice*. Hoboken, NJ, USA: Wiley, 2019.
[156] T. Wilkinson and R. Reinhardt, "Technology in counselor education: HIPAA and HITECH as best practice," *Prof. Counselor*, vol. 5, no. 3, pp. 407–418, Jun. 2015.
[157] J. M. Kiel, F. A. Ciamacco, and B. T. Steines, "Privacy and data security: HIPAA and HITECH," in *Healthcare Information Management Systems*. Cham, Switzerland: Springer, 2016, pp. 437–449.
[158] N. Yaraghi and R. D. Gopal, "The role of HIPAA omnibus rules in reducing the frequency of medical data breaches: Insights from an empirical study," *Milbank Quart.*, vol. 96, no. 1, pp. 144–166, Mar. 2018.
[159] J. Pipersburgh, "The push to increase the use of EHR technology by hospitals and physicians in the United States through the HITECH Act and the Medicare incentive program," *J. Health Care Finance*, vol. 38, no. 2, pp. 54–78, 2011.
[160] W. Moore and S. A. Frye, "Review of HIPAA, part 1: History, protected health information, and privacy and security rules," *J. Nucl. Med. Technol.*, vol. 47, no. 4, pp. 269–272, 2019.
[161] C. J. Wang and D. J. Huang, "The HIPAA conundrum in the era of mobile health and communications," *J. Amer. Med. Assoc.*, vol. 310, no. 11, pp. 1121–1122, 2013.
[162] A. S. Kesselheim and J. Avorn, "New '21st century cures' legislation: Speed and ease vs science," *J. Amer. Med. Assoc.*, vol. 317, no. 6, pp. 581–582, 2017.
[163] C. T. Lye, H. P. Forman, J. G. Daniel, and H. M. Krumholz, "The 21st century cures act and electronic health records one year later: Will patients see the benefits?" *J. Amer. Med. Inform. Assoc.*, vol. 25, no. 9, pp. 1218–1220, 2018.
[164] D. Mohammed, "US healthcare industry: Cybersecurity regulatory and compliance issues," *J. Res. Bus., Econ. Manage.*, vol. 9, no. 5, pp. 1771–1776, 2017.
[165] S. M. Ahmed and A. Rajput, "Threats to patients' privacy in smart healthcare environment," in *Innovation in Health Informatics*. Amsterdam, The Netherlands: Elsevier, 2020, pp. 375–393.

[166] I. G. Cohen and M. M. Mello, "HIPAA and protecting health information in the 21st Century," *J. Amer. Med. Assoc.*, vol. 320, no. 3, pp. 231–232, 2018.
[167] W. Moore and S. Frye, "Review of HIPAA, part 2: Limitations, rights, violations, and role for the imaging technologist," *J. Nucl. Med. Technol.*, vol. 48, no. 1, pp. 17–23, Mar. 2020.
[168] B. S. Lee, J. Walker, T. Delbanco, and J. G. Elmore, "Transparent electronic health records and lagging laws," *Ann. Internal Med.*, vol. 165, no. 3, pp. 219–220, 2016.
[169] S. T. Rosenbloom, J. R. L. Smith, R. Bowen, J. Burns, L. Riplinger, and T. H. Payne, "Updating HIPAA for the electronic medical record era," *J. Amer. Med. Inform. Assoc.*, vol. 26, no. 10, pp. 1115–1119, Oct. 2019.

SHAHID MUNIR SHAH received the M.Sc. degree in electronics and M.S. degree in telecom. He received the Ph.D. degree in information technology (IT) from the University of Sindh, Jamshoro. He has vast teaching experience in mathematical and applied sciences background in national and international organizations of repute. He is currently working as an Assistant Professor with Barrett Hodgson University, Karachi, Pakistan. His research interests include machine learning, signal processing, natural language processing, and pattern recognition.

RIZWAN AHMED KHAN received the Ph.D. degree in computer vision from Université Claude Bernard Lyon 1, France, in 2013. He has worked as Postdoctoral Research Associate with Laboratoire d'InfoRmatique en Image et Systemes d'information (LIRIS), Lyon, France. He is currently working as a Professor with Barrett Hodgson University, Karachi, Pakistan. His research interests include machine learning, computer vision, image processing, pattern recognition, and human perception.