NFSU – Journal of Cyber Security and Digital Forensics
Special Issue – 1, 2024
E – ISSN – 2583-7559

Blockchain Revolution: Securing Electronic Health Records for Patient Empowerment and Innovative Healthcare

Shruti Saxena
School of Cyber Security and Digital ForensicsNational Forensics Sciences University Gandhinagar, Gujarat, India.

Nikunj Tahilrahamani
School of Cyber Security and Digital ForensicsNational Forensics Sciences University Gandhinagar, Gujarat, India.

***Abstract*— Blockchain technology has emerged as a promising technology for healthcare, enabling the development of a secure, efficient, and patient-centric healthcare system. This paper examines how blockchain technology might completely change how electronic health records (EHRs) are managed and how healthcare is provided. The results underscore the noteworthy advantages that blockchain provides, such as improved security, patient empowerment, and novel opportunities for data analysis and prediction. Blockchain has the power to transform EHR administration, encourage patient autonomy, and spur innovation in the medical field.**

***Index Terms*— Blockchain, Electronic health record (EHR), Artificial Intelligence, Health-care, Medical data, Sharing and protection, Smart-contract.**

I. Introduction

Healthcare applications encompass electronic health records (EHR), clinical data, decision-making, genetics, neurology, biomedicine, and medicines. IoT technology, facilitated by communication protocols and data standard- ization, holds promise for effective healthcare services. Improving connection, user interfaces, patient data secu- rity, and interoperability can enhance healthcare delivery. Current research focuses on developing reliable health- care apps for both the public and the healthcare sector. Maintaining up-to-date, secure access to health records is crucial for multiple parties, including hospitals, phar-macies, and patients. Blockchain technology, depicted in Figure 1, emerges as a solution to address healthcare sector challenges.

Healthcare researchers stress the importance of multi-source data analysis for identifying community health hazards and improving patient care quality [1]. Integrat- ing blockchain, AI, and other technologies is crucial for business success [3]. To enhance medical exploration,

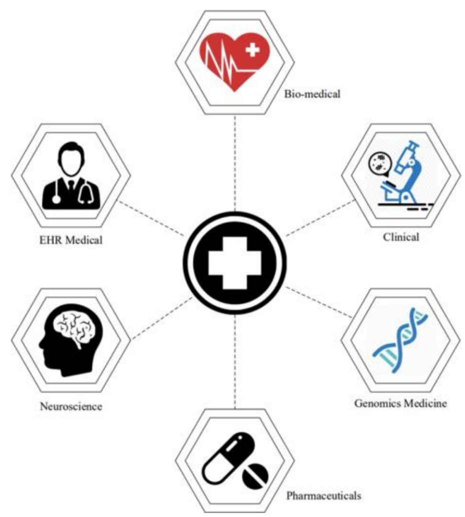

Fig. 1: Applications of Blockchain in Health-care

the industry needs technology for client-centric interfaces and data-driven insights [2], [5]. AI prioritizes cases for medicine monitoring and development, optimizing drug formulations [3]. Blockchain enables secure data analysis, but challenges arise in storing massive records, addressed by integrating AI algorithms [4]. Despite AI's powerful capabilities, reluctance exists due to concerns about rapid execution of dynamic processes [6]. The machine sector demonstrates AI's prowess in applications like autonomous vehicles and financial analysis [6].

A. Perks of using Artificial Intelligence with Blockchain

The fight against the coronavirus pandemic saw the effective use of blockchain and AI. Blockchain securely tracked patient data, aiding in early threat identification and preventing unauthorized disclosure. It streamlined vaccine trials, ensured transparency in fundraising, and assisted in epidemic control. AI analyzed extensive data on environmental factors, healthcare access, and transmission patterns, predicting disease outbreaks and distinguishing coronavirus cases from similar conditions. Genome-based neural networks effectively managed complications, con- sidering individual immunity and genetic structure. AI ac- celerated the development of vaccines for emerging strains and created automated models correlating medical records with treatment outcomes. Recognizing these benefits, the White House requested the use of AI in the US govern- ment's pandemic response [4], [5], [7].

Blockchain's decentralization, preventing a central au- thority from solely controlling data, processing, and model validation, offers advantages [9]. It reduces processing time, errors, and costs, enabling the development of predictive models for clinical practice and pandemic response. The immutability of blockchain-verified transactions ensures data authenticity and enhances system security [9]. The cryptographic nature and immutability of data, coupled with the absence of a central authority, build trust among participants, minimizing the need for individual verifica- tion [10].

B. The Imminentities of the Healthcare Systems

Blockchain's potential in healthcare is undeniable, yet widespread adoption faces significant roadblocks. Three key challenges—scalability, interoperability, and regulatory ambiguity—pose critical hurdles.

- Scalability
  Current blockchain networks often grapple with the immense volume of healthcare data, especially with extensive Electronic Health Records (EHRs) covering medical history, diagnoses, medications, and more. When multiplied by millions of patients, the data exceeds the processing capacity of many existing blockchain platforms. It is essential to develop scal- able solutions that can efficiently manage this data deluge for the widespread adoption of blockchain in healthcare.
- Interoperability
  Healthcare systems are known for their fragmentation, featuring numerous incompatible software and data formats. Successful integration of blockchain tech- nology into this complex ecosystem requires seam- less interoperability with existing systems. Without it, clinicians may encounter duplicative data entry, fragmented patient records, and limited communica- tion among healthcare providers. Overcoming inter-operability challenges demands collaboration among technology developers, healthcare institutions, and policymakers to establish standardized formats and protocols.
- Regulatory Ambiguity
  Blockchain technology in healthcare operates in a legal grey area, with uncertainties surrounding data privacy, ownership, and security protocols hindering widespread adoption. Governments and regulatory bodies must develop clear legal frameworks to ad- dress these concerns and provide guidelines for the responsible use of blockchain in healthcare data man- agement.

Overcoming hurdles in scalability, interoperability, and regulatory challenges is essential to unlock the full potential of blockchain in healthcare. Addressing these issues can pave the way for a future where secure, transparent, and patient-centric healthcare data management becomes a reality, ultimately leading to improved healthcare delivery and patient outcomes.

The remaining paper is organized as follows: Section II introduces blockchain technology. Section III explores AI in health records management. Section IV examines the synergy of blockchain and AI. Section V focuses on privacy and security in blockchain-based Electronic Health Record (EHR) management. Section VI delves into using blockchain for privacy in healthcare records. Section VII addresses hurdles in implementing AI-based blockchain technology in healthcare. Finally, Section VIII concludes, highlighting the transformative potential of blockchain and AI in health records management.

II. Blockchain's history within the Health Records Management System

A. Blockchain

Blockchain is a distributed database that aggregates data blocks in a chronological manner and primarily addresses trust and security concerns related to transactions. The blockchain can be broadly classified into three classes: consortium, private, and public blockchains [11], [12]. As new blocks are consistently added to the chain, it grows. To ensure ledger integrity and user security, blockchain technology combines two technologies: P2P distributed consensus and asymmetric cryptography. As a result, a cryptographic hash connects these time-stamped blocks [13].

Every blockchain, as seen in Figure 2, is made up of numerous blocks, each of which has a block body and a block header. Multiple meta-information about the current block is contained in the block header. Examples include the timestamp, nonce for the blockchain body, hash of current block and a hash value for the previous block. Typically, block bodies are used to capture the actual data of ongoing transactions.

- Time stamp: the timestamp is the moment of creation of the current block.
- Nonce: each time a miner hashes a piece of code, they change a four-byte random field to solve a PoW problem
- Previous block hash: the value of the previous block's hash.
- Hash of the current block: A hash algorithm like SHA-256 is used to create a distinct, cryptographically

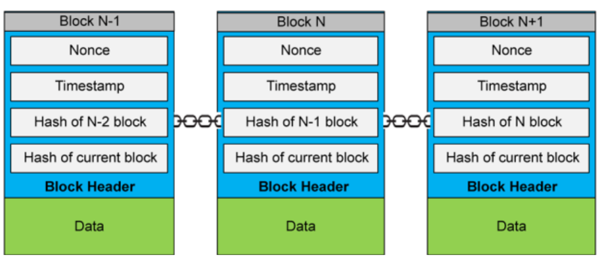

Fig. 2: The structure of the blockchain

secure hash for each block in a blockchain. By serving as the block's "fingerprint," this hash guarantees the block's integrity and immutability. Any modification to the data within the block would produce an entirely new hash, which would make it simple to identify any manipulation or tampering.

The following are the blockchain's primary characteristics[12]:

1) Decentralization
every node is equal and there isn't a central node. Many nodes spread over various locations perform the task of recording transaction information, and each node maintains a comprehensive account. Every node has the ability to oversee the transaction and affirm it together.
2) Resistance against tampering
the subsequent block contains the hash value of the preceding block. Every subsequent block will require recalculation in the event that one is altered. It istherefore invalid for one node to have modified the database.
3) Openness
The blockchain's data is accessible to everyone, in ad-dition to encrypting the private information of each party to the transaction. Anyone can use the public interface to access block data and create pertinent apps.
4) Autonomy
All nodes in the system can freely and securely exchange data since the blockchain uses a consensual protocol (like an open and transparent algorithm). So, a person will not get involved.
5) Anonymity
Since the exchange between nodes is based on a pre-determined mechanism, the other side does not need to establish confidence by revealing their identify inpublic.

In the Merkle tree, each leaf node represents a transac- tion, and non-leaf nodes store the hash value of the con- catenated child nodes. To verify a transaction's existence and integrity, nodes only need to check the hash value of the two concatenated child nodes in the leaf node. Any modification to a transaction generates a new hash value at the top layer, resulting in a single root hash. The transaction size impacts the maximum transactions per block and the overall block size. All blocks are linked through the hash function, preventing changes or removal of verified data. Each block modification creates a new block with a new hash value and link. Security and immutability are crucial aspects of blockchain technology.

1) **Digital signature:** In an untrustworthy environment, asymmetric cryptography is frequently used for transaction authentication [13]. An essential part of the Blockchain is asymmetric cryptography, which is used to transmit and confirm the validity of transactions. Before being received in a P2P network, transactions are signed using the private key of the person who initiated them. The elliptic curve digital signature algorithm (ECDSA) is employed by the majority of modern blockchains [14].

Peers in a peer-to-peer (P2P) network have equal priv- ileges, and when a transaction is requested or started, a block describing that transaction is generated and broad- cast to all neighbouring nodes. Other nodes will receive this block. Using predetermined block validation procedures, the sender's public key is utilized to confirm the legitimacy of the received block.

The block will be sent to more nodes until they have all confirmed it if it is authentic. If not, it will be thrown away while the process is underway. The blockchain network can only store and adds legitimate blocks.

B. Consensus Algorithms: Building Trust in Decentralized Networks

In the world of blockchain and other decentralized sys- tems, where trust is paramount but central authorities are absent, consensus algorithms step up to the plate. These sophisticated protocols allow nodes in a network to agree on a single version of the truth, even amidst unreliable or malicious actors. Now let's explore a few important consensus algorithms:

1) Proof of Work (PoW)
   The granddaddy of consensus algorithms, PoW re-lies on computational power to validate transactions. Miners compete to solve cryptographic puzzles, and the first to find the solution gets to add a block to the blockchain. This energy-intensive method secures the network but faces scalability concerns. [15][16][17]
2) Proof of Stake (PoS)
   A greener alternative to PoW, PoS rewards nodes based on their stake in the network's cryptocurrency. Instead of solving puzzles, validators lock up their coins, and the system randomly selects one to vali- date the next block. This incentivizes honest behavior and reduces energy consumption. [18]
3) Byzantine Fault Tolerance (BFT)
   Designed for tolerating even malicious actors, BFT al- gorithms employ complex communication protocols to ensure agreement. Nodes exchange messages and compare versions of the ledger, eventually converging on a single version even if some nodes act out of spite. However, BFT can be computationally expen- sive for large networks. [20]
4) Delegated Proof of Stake (DPoS)
   A variant of PoS, DPoS delegates the validation power to a limited set of elected validators. Stake holders vote for their representatives, who then secure the network. This can improve efficiency and scalability but raises concerns about centralization. [19]
5) Proof of Authority (PoA)
   Nodes that have the capacity to create new blocks are allowed. Pre-authentication is the initial step that every node needs to complete. However, this process results in a design that is oriented toward the natural world. [18]
6) Proof of Capacity (PoC)
   This consensus method uses hard disc space that is accessible instead of computational resources to reach consensus [21]. You could store more solutions if you had more storage space, which would raise the possibility that a new block will be created.

To increase performance in a range of applications, there is a growing trend toward combining multiple consensus algorithms rather than relying just on one.

C. Smart Contract

Self-executing programs, known as smart contracts, op- erate on the blockchain, finding applications in govern- ment, healthcare, and finance. These contracts execute programmable functions when a transaction is sent to the designated contract address, automatically fulfilling the predefined terms within the secure container. Ethereum, an open-source blockchain platform, introduced Turing- complete smart contract languages, allowing the creation of decentralized applications (Dapps). Dapps, such as those facilitating patient-physician communication on the Ethereum network, empower patients with greater control over their records [22].

D. Essentials for a Medical Data Share and Protection Plan

The ideal system for exchanging and protecting medical data should meet the following fundamental needs [23], [24]:

1) Security and privacy protection
   no one could utilize medical data improperly. Ma- licious attacks should be able to be resisted by the plan, and illicit activity may be tracked down.
2) Data access
   Patients can view all of their medical records after gaining authorization, and clinicians can examine past medical records with patients' consent.
3) Patient control
   The patient was in charge of their own medical

history; that is, no one else could access their medical history without their consent.

4) Unified standard

To implement data exchange and increase system stability, all players in the model should utilize the unified data standard and management scheme.

III. A Review of AI in the System for Managing Health Records

AI systems in healthcare commonly employ either su- pervised or unsupervised methods. Supervised learning uses labeled data to train machines in predicting classi- fications based on input predictors. On the other hand, unsupervised methods identify patterns and correlations without relying on labeled data or anticipating specific outcomes. The majority of machine learning structures in healthcare, especially those developing electronic pheno- type algorithms, predominantly utilize supervised learning methods. This section provides a brief overview of various machine learning approaches extensively used for cate- gorizing clinical outcomes from electronic health records, including random forests, support vector machines, as well as both supervised and unsupervised models for deep learning and neural networks [26], [27], [28].

Support vector machines establish an optimal discon- nected hyperplane in the co-variate space between ob- servations of distinct outcome groups to identify vari- ables [26]. This hyperplane, with the maximum margin or distance separation from the nearest observations on either side, is determined as the 'support vectors' [31]. On the contrary, random forests function as regulation-based batch classifiers that identify inputs by averaging estimates across a group of decision tree models [32]. Each tree in the random forest classifier is trained with a bootstrap sample of data points, and the sample is split at each node based on the most descriptive among a randomized subset of potential predictors [26]. In recent years, the popularity of deep learning models and neural networks has surged, especially in diagnostic imaging and forecasting activities [26]. When applied in the supervised learning approach, these models are considered progressively complex exten- sions of the classic regression paradigm [26].

Conventional logistic regression models consist of an input and output layer, with a node in the input layer for each parameter and a set of relation weights linking input nodes to coefficients or the output node. The output node computes the total from each parameter multiplied by the corresponding relation weight, known as the in- put nodes' weighted sum, and processes it through the activation function, such as the logistic function. Neural networks enhance this structure by incorporating a hidden layer between the output and input layers, allowing the simulation of more complex and non-linear relationships between response and predictor factors [25]. Subsequently, deep learning models improve upon this by introducing multiple hidden layers between the output and input layers to identify even more subtleties in the data [28].

Despite their ability to autonomously simulate intricate interactions and relationships in data points, these algo- rithms are challenging to interpret, susceptible to overfit- ting, and often require substantial training data to yield accurate results [30]. To justify and optimize the application of machine learning for categorizing health responses from electronic health record (EHR) data, researchers should carefully consider when these techniques are most appro-priate and for which tasks they should be employed [31].

A. The Difficulties in Applying AI to a Health Records System

1) **Transparency:** Transparency is a critical aspect con-cerning the intricate machine learning algorithms like deep learning, which poses challenges in their application to phenotyping tasks, especially when high stakes and end- user confidence are paramount. Clinicians, for instance, prefer algorithms to complement their expertise rather than dictate decision-making processes outright [33]. Regulatory authorities demand decipherability for transparency, as classification systems may have significant legal or financial implications [34]. Enhancing the interpretability of these "black box" models is pivotal, as demonstrated in prior re- search utilizing novel approaches like 'saliency' to interpret deep learning model predictions [34, [35]. These initiatives, such as creating heatmaps based on mappings of class activation, improve the transparency and interpretability of complex machine learning models, thereby strengthening trust among physicians and end-users and encouraging broader applications [36].

2) **Interpretability and Explainability :**

Interpretability and explainability are closely related concepts that address the need for AI algorithms to provide understandable and meaningful insights into their decision-making processes. Interpretability refers to the ability to comprehend and make sense of the outputs generated by AI algorithms, while explainability involves providing clear and intelligi- ble explanations for those outputs. In healthcare, where decisions directly impact patient well-being, understanding why an AI system recommends a particular course of action is paramount. However, many AI models used in health records systems, such as deep learning neural networks, are often complex and difficult to interpret or explain. This lack of interpretability and explainability hinders the adoption of AI in healthcare settings, as clinicians may be hesitant to rely on AI-driven recommendations without a clear understanding of how they were generated.

3) **Transportability:** Electronic phenotyping research, often conducted in a single-site environment due to privacy concerns and administrative constraints, faces challenges in achieving transportability [37], [38], [39]. Despite this, there is a growing interest in exchanging algorithms among researchers and healthcare bodies to enhance versatility [40]. Platforms like the Phenotype Knowledgebase (PheKB) facilitate the development, sharing, and validation of elec- tronic phenotyping algorithms, showing progress in this domain [40]. To build scalable phenotyping algorithms, external verification is essential for determining portability, followed by modifications if needed to account for site- specific idiosyncrasies [39]. The "validation-adaptation" strategy proves valuable, especially for NLP-based pheno- typing algorithms, though it may be laborious. External validation before implementation in different settings is crucial to address overfitting concerns, particularly for machine learning-based phenotyping algorithms [41]. The portability of deep learning models for health responses from imaging procedures may be more feasible compared to those used for NLP tasks due to the lower heterogeneity in medical images across different sites.

4) **Training Size:** To achieve optimal efficiency, machine learning algorithms, particularly deep learning frameworks, require a substantial amount of labeled training data [42]. The CheXNet algorithm, for example, achieved expert-level results after being trained on over 100,000 labeled images [35]. However, many researchers face limitations in time or resources for data annotation, and pooling labeled data across different locations may not be practical due to privacy and administrative hurdles [42], [38]. Innovative solutions like image data enhancement and active learning, where the most insightful instances are successively se-lected for training, can help reduce the amount of training data needed for satisfactory performance [43], [44]. In the context of electronic phenotyping for rheumatoid arthritis, combining active learning with support vector machines lowered the annotated samples needed to achieve an AUC of 0.95 by 68

B. An AI Algorithm for Health-Care Systems

This article facilitates a collaborative discussion between machine learning and data mining as both fields are rooted in data science and often intersect [45]. Despite this convergence, there exist fundamental distinctions between the two disciplines. Machine learning, characterized by its focus on methods capable of autonomously extracting information, encompasses the study of such processes [45]. To predict future events, machine learning relies on two datasets, namely training data and test data. On the contrary, data mining represents an iterative exploration of diverse and valuable patterns within datasets. While data mining can incorporate machine learning, it is notlimited to this technique alone, utilizing various other methods to unveil novel patterns. The healthcare sector extensively employs machine learning and data mining technologies to extract insights from extensive electronic health data. The inclusion of both machine learning anddata mining approaches in the analysis conducted in [46] stems from their shared mechanisms in disease prediction, with literature often intertwining discussions of the two.

1) **Supervised Algorithm:**

A. Artificial Neural Network (ANN)

Artificial Neural Networks (ANNs), initially proposed by McCulloch and Pitts [46] and popularized in the 1980s [48], have become instrumental in addressing various categorization challenges. The term 'neutral' in their nomenclature reflects their design to emu-late brain-inspired systems, mirroring the way hu- man brains learn categories. ANNs are modeled after the human brain's vast network of interconnected neurons with the ability to strengthen or weaken connections through reinforced labeled training data, resembling biological learning. These neuronal con- nections are represented by a weighted matrix, anal- ogous to cortical layers in the brain, where training data functions as a form of 'biological learning' for individuals. ANNs typically incorporate one or more hidden layers, supplementing the input and output layers, and are trained to generate outputs based on given input variables. While numerous studies have explored ANNs in the context of survival prediction, limited research utilizing electronic health data is available.

Deep learning, a subset of machine learning inspired by ANNs [49], has seen increased utilization in modeling illness symptoms and hazards. Liu et al.

[50] developed a deep learning-based technique for early Alzheimer's disease and mild cognitive impairment identification in 2014, leveraging neuroimaging data from the Alzheimer's Disease Neuroimaging Initiative database. To overcome limitations, they employed stacked autoencoders. Another study by Cheng et al. [51] proposed a method for phenotyping patient electronic health records (EHRs) using deep learning. Each patient's EHR was initially represented as a temporal matrix, with time and events as axes. They constructed a four-layer Convolutional Neural Network (CNN) to extract phenotypes and predict risks. In a recent development, [52] introduced the Heterogeneous Convolutional Neural Network (HCNN), a predictive learning model representing EHRs as graphs with diverse properties like diagnosis, procedures, and medications. This information was harnessed to construct a novel risk prediction model for multiple comorbid conditions. The integration of ANNs and deep learning methods holds promise in advancing healthcare predictive modeling, though further exploration in utilizing electronic health data is warranted.

B. Support Vector Machine (SVM)

Support Vector Machine (SVM) stands out as a pop- ular supervised learning approach proficient in clas-sifying both linear and non-linear data. SVMs oper-ate by transforming the input vector into a higher- dimensional feature space, seeking to identify the hyperplane that effectively segregates data points into two distinct groups. In the context of classification tasks, SVM aims to maximize the marginal distance between the two classes while minimizing classifi- cation errors. The marginal distance is defined as the space between the decision hyperplane and its nearest instance belonging to a specific class [53]. In more technical terms, each data point is initially represented as a point in an n-dimensional space (where n denotes the number of features), with each feature value serving as a coordinate. The key ob-jective is to identify the hyperplane that maximizes the separation between the two classes, creating the most significant margin for effective classification. SVM has found applications in various domains, including bioinformatics and healthcare, where its robust categorization capabilities have been utilized [53]. The versatility of SVM makes it a valuable tool in diverse fields, emphasizing its efficacy in handling complex classification tasks across different domains. As technological advancements continue, the synergy between SVM and emerging technologies such as Explainable AI (XAI) becomes increasingly relevant. Ensuring the interpretability of SVM models in healthcare, for instance, is a critical consideration for fostering trust among healthcare professionals and stakeholders [54]. This intersection of SVM with interpretability initiatives aligns with the evolving landscape of responsible and ethical AI adoption.

C. Decision Tree Random Forest

A Decision Tree (DT) represents a sophisticated and deterministic data structure resembling a tree, where internal nodes depict input variables or attributes, and leaves signify decision outcomes. This hierarchi- cal arrangement of nodes and leaves is orchestrated to formulate a strategy for achieving a specific cate-gorization goal. The leaves, situated at the terminal level of the relevant branch, house the final decision outcomes, while nodes can be distributed across multiple levels. The root node marks the inception of the tree, akin to a flowchart where each non-leaf symbolizes a test on a single property, each branch indicates the outcome of the test, and each leafsignifies the associated class label.

Decision trees find extensive application in the healthcare sector, where academics leverage them for various purposes. For instance, a decision tree- based prognostic approach was proposed to quantify disease recurrence and predict survival in breast can-cer patients [53]. This model, designed for predicting breast cancer survival, utilized two machine learning techniques (Artificial Neural Network and decision tree) along with a statistical approach (logistic regres- sion). The SEER breast cancer database, recognized as a vital population-based repository for evaluating cancer care quality, played a pivotal role in training and evaluating the model's performance.

Random Forest [53], an ensemble classifier compris- ing multiple decision trees, has gained prominenceas one of the most accurate machine learning-based algorithms. Combining Breiman's 'Bagging' concept with the random selection of features, this method creates a collection of decision trees characterized by controlled variation. In a study by [55], researchers proposed a classifier based on the random forest algorithm to estimate illness risk among individuals,

employing the Healthcare Cost and Utilization Project (HCUP) dataset for their research. Furthermore, [56] developed a diabetes risk prediction model using a scalable random forest classification algorithm based on administrative data.

The versatility of decision trees and their ensemble counterparts, such as random forest, showcases their efficacy in handling complex classification tasks in healthcare. Their interpretability and abilityto capture non-linear relationships make them valuable tools for researchers and practitioners alike. As machine learning continues to advance, decision trees remain integral components in predictive modelling within the healthcare domain, contributing significantly to advancements in disease prognosis, risk estimation, and personalizedhealthcare.

D. K-Nearest Neighbors (k-NN)

K-Nearest Neighbours (k-NN) is a versatile and intu- itive supervised learning algorithm that operates onthe principle of proximity. The core idea behind k-NNis to classify an unknown data point based on the ma- jority class of its k-nearest neighbours in the feature space. The distance metric, often Euclidean distance, determines the proximity between data points.

In the context of healthcare applications, k-NN has demonstrated effectiveness in various scenarios. For instance, researchers [57] have applied k-NN for pre- dicting patient outcomes, leveraging electronic health records (EHRs) data. The algorithm considers the similarity between a patient's features and those of its k-nearest neighbours to predict potential healthoutcomes.

The simplicity of k-NN makes it particularly suitable for healthcare tasks where interpretability is crucial. Healthcare professionals can easily understand andtrust the predictions made by k-NN, as it aligns with their intuitive understanding of patient similarities and outcomes.

As healthcare datasets grow in complexity and size, the ability of k-NN to provide reliable predictions based on local neighbourhoods becomes increasingly valuable. Its adaptability to different data types and straightforward implementation contribute to its widespread use in healthcare analytics.

E. Gradient Boosting

Gradient Boosting is an ensemble learning method that combines the predictive power of multiple weak learners, typically decision trees, to create a ro- bust and accurate model. It operates by sequentially adding trees to correct the errors of the precedingones. The boosting process assigns more weight to instances that were misclassified, emphasizing their significance in subsequent iterations. In healthcare applications, Gradient Boosting has proven effective in tasks such as disease prediction and risk assessment. Researchers [58][59] have uti- lized Gradient Boosting to develop models for pre- dicting the progression of chronic diseases based on longitudinal patient data.

The strength of Gradient Boosting lies in its ability to handle complex relationships in data, making it suit- able for healthcare datasets with intricate patterns. Moreover, its capacity to deal with missing values and outliers enhances its applicability in real-world healthcare scenarios. As a part of the ensemble learning family, Gradient Boosting complements other algorithms like Random Forest. Its iterative nature and emphasis on correcting errors make it a powerful tool for building accurate predictive models in healthcare analytics.

2) **Unsupervised Algorithm:**

A. Association Analysis

Association analysis is a widely utilized technique in the realms of data mining and machine learning, particularly for prediction purposes, as it has the capability to extract concealed and pertinent infor- mation from extensive datasets [53]. This method generates a set of association rules for dataset items [56]. An association rule takes the form of X → Y, where X and Y are disjoint item sets (i.e., X Y = Ø). This implies that the presence of items in set X in current transactions may lead to the appearance of one or more items in set Y in future transactions. Consequently, association analysis has found exten- sive application in market basket data to predict retail sales behaviour, where each item reflects a customer's purchase [53]. In a medical context, if items represent diseases and the item set is defined as a patient's set of conditions up to the present moment, this method can be employed to predict future disease risk.

The Hierarchical Association Rule Model (HARM), as introduced in [60], utilizes association analysis and Bayesian estimation to predict illness risk from medical data. In this modelling technique, a set of association rules is established through association analysis methods, and these rules are subsequently ranked using Bayesian estimation. HARM has the capability to anticipate a patient's potential future medical issues based on their past and current history of reported ailments, assuming regular consultations with healthcare professionals. This approach demonstrates the adaptability of association analysis in extracting meaningful patterns from diverse datasets, making it a valuable tool in predictive modelling within the medical domain.

B. Hierarchical Clustering

Hierarchical Clustering, an unsupervised algorithm in data analysis, groups similar data points into clusters, forming a dendrogram with leaves representing individual data points. In healthcare, it proves valuable, as shown by researchers [61] categorizing patients based on EHRs to identify patterns and contribute to personalized medicine. The flexibility of hierarchical clustering extends to various healthcare data types, making it versatile for exploring complex datasets and revealing hidden structures. As healthcare datasets grow in size and complexity, hierarchical clustering becomes crucial for uncovering intricate patterns within patient populations. Its unsupervised nature allows exploration without predefined labels, making it instrumental in revealing hidden insights in complex medical datasets.

3) **Reinforcement Learning:** Reinforcement Learning (RL) stands as a dynamic and adaptive paradigm within machine learning, where an agent learns through interac- tion with an environment. This learning process involves the agent taking actions and receiving feedback, typically in the form of rewards or penalties, with the goal of maximizing cumulative rewards over time. In the health- care domain, RL finds versatile applications in optimiz-ing decision-making processes, treatment strategies, and resource allocation.

Applications in Healthcare:

1) Treatment Optimization
RL fine-tunes treatment plans by dynamically adapt-ing and personalizing strategies based on patient re- sponses. Electronic health records (EHRs) are utilized for treatment optimization, incorporating a value- decomposition method for improved results [62].
2) Clinical Trial Design
RL significantly contributes to designing efficient and adaptive clinical trials. Reinforcement learning is em- ployed for learning dosing policies, utilizing action- derived rewards to tailor dosing regimens in trials, including those with TMZ and PCV treatments [63].
3) Resource Allocation
RL optimizes resource allocation in hospital

settings, enhancing operational efficiency and patient out- comes. A decentralized resource allocation manage- ment model for Peer-to-Peer (P2P) communications employs multi-agent deep reinforcement learning to address delay constraints [64].

4) Chronic Disease Management
RL plays a pivotal role in the dynamic management of chronic diseases. The algorithm adapts treatment plans over time, taking into account patient progress and responses. This dynamic approach is particularly beneficial for conditions with a gradual onset. [65]

5) Surgical decision-making
RL is valuable in surgical decision-making, address- ing challenges in formulating reward functions and defining patient states. Personalized reinforcement learning models can enhance surgical care, with con- siderations for privacy-focused alternatives, including on-device machine learning and federated learning. Adopting a partially observable Markov Decision Pro-cess (MDP) approach may offer superior modeling of patient states. Challenges persist in ensuring data quality and consistency in electronic health records, emphasizing the importance of understanding the impact of noise and bias on patient states [66].

Implementing RL in healthcare introduces challenges related to safety, ethics, and the interpretability of decision-making processes. Addressing these challenges requires careful design and validation of RL algorithms to ensure their effectiveness, reliability, and alignment with ethical considerations in healthcare. The adaptability and learning capabilities of RL hold promise for transformative advancements in healthcare decision-making and personalized interventions.

4) **Network Methods:** A network is a graphical rep- resentation composed of nodes (vertices or actors) and edges (ties or links), where edges denote relationships and nodes represent entities. Various scientific problems, including disease prediction in healthcare, can be modeled as graphs, and the literature presents numerous graph theory approaches and algorithms for analysis. Conven- tional statistical and data mining methods often overlookthe links between diseases and symptoms when predict-ing disease risk from healthcare data. Chronic and non-communicable diseases typically share risk factors, whethergenetic, environmental, or behavioral, complicating indi- vidualized study predictions. A network-based approach,encompassing graph theory and statistical methods, is more suitable in such scenarios. Social Network Analysis (SNA) is a comparable approach rooted in network and graph theories, examining relationship patterns among entities in a dataset. In a healthcare context, SNA proves valuable when analyzing interactions among clinicians, pa- tients, pharmacists, nurses, and medical technicians. Each entity in healthcare data is represented as a node in a social network, with edges denoting relationships. SNA hasbeen applied to study physician-patient partnerships and collaborations within hospital networks. Uddin et al. [67] introduced an SNA framework for analyzing physician col- laboration and hospital coordination, emphasizing patient- centric care, hospital-rehab coordination, and physician collaboration networks. In obesity research, SNA is em- ployed to understand research trends and map knowledge structures. Large-scale studies exploring comorbidities and predicting chronic diseases using electronic health data hold significant potential. [68] proposed a novel strategy utilizing graph theory and SNA to analyze chronic disease progression, constructing a baseline network based on hospital records. Expanding this approach to study co- morbidities related to type 2 diabetes, they introduced the concept of a ’comorbidity network’ for predicting chronic illness risk [69], [70].

IV. Handling EHR using Blockchain and AI

Machine learning holds the potential to optimize health- care systems and enhance the delivery of intelligent ser- vices. However, ensuring the secure storage, exchange, and training of sensitive datasets poses a significant challenge for practical machine learning implementations. The inte- gration of machine learning and blockchain technologies is emerging as a solution to enhance the security and privacyof datasets [71], [72].

Federated learning, a machine learning technique con- ducted across multiple computing nodes while safeguard- ing the confidentiality of sensitive data during sharing, is a promising approach. Through the exchange of encrypted datasets, different medical organizations can collaborate to develop highly accurate prediction models. Blockchain, serving as a regulator, can record associated training trans- actions in an immutable and transparent manner, fostering accountability and reliable cooperation. This approach en- courages medical organizations and researchers to share encrypted datasets, thereby advancing medical treatment and public health.

Blockchain acts as a robust foundation for machine learning algorithms, ensuring the security of data input. The sharing of large datasets across various applications and domains, a challenge highlighted by [73], is addressed through homomorphic encryption. Although homomor- phic encryption incurs significant computational costs, future advancements may enable the encryption of sensi- tive data without compromising the efficiency of machine learning for intelligent services.

Blockchain can also be leveraged to store rollback models in instances of high rates of erroneous predictions. Pointers to essential data of retrained models are securely and immutably stored on the blockchain. In the context of erratic arrhythmia alarm rates, [74] suggests that retraining models indexed by pointers on the blockchain can enhance accuracies for continuous remote systems. Additionally, AI can automate the creation of smart contracts, enhancing processes' security and adaptability.

The intersection of blockchain and healthcare has gar- nered increasing interest in both academic research and industry, with startups exploring innovative applications [75], [76], [77], [78], [79], [80]. While [81] claims ownership of the first attempt to utilize blockchain for an Electronic Health Record (EHR) management system, the presented system remains conceptual and awaits deployment and testing. MedRec, a decentralized EHR management frame- work, was developed by [75], utilizing blockchain technol- ogy for security and ease of transfer. [79], on the other hand, implemented a closed, access-controlled blockchain EHR system, incorporating cloud storage and access key transfers for encryptions. The synergy of AI and blockchain in the context of EHR management systems is an evolving area, as seen in recent literature [80], [81]. The advantages of combining AI with both permission less and permissioned prototypes of blockchain EHR systems are yet to be fully realized.

[82] discusses the fundamental role of blockchain tech- nology in monitoring health records, emphasizing its po- tential for meaningful results in drug tracking, treatment effectiveness, safe patient management, and improved clin- ical outcomes when combined with healthcare data and big data. Conversely, [27] focuses on critical aspects of an EHR, such as comprehensive reporting, quality assurance, monitoring health-related expenses, billing details, and confidentiality. It also highlights the inefficiencies of cur- rent systems, characterized as slow, inflexible, and insecure. Furthermore, [83] underscores the importance of patient records availability and the challenges in accessing vital as- pects of a patient's medical history due to time constraints and lack of expertise among healthcare professionals.

V. Dimensions of Blockchain-based EHR Management: Privacy, Security, Accessibility, and Storability

A. Privacy

The principle of privacy involves an individual's enti- tlement to control the access, modification, and sharing of their electronic health records (EHRs) [84]. Concerns arise from the potential misuse of EHRs by healthcare providers, either intentional or unintentional, resulting in privacy violations [85]. A study [86] reveals widespread patient apprehension regarding the security of their EHRs, with around half of the respondents [87] expressing con- cerns that sharing health data might compromise personal information. Consequently, privacy stands as a critical factor in evaluating blockchain-based solutions that claim to enhance EHR privacy.

B. Security

EHR security pertains to the degree to which an individ- ual's electronic health records are confined to authorized individuals. Approximately half of patients express worries about EHR security, particularly as these records traverse the internet [88]. Interestingly, doctors prioritize EHR se- curity more than patients [89], and some physicians still favour paper records, perceiving them as safer than digital alternatives. Yet, the transition to digital records increases vulnerability to security breaches compared to traditional paper-based systems [90]. Addressing security concerns, Liu et al. [91] emphasize the necessity of in-depth studies on securing EHRs, underscoring the importance of robust security measures.

C. Accessibility

Efficiently controlling and managing access to sensi- tive data is a fundamental aspect of accessibility. Com- mon strategies for healthcare systems include role-based, attribute-based, and identity-based access control [92].

Given the sensitivity of patient health data in EHRs, im- plementing effective access management is imperative.

D. Storability

The scalability challenge of blockchain technology has evolved over the years. Bitcoin's initial block had a mere 1MB storage limit, but with the blockchain's growing popularity, participants, and blocks, scalability became a concern [93]. Blockchain applications address scalability through on-chain and off-chain solutions. On-chain storage involves directly storing all user-uploaded data on the blockchain, requiring substantial storage capacity. In con- trast, off-chain storage keeps the actual data elsewhere but links it to the main chain, albeit with less robust security. Consideration for storing EHRs on-chain must weigh the need for substantial storage capacity against the imperative to ensure data safety and security [93].

VI. Blockchain Implementation for Privacy and Anonymity in Healthcare Records

The degree of privacy and anonymity in healthcare records can be tailored based on the type of blockchain implementation, whether it's public, private, or licensed [94]. For instance, Corda enhances transaction privacy by validating only through participating individuals, em- phasizing a more controlled and secure environment [95][96]. In the context of Industry 4.0, the Secure Mu- tual Authentication System (BSeIn) employs blockchain to ensure privacy and security, demonstrating scalability through Smart Contracts [97]. Anonymization techniques, like pseudonymization, are introduced to protect identities, acknowledging that blockchain may not ensure absolute anonymity [98][99][100]. The application of cryptography is crucial in maintaining participant anonymity, offering varying levels of privacy based on cryptographic methods [115]. To address the challenges posed by public exposure of health data, approved blockchain technology and out-of- chain solutions are suggested, striking a balance between security and practicality [100][101].

The need for confidence and privacy extends beyond healthcare, as seen in the Blockchain-Based Anonymous Reputation System (BARS) proposed for Vehicular Ad Hoc Networks (VANETs) [154]. This system, designed to pro- tect the privacy of VANET users, could be adapted for Electronic Health Records (EHRs) by leveraging similar features. Additionally, the broadcast encryption and multi- receiver encryption used in the BSeIn approach ensure secure communication, providing a model for privacy in transactions [97].

A. Privacy by Design for Healthcare Records

Implementing Privacy by Design (PbD) is crucial in ensuring inherent data protection and privacy features in healthcare systems [102]. PbD, introduced by Ann Cavoukian in the mid-1990s, has become an international privacy standard, incorporated into regulations like the GDPR [103][104][105][106]. Its incorporation into the U.S.Commercial Privacy Bill of Rights Act underscores its global acceptance [107]. EHR systems can significantly benefit from PbD to maintain robust privacy protection in the face of evolving information technology.

B. Strategies for Privacy by Design

1) **Data-Driven Approaches:** Keep it Simple: Minimizing the processing of personal data ensures the least impact on privacy, emphasizing the avoidance of unnecessary data collection [108].

2) **Design Patterns:** Approaches like "choose before you collect" and the use of pseudonyms and anonymization are essential design patterns aligned with the strategy of simplicity [109][110].

3) **Hide:** Concealing personal information prevents var- ious abuses, with design patterns like data encryption contributing to maintaining privacy [111][110].

4) **Design Patterns:** Encryption strategies play a crucial role within the "hide" strategy, providing security by con- verting data to an unreadable format [110].

5) **Separate:** Storing personal data in distinct partitions and segregating the storage and processing ensure privacy protection, demanding a distributed processing solution [112].

6) **Aggregate:** Managing personal data with minimal details at a higher aggregate level reduces sensitivity, em- ploying strategies like the "Granularity of Location" and "K-anonymity" design patterns [113][114][115].

VII. Navigating Challenges in Implementing AI-

basedBlockchain Technology in Healthcare

Integrating AI-driven blockchain technology into the healthcare landscape presents a transformative potential, yet it encounters several substantial challenges. The fol- lowing points shed light on critical considerations:

A. Technological Challenges:

- Scalability
  The extensive computational power required by both blockchain and AI models faces a challenge when deal-ing with the vast volumes of healthcare data. Achievingscalability without compromising performance or se- curity is a formidable task.
- Interoperability
  Healthcare data fragmentation across various institu- tions demands seamless interoperability between sys- tems and blockchain platforms. Standardization and collaborative efforts are imperative for success.
- Data Privacy and Security
  Striking a balance between the transparency and ac- cessibility features of blockchain and the paramount importance of patient privacy demands robust encryp-tion and access control mechanisms.

B. Regulatory and Ethical Challenges:

- Regulatory Uncertainty
  Evolving regulations regarding data ownership, access, and governance on blockchain in healthcare create uncertainties that may impede widespread adoption.
- Ethical Considerations
  The ethical dimensions surrounding algorithmic bias, fairness, and explainability of AI models used in healthcare decisions necessitate careful attention to ensure responsible and ethical utilization.
- Consent and Trust
  Garnering patient trust and informed consent for utilizing their data on blockchain platforms calls for transparent communication about potential risks and benefits.

C. Other Challenges

- Cost and Technical Expertise
  The implementation and maintenance of these sophis- ticated technologies come with a significant financial burden and require specialized technical expertise, which might not be universally available in healthcaresettings.
- Change Management and Workforce Training
  Effectively integrating these technologies entails sub-stantial workflow changes and demands comprehen-sive training for healthcare professionals regardingtheir use and implications.
- Limited Awareness and Understanding
  Widening acceptance of AI-based blockchain technol- ogy in healthcare hinges on raising awareness and edu-cating stakeholders about both its potential advantages and challenges.

VIII. Conclusion

This study explores how blockchain technology can revo- lutionize electronic health records (EHRs) management and healthcare delivery, highlighting benefits such as enhanced security and patient empowerment. By decentralizing data storage and integrating with AI and IoT, blockchain offers transparent and secure EHR management. Despite chal- lenges like interoperability and scalability, ongoing research holds tremendous potential for the future of healthcare. Blockchain has the capacity to transform EHR administra- tion, promote patient autonomy, and drive innovation in healthcare. Addressing challenges requires further investi- gation and collaboration to fully realize its revolutionary potential.

References

1) A. Kumari, S. Tanwar, S. Tyagi, and N. Kumar, "Fog computing for 2242 healthcare 4.0 environment: Op- portunities and challenges," Comput. Electr. Eng., vol. 72, pp. 1–13, Nov. 2018.
2) P. Campanella, E. Lovato, C. Marone, L. Fallacara, A. Mancuso, W. Ricciardi, and M. L. Specchia, "The impact of electronic health records on healthcare quality: A systematic review and meta-analysis," Eur. J. Public Health, vol. 26, no. 1, pp. 60–64, Feb. 2016.
3) S. Tanwar, K. Parekh, and R. Evans, "Blockchain-based electronic health- care record system for healthcare 4.0 applications," J. Inf. Secur. Appl., vol. 50, Feb. 2020, Art. no. 102407.
4) P. Tagde, S. Tagde, T. Bhattacharya, P. Tagde, H. Chopra, R. Akter, D. Kaushik, and M. Rahman, "Blockchain and artificial intelligence technology in e-health," Environ. Sci. Pollut. Res., vol. 28, no. 38, pp. 52810–52831, Oct. 2021.
5) A. A. Siyal, A. Z. Junejo, M. Zawish, K. Ahmed, A. Khalil, and G. Soursou, "Applications of blockchain technology in medicine and healthcare: Challenges and future perspectives," Cryptography, vol. 3, no. 1,p. 3, Jan.

2019.

6) A. Shahnaz, U. Qamar, and A. Khalid, "Using blockchain for electronic health records," IEEE Ac- cess, vol. 7, pp. 147782–147795, 2019.
7) S. Tanwar, K. Parekh, and R. Evans, "Blockchain-based electronic health-care record system for healthcare 4.0 applications," J. Inf. Secur. Appl., vol. 50, Feb. 2020, Art. no. 102407.
8) J. Vora et al., "Ensuring privacy and security in e-health records," in Proc. Int. Conf. Comput. Inf. Telecommun. Syst. (CITS), Jul. 2018, pp. 1–5.
9) T. Kubo, A. Yanasan, T. Herbosa, N. Buddh, F. Fer- nando, and R. Kayano, "Health data collection before, during and after emergencies and disasters—The re- sult of the Kobe expert meeting," Int. J. Environ. Res. Public Health, vol. 16, no. 5, p. 893, Mar. 2019.
10) J. M. Puaschunder, "The potential for artificial intelli- gence in health-care," Future Healthcare J., vol. 6, no. 2, p. 94, 2019.
11) M. Mettler, "Blockchain technology in healthcare: The revolution starts here," in Proc. IEEE 18th Int. Conf. e-Health Netw., Appl. Services (Healthcom), Munich, Germany, Sep. 2016, pp. 1–3.
12) Z. Zheng, S. Xie, H. Dai, X. Chen, and H. Wang, "An overview of blockchain technology: Architecture, consensus, and future trends," in Proc. IEEE Int. Congr. Big Data (BigData Congr.), Honolulu, HI, USA, Jun. 2017, pp. 557–564.
13) Q. Feng, D. He, S. Zeadally, M. K. Khan, and N. Kumar, "A survey on privacy protection in blockchain system," J. Netw. Comput. Appl., vol. 126, pp. 45–58, Jan. 2019.
14) W. Wang, D. T. Hoang, P. Hu, Z. Xiong, D. Niyato, P. Wang, Y. Wen, and D. I. Kim, "A survey on consen- sus mechanisms and mining strategy management in blockchain networks," IEEE Access, vol. 7, pp. 22328–22370, 2019.
15) X. Li, P. Jiang, T. Chen, X. Luo, and Q. Wen, "A survey on the security of blockchain systems," Future Gener. Comput. Syst., vol. 107, pp. 841–853, Jun. 2020.
16) N. Nakamoto. Centralised Bitcoin: A Secure and High Performance Electronic Cash System. [On- line]. Available: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3065723
17) G. Wood. (Dec. 2021). Ethereum: A Secure Decentralised Generalised Transaction Ledger.[Online]. Available: https://files.gitter.im/ethereum/yellowpaper/VIyt/Paper.pdf
18) I. Bentov, C. Lee, A. Mizrahi. (Jan. 2022). Proof of Activity: Extending Bitcoin's Proof of Work Via Proofof Stake Y. [Online]. Available: https://dl.acm.org/doi/abs/10.1145/2695533.2695545
19) C. Li and B. Palanisamy, "Incentivized blockchain- based social media platforms: A case study of Steemit," in Proc. 10th ACM Conf. Web Sci., Jun. 2019,pp. 145–154.
20) L. Lamport, R. Shostak, and M. Pease, "The Byzantine generals problem," ACM Trans. Program. Lang. Syst., vol. 4, no. 3, pp. 382–401, Jul. 1982.
21) F. Tschorsch. (Jan. 2022). Bitcoin and Beyond:A Technical Survey on Decentralized Digital Cur- rencies. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/7423672/
22) C. Villani, For a Meaningful Artificial Intelligence: Towards a French and European Strategy (AI for Humanity). 2018.
23) T. F. Xue, Q.-C. Fu, C. Wang, and X.-Y. Wang, "A medical data sharing model via blockchain," Acta Automatica Sinica, vol. 43, no. 9, pp. 1555–1562, 2017
24) Q. Xia, E. B. Sifah, A. Smahi, S. Amofa, and X. S. Zhang, "BBDS: Blockchain-based data sharing for electronic medical records in cloud environments," Information, vol. 8, no. 2, p. 44, 2017.
25) J. Gareth, W. Daniela, H. Trevor, and T. Robert. (Sep. 22, 2013). An Introduction to Statistical Learning: With Applications in R. [Online]. Available: http://dspace.agu.edu.vn:8080/handle/AGU_Library/13322
26) A. Tekkeşin, "Artificial intelligence in healthcare: Past, present and future," Anatolian J. Cardiol., vol. 2, no. 4, pp. 230–243, 2019. [Online]. Available: https://svn.bmj.com/content/2/4/230
27) Y. LeCun, Y. Bengio, and G. Hinton, "Deep learn- ing," Nature, vol. 521, no. 7553, pp. 436–644, 2015. [Online]. Available: https://www.nature.com/articles/nature14539
28) M. Resta, M. Sonnessa, E. Tànfani, and A. Testi, "Un- supervised neural networks for clustering emergent patient flows," Oper. Res. Health Care, vol. 18, pp. 41–51, Sep. 2018.
29) W. Raghupathi and V. Raghupathi, "Big data ana- lytics in healthcare: Promise and potential," Health Inf. Sci. Syst., vol. 2, no. 1, pp. 1–10, Dec. 2014. [Online]. Available: https://link.springer.com/article/10.1186/2047-2501-2-3
30) A. L. Beam and I. S. Kohane, "Big data and machine learning in health care," JAMA, vol.319, no. 13, p. 1317, Apr. 2018. [Online]. Avail- able: https://jamanetwork.com/journals/jama/fullarticle/2675024
31) J. Kaye, E. A. Whitley, D. Lund, M. Morrison, H. Teare, and K. Melham, "Dynamic consent: A patient interface for twenty-first century research networks," Eur. J. Hum. Genet., vol. 23, no. 2, pp. 141–146, Feb. 2015.
32) D. V. Dimitrov, "Blockchain applications for health- care data management," Health Inform. Res., vol. 25, no. 1, pp. 6–51, 2019.
33) L. Rundo, R. Pirrone, S. Vitabile, E. Sala, and O. Gam- bino, "Recent advances of HCI in decision-making tasks for optimized clinical work- flows and precision medicine," J. Biomed. Informat., vol. 108, Aug. 2020, Art. no. 103479.
34) S. Gehrmann, F. Dernoncourt, Y. Li, E. T. Carlson, J. T. Wu, J. Welt, J. Foote, E. T. Moseley, D. W. Grant, P. D. Tyler, and L. A. Celi, "Comparing deep learning and concept extraction-based methods for patient

phenotyping from clinical narratives," PLoS ONE, vol. 13, no. 2, Feb. 2018, Art. no. e0192360.

35) P. Rajpurkar, J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C. Langlotz, K. Shpanskaya, M. P. Lungren, and A. Y. Ng, "CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning," 2017, arXiv:1711.05225.

36) B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba. (Sep. 22, 2020). Learning Deep Features for Discriminative Localization. [Online]. Available: http://cnnlocalization.csail.mit.edu

37) E. Ford, J. A. Carroll, H. E. Smith, D. Scott, and J. A. Cassell, "Extracting information from the text of electronic medical records to improve case detection: A systematic review," J. Amer. Med. Inform. Assoc., vol. 23, no. 5, 1007–1015, Sep. 2016. [Online]. Available: https://academic.oup.com/jamia/article/23/5/1007/2379833

38) C. Shivade, P. Raghavan, E. Fosler-Lussier, P. J. Embi, N. Elhadad, S.B. Johnson, and A. M. Lai, "A review of approaches to identifying patient phenotype cohorts using electronic health records," J. Amer. Med. Inform. Assoc., vol. 21, no. 2, pp. 221–230, Mar. 2014. [Online].

39) D. S. Carrell, R. E. Schoen, D. A. Leffler, M. Morris, S. Rose, A. Baer, S. D. Crockett, R. A. Gourevitch, K. M. Dean, and A. Mehrotra, "Challenges in adapting existing clinical natural language processing systems to multiple, diverse health care settings," J. Amer. Med. Inform. Assoc., vol. 24, no. 5, pp. 986–991, Sep. 2021. [Online]. Available: https://academic.oup.com/jamia/article/24/5/986/3737804

40) J. C. Kirby, P. Speltz, L. V. Rasmussen, M. Basford, O. Gottesman, P. L. Peissig, J. A. Pacheco, G. Tromp, J. Pathak, D. S. Carrell, and S. B. Ellis, "PheKB: A catalog and workflow for creating electronic phenotype algorithms for transportability," J. Amer. Med. Inform. Assoc., vol. 23, no. 6, pp. 1046–1052, 2016. [Online]. Available: https://academic.oup.com/jamia/article/23/6/1046/2399228

41) K. R. Foster, R. Koprowski, and J. D. Skufca, "Machine learning, medical diagnosis, and biomedical engineering research–commentary," Biomed. Eng. OnLine, vol. 13, no. 1, p. 94, 2014. [Online]. Available: https://link.springer.com/articles/10.1186/1475-925X-13-94

42) A. Asperti and C. Mastronardo, "The effectiveness of data augmentation for detection of gastrointestinal diseases from endoscopical images," in Proc. 11thInt. J. Conf. Biomed. Eng. Syst. Technol., 2017, pp. 199–205.

43) S. C. Wong, A. Gatt, V. Stamatescu, and M. D. McDonnell, "Understanding data augmentation for classification: When to warp?" in Proc. Int. Conf. Digit. Image Comput. Tech. Appl., Dec. 2016.

44) R. Kemp and V. Prasad, "Surrogate endpoints in oncology: When are they acceptable for regulatory and clinical decisions, and are they currently overused?" BMC Med., vol. 15, no. 1, pp. 1–7, Jul. 2017. [Online]. Available: https://bmcmedicine.biomedcentral.com/articles/10.1186/s12916-017-0902-9

45) R. Lorbieski and S. M. Nassar, "Performance assess- ment of multi-level ensemble for multi-class prob- lems," Int. J. Comput. Inf. Eng., vol. 12, no. 3, pp.149–155, 2018.

46) I. Kavakiotis, O. Tsave, A. Salifoglou, N. Maglaveras, I. Vlahavas, and I. Chouvarda, "Machine learning and data mining methods in diabetes research," Comput. Struct. Biotechnol. J., vol. 15, pp. 104–116, Jan. 2017.

47) W. Mcculloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," Bull. Math. Biol., vol. 52, nos. 1–2, pp. 99–115, 1990. [Online]. Available: https://link.springer.com/article/10.1007/BF02478259

48) D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating er- rors," Nature, vol. 323, no. 6088, pp. 533–536, Oct. 1986. [Online]. Available: https://www.nature.com/articles/323533a0

49) J. Schmidhuber, "Deep learning in neural networks: An overview," Neural Netw., vol. 61, pp. 85–117, Jan. 2015.

50) S. Liu, S. Liu, W. Cai, S. Pujol, R. Kikinis, and D. Feng, "Early diagnosis of Alzheimer's disease with deep learning," in Proc. IEEE 11th Int. Symp. Biomed. Imag., Jul. 2014, pp. 1015–1018.

51) Y. Cheng, F. Wang, P. Zhang, and J. Hu, "Risk prediction with electronic health records: A deep learning approach," in Proc. SIAM Int. Conf. Data Mining, 2016, pp. 432–440. [Online]. Available: https://epubs.siam.org/page/terms

52) J. Zhang, J. Gong, and L. Barnes, "HCNN: Heterogeneous convolutional neural networks for comorbid risk prediction with electronic health records," in Proc. IEEE/ACM Int. Conf. Connected Health, Appl., Syst. Eng. Technol. (CHASE), Jul. 2017, pp. 214–221.

53) M. E. Hossain, A. Khan, M. A. Moni, and S. Uddin, "Use of electronic health data for disease prediction: A comprehensive literature review," IEEE/ACM Trans. Comput. Biol. Bioinf., vol. 18, no. 2, pp. 745–758, Mar. 2021.

54) Alejandro Barredo Arrieta, Natalia Díaz-Rodríguez , Javier Del Ser, Adrien Bennetot, Siham Tabik , Alberto Barbado , Salvador Garcia , Sergio Gil-Lopez , Daniel Molina , Richard Benjamins , Raja Chatila , Francisco Herrera, "Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI", Available: https://doi.org/10.1016/j.inffus.2019.12.012

55) I. Kavakiotis, O. Tsave, A. Salifoglou, N. Maglaveras, I. Vlahavas, and I. Chouvarda, "Machine learning and data mining methods in diabetes research," Comput. Struct. Biotechnol. J., vol. 15, pp. 104–116, Jan. 2017.

56) S. Rallapalli and T. Suryakanthi, "Predicting the risk of

diabetes in big data electronic health records by using scalable random forest classification algorithm," in Proc. Int. Conf. Adv. Comput. Commun. Eng., pp. 281–284.

57) S. Tayeb et al., "Toward predicting medical conditions using k-nearest neighbors," 2017 IEEE International Conference on Big Data (Big Data), Boston, MA, USA, 2017, pp. 3897-3903,
58) Xing Song; Lemuel R Waitman; Alan SL Yu ; David C Robbins; Yong Hu; Mei Liu; "Longitudinal Risk Prediction of Chronic Kidney Disease in Diabetic Patients Using a Temporal-Enhanced Gradient Boosting Machine: Retrospective Cohort Study",
59) Rufo, D.D.; Debelee, T.G.; Ibenthal, A.; Negera, W.G. Diagnosis of Diabetes Mellitus Using Gradient Boosting Machine (LightGBM). Diagnostics 2021, 11, 1714.
60) T. McCormick, C. Rudin, and D. Madigan, "A hierarchical model for association rule mining of sequential events: An approach to automated medical symptom prediction," 2011.
61) Ogbuabor, G. and Ugwoke, F.N., 2018. Clustering algorithm for a healthcare dataset using silhouette score value. Int. J. Comput. Sci. Inf. Technol, 10(2), pp.27-37.
62) Li, Tianhao, et al. "Electronic health records based reinforcement learning for treatment optimizing." Information Systems 104 (2022): 101878.
63) Yauney, Gregory, and Pratik Shah. "Reinforcement learning with action-derived rewards for chemotherapy and clinical trial dosing regimen selection." Machine Learning for Healthcare Conference. PMLR, 2018.
64) Alelaiwi, Abdulhameed. "Resource Allocation Management in Patient-to-Physician Communications Based on Deep Reinforcement Learning in Smart Healthcare Services." 2020 IEEE International Conference on Multimedia & Expo Workshops (ICMEW). IEEE, 2020.
65) Dhote, S., Roberts, M.K., Sridhar, K.P. (2024). Reinforcement Learning Method for Identifying Health Issues for People with Chronic Diseases. In: Mehta, G., Wickramasinghe, N., Kakkar, D. (eds) Innova-tions in VLSI, Signal Processing and Computational Technologies. WREC 2023. Lecture Notes in Electrical Engineering, vol 1095. Springer, Singapore.
66) Datta, Shounak, et al. "Reinforcement learning in surgery." Surgery 170.1 (2021): 329-332.
67) S. Uddin, A. Khan, and M. Piraveenan, "Administrative claim data to learn about effective healthcare collaboration and coordination through social network," in Proc. 48th Hawaii Int. Conf. Syst. Sci., Jan. 2015, pp. 3105–3114.
68) A. Khan, S. Uddin, and U. Srinivasan, "Adapting graph theory and social network measures on healthcare data-a new framework to understand chronic disease progression," in Proc. Australas. Comput. Sci. Week Multiconf., Feb. 2016, pp. 1–7.
69) A. Khan, S. Uddin, and U. Srinivasan, "Chronic disease prediction using administrative data and graph theory: The case of type 2 diabetes," Expert Syst.Appl., vol. 136, pp. 230–241, Dec. 2019.
70) E. Kang, S. Kim, Y. E. Rhee, J. Lee, and Y. H. Yun, "Self-management strategies and comorbidi- ties in chronic disease patients: Associations with quality of life and depression," Psychol., Health Med., vol. 26, no. 8, pp. 1031–1043, Sep. 2020. [Online]. Available: https://www.tandfonline.com/doi/abs/10.1080/13548506.2020.1838585
71) X. Zheng et al., "Blockchain-based personal health data sharing system using cloud storage", Proc. IEEE 20th Int. Conf. Healthcom, pp. 1-6, 2018.
72) S. H. Lee and C. S. Yang, "Fingernail analy-sis management system using microscopy sensor and blockchain technology", Int. J. Distrib. Sensor Netw., vol. 14, no. 3, Mar. 2018, [online] Avail- able: https://journals.sagepub.com/doi/full/10.1177/1550147718767044.
73) S. Yaji, K. Bangera and B. Neelima, "Privacy preserving in blockchain based on partial homomorphic encryption system for AI applications", Proc. IEEE 25th Int. Conf. High Perform. Comput. Workshops (HiPCW), pp. 81-85, Dec. 2018.
74) A. Juneja and M. Marefat, "Leveraging blockchain for retraining deep learning architecture in patient-specific arrhythmia classification", Proc. IEEE Int. Conf. Biomed. Health Informat. (BHI), pp. 393-397, Mar. 2018.
75) A. Ekblaw, A. Azaria, J. D. Halamka, A. Lippman and T. Vieira, "A case study for blockchain in healthcare: 'MedRec' prototype for electronic health records andmedical research data", Proc. IEEE Open Big Data Conf., pp. 13, Aug. 2016.
76) X. Yue, H. Wang, D. Jin, M. Li and W. Jiang, "Healthcare data gateways: Found healthcare intelligence on blockchain with novel privacy risk control", J. Med. Syst., vol. 40, no. 10, pp. 1-8, 2016, [online] Available: https://link.springer.com/article/10.1007/s10916-016-0574-6.
77) GEM [Internet], Sep. 2021, [online] Available: https://gem.co/health/.
78) P. Beninger and M. A. Ibara, "Pharmacovigilance and biomedical informatics: A model for future development", Clin. Therapeutics, vol. 38, no. 12, pp. 2514-2525, Dec. 2016.
79) A. Dubovitskaya, Z. Xu, S. Ryu, M. Schumacher and F. Wang, "Secure and trustable electronic medical records sharing using blockchain", Proc. AMIA Annu. Symp., pp. 650, Sep. 2017, [online] Available: https://pmc/articles/PMC5977675/.
80) D. Randall, P. Goel and R. Abujamra, "Blockchain applications and use cases in health information technology", J. Health Med. Inform., vol. 8, no. 3, pp.

8-11, 2017.

81) N. Rifi, E. Rachkidi, N. Agoulmine and N. C. Taher, "Towards using blockchain technology for eHealth data access management", Proc. 4th Int. Conf. Adv. Biomed. Eng. (ICABME), pp. 1-4, Oct. 2017.

82) Z. Wang and M. O'Boyle, "Machine learning in compiler optimization", Proc. IEEE, vol. 106, no. 11, pp. 1879-1901, Nov. 2018.

83) A. Khan, S. Uddin and U. Srinivasan, "Comorbidity network for chronic disease: A novel approach to understand type 2 diabetes progression", Int. J. Med. Informat., vol. 115, pp. 1-9, Jul. 2018.

84) I. Keshta and A. Odeh, "Security and privacy of electronic health records: Concerns and challenges", Egyptian Informat. J., vol. 22, no. 2, pp. 177-183, Jul. 2021.

85) M. Cifuentes, M. Davis, D. Fernald, R. Gunn, P. Dickinson and D. J. Cohen, "Electronic health record challenges workarounds and solutions observed in practices integrating behavioral health and primary care", J. Amer. Board Family Med., vol. 28, no. 1, pp. S63-S72, Sep. 2015.

86) K. T. Win, "A review of security of electronic health records", Health Inf. Manage., vol. 34, no. 1, pp. 13-18, Mar. 2005.

87) J. S. Ancker, M. Silver, M. C. Miller and R. Kaushal, "Consumer experience with and attitudes toward health information technology: A nationwide survey", J. Amer. Med. Inform. Assoc., vol. 20, no. 1, pp. 152-156, Jan. 2013.

88) G. Perera, A. Holbrook, L. Thabane, G. Foster and D. J. Willison, "Views on health information sharing and privacy from primary care practices using electronic medical records", Int. J. Med. Informat., vol. 80, no.2, pp. 94-101, Feb. 2011.

89) S. B. Wikina, "What caused the breach? An examination of use of information technology and health data breaches", Perspectives Health Inf. Manage., vol. 11, pp. 1-16, Oct. 2014.

90) C. S. Kruse, B. Smith, H. Vanderlinden and A.Nealand, "Security techniques for the electronic health records", J. Med. Syst., vol. 41, no. 8, pp. 1-9, Aug. 2017.

91) V. Liu, M. A. Musen and T. Chou, "Data breaches of protected health information in the United States", Jama, vol. 313, no. 14, pp. 1471-1473, 2015.

92) B. Yüksel, A. Küpçü and Ö. Özkasap, "Research issues for privacy and security of electronic health services", Future Gener. Comput. Syst., vol. 68, pp. 1-13, Mar. 2017.

93) S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system. Decentralized bus", Rev., vol. 2008, Jun. 2008.

94) M. Tzanou, The Fundamental Right to Data Protec- tion: Normative Value in the Context of Counter Ter- rorism Surveillance, London, U.K.:Bloomsbury, 2017.

95) A. Ahmed, M. M. R. Parvez, M. H. Hasan, F. N. Nur, N. N. Moon, A. Karim, et al., "An intelligent and secured tracking system for monitoring school bus", Proc. Int. Conf. Comput. Commun. Informat. (ICCCI), pp. 23-25, Jan. 2019.

96) J. H. Hoepman, "Privacy design strategies", Proc. 29th IFIP Int. Inf. Conf., pp. 2-4, Jun. 2014.

97) M. van Lieshout, L. Kool, B. van Schoonhoven and M. de Jonge, "Privacy by design: An alternative to existing practice in safeguarding privacy", Info, vol. 13, no. 6, pp. 55-68, Sep. 2011.

98) L. Sweeney, "Privacy-preserving surveillance using selective revelation", IEEE Intell. Syst., vol. 1, pp. 83-84, 2005.

99) L. Lyu, K. Nandakumar, B. Rubinstein, J. Jin, J. Bedo and M. Palaniswami, "PPFA: Privacy preserving fog- enabled aggregation in smart grid", IEEE Trans. Ind. Informat., vol. 14, no. 8, pp. 3733-3744, Aug. 2018.

100) G. Skinner, E. Chang, M. Mcmahon, J. Aisbett and M. Miller, "Shield privacy hippocratic security method for virtual community", Proc. 30th Annu. Conf. IEEE Ind. Electron. Soc., pp. 2-6, Nov. 2004.

101) A. Gkoulalas-Divanis and G. Loukides, Medical Data Privacy Handbook, 2015.

102) G. Iachello and J. Hong, "End-user privacy in human-computer interaction", Found. Trends Hum.-Comput. Interact., vol. 1, no. 1, pp. 1-137, 2007.

103) Privacy by Design: Effective Privacy Managementin E Victorian Public Sector, Dec. 2019, [online] Available: https://ovic.vic.gov.au/resource/privacy-by-designeffective-privacy-management-in-the-victorian-public-sector/.

104) A. Cavoukian, "Privacy by design [leading edge]", IEEE Technol. Soc. Mag., vol. 31, no. 4, pp. 18-19, Dec. 2021.A. Cavoukian and D. Tapscott, Who Knows: Safe- guarding Your Privacy in a Networked World, New York, NY, USA:McGraw-Hill, 1996.

105) J. Spitzer, 6.1M Healthcare Data Breach Victims in 2018: 5 of the Biggest Breaches So Far, 2019.

106) C. Donnelly, The GDPR: Why You Need to Adopt the Principles of Privacy by Design, Dec. 2019, [online] Available: https://www.itgovernance.eu/blog/en/the-gdpr-why-you-need-to-adoptthe-\principles-of-privacy-by-design.

107) S. Gürses, C. Troncoso and C. Diaz, "Engineering privacy by design", Comput. Privacy Data Protection, vol. 14, pp. 25, Jan. 2011.

108) B. Jacobs, "Select before you collect", Ars Aequi, vol. 54, pp. 1006-1009, Jun. 2005.

109) A. Pfitzmann and M. Hansen, "Anonymity unlinkability undetectability unobservability pseudonymity and identity management—A consolidated proposal for terminology", 2010.

110) ISO/IEC, 29100, "Privacy Framework Technologies 29100", 2011.

111) S. D. Warren and L. D. Brandeis, "The right to privacy", Harv. Law Rev., vol. 5, pp. 193-220, Mar. 1990.

112) J. H. Hoepman, "Privacy design strategies", Proc. IFIP

Int. Inf. Secur. Conf., vol. 428, pp. 446-459, 2014.

113) L. Sweeney, "K-anonymity: A model for protecting privacy", Int. J. Uncertainty Fuzziness Knowl.-Based Syst., vol. 10, no. 5, pp. 557-570, Oct. 2002.

114) Australian Governemnt-Office of the Australian Information Commissioner, Jan. 2019,[online] Available: https://www.oaic.gov.au/privacy/privacy-for-organisations/privacy-by-design/.