Background

Several studies have been conducted that have focused on the use of AI and ML in the field of healthcare. While dealing with patient data and medical records, it is important to uphold the privacy of the customers, in global settings the number of frameworks developed that respect the confidentiality of EHR are limited. The following section would discuss some of these studies that have highlighted frameworks while keeping the patient's privacy under consideration. Moreover, the literature review also aims to highlight some of the other developments that make use of AI for processing patient data.

Background For EHR

The way data is being processed in the field of healthcare is based on the system of EHR for managing the data records. Digitizing the patient’s health information and medical history has now become one of the most important aspects for automating healthcare with the advent of the industrial revolution. With the help of EHR, numerous benefits for processing patient information and enabling safe access to the information are possible. By replacing traditional paper-based records, EHRs facilitate better clinical decision-making, and reduce the number of human errors that may be present in the system. The widespread adoption of automation of the patient record processes also raises concerns of patient privacy as according to the guidelines set by global policies, sharing private data needs to be monitored.

Need for Privacy Protection

EHRs contain sensitive personal health information that must be safeguarded to maintain patient confidentiality, unauthorized access to the system database needs to be always monitored as lack

of consideration may result in an increase in the potential of attack, privacy breaches or data theft. Moreover, the stakeholders that are involved in dealing with confidential patient data need to ensure that privacy is always maintained, and the data sharing is regulated under the clauses of HIPAA and GDPR, if a business owner or a stakeholder dealing with confidential medical data fails to comply with these policies, he may be penalized as the trust of the patient can get violated.

Machine learning techniques are often used to improve the privacy module of EHR, the biggest use of machine learning in healthcare records can be used to de-identify the patient data by removing or encrypting personally identifiable information to preserve the utility of the data for research and analysis. ML algorithms are capable enough to detect and mask sensitive information to minimize the risk of identification of record owner. Moreover, another use case where ML is actively being used is to monitor the access of user records, access control mechanisms can be enabled to analyze user behavior patterns and access patterns to detect anomalies and potential security breaches. ML algorithms can be used to identify the users accessing the resources to enable privileged user management while maintaining the accessibility of the records. There is numerous research that has made use of machine learning to maintain the privacy of healthcare records, the literature review below would present an analysis of some papers in this domain.

In a research conducted by William Briguglio, a model was prepared to address the security issues that were found in the privacy in precision medicine, and handling the sensitive health-related data and EHRs. The importance of privacy preserving was identified and how can machine learning be used to enable predictive model that ensured patient confidentiality and diagnose treatment of the diseases. They introduced generic machine learning with encryption

framework, which serves as the foundation for building ML models that predict cancer using comprehensive genomics datasets. The framework can use a machine learning model to diagnose a disease while ensuring end-to-end encryption. To facilitate validation and replication of their work, the authors provided an open-source repository for the design, implementation, and code for the MLE framework that could be deployed in a cloud server. The research also proposed a four-component system architecture to accommodate encryption scenarios, to meet the privacy-preserving requirements for ML-based precision medicine (Briguglio et al., 2021).

Advancements in Electronic Health Records: Privacy-Preserving Machine Learning and Federated Learning in Healthcare

The use of machine learning in the domain of healthcare has been one of the biggest research areas in recent years. As the compute costs are now becoming cheaper with the coming age, more frameworks are now being developed that involve smart healthcare opportunities, Moreover, the Excessive availability of healthcare records of the patients for analysis offers opportunities for insights into disease diagnosis and healthcare system optimization. The research area may seem promising for improved patient care, but patient data confidentiality is something that needs to be upheld and protected from unauthorized access in order to comply with international standards.

Researchers and healthcare practitioners have used Machine learning models for extracting actionable knowledge from EHRs while addressing privacy concerns. By using algorithms such as federated learning and privacy-preserving machine learning models, researchers can further use the value obtained from EHR data without compromising patient privacy. The following

section would provide an overview of some research that have made use of patient data for improving the healthcare delivery, and the latest practices that ensures privacy of the users while processing the data is maintained along with addressing other challenges that may be present while using healthcare data.

Privacy-Preserving Techniques in Electronic Health Records

EHR systems have transformed the data processing methods for patient data management, as there are several repositories available that can be used for processing patient data, medical research can advance with ease. As healthcare organizations are now relying on EHRs to store and exchange medical information, patient confidentiality needs to be addressed. The following section would provide an overview of research that has highlighted strategies and methods for improving privacy preservation methods with the use of Machine Learning. The review would aim to provide insights on how privacy preservation can be improved for EHRs.

In research presented by Jinsung Yoon, a generative modeling framework called EHR-Safe is proposed to address privacy concerns associated with Electronic Health Records. EHR-Safe makes use of sequential encoder-decoder networks and generative adversarial networks to generate highly realistic synthetic EHR data that closely resembles real-world EHR data. The primary reason for synthesizing the EHR data is to address the key challenges of data privacy including heterogeneity, sparsity, coexistence of numerical and categorical features. By generating synthetic data that accurately captures the complexities of real EHR data, EHR-Safe enables the safe sharing of EHR data for machine learning and data analytics applications while preserving patient privacy. From the Evaluations that were presented in the research, it was observed that EHR-Safe achieves almost-identical fidelity to real data with minimal accuracy

differences at less than 3 percent for the models that were trained using the data when compared with actual data in real environments. This research helps in improving healthcare systems by providing a robust solution for privacy-preserving data sharing to develop healthcare solutions without compromising patient privacy (Yoon et al., 2023).

In another research conducted by Raza Nowrozy et al, a privacy model for Electronic Health Records was proposed that was based on the concepts of privacy ontology and Machine Learning. The study addresses the challenges faced by EHR systems, that includes privacy, accessibility, and legal compliances. To tackle these challenges, a privacy model is developed in this research to efficiently manage and share patients' personal and sensitive data across different platforms, such as MHR and NHS systems. The research employs various BERT techniques to distinguish between legitimate and illegitimate privacy policies with Distil BERT emerging as the most accurate. This paper highlights the potential of ML-based approaches in effectively identifying inadequate privacy policies within EHR systems. The paper also outlines future research directions on how Smart EHR frameworks can be evaluated while upholding the ethical considerations as well. Overall, this research aims to present an approach to enhance healthcare information privacy, by laying the groundwork for future advancements in the field with the help of ML and other new technologies, the privacy of the EHR can be improved to make the data management easier (Nowrozy et al., 2023).

In another research conducted by N. Khalid et al, The privacy of EHR is maintained through the implementation of privacy-preserving techniques in AI based applications. The study highlights the importance of protecting patient privacy in the healthcare sector, especially when handling sensitive data for algorithm development. To address this concern, the researchers explore different privacy-preserving methods, including Federated Learning and Hybrid Techniques, the

paper provided an analysis of existing literature that covered topics from EHR to medical imaging and how the online databases can be anonymized prior to using them for training. The techniques that have been included in the research aimed to improve the healthcare services by upholding the privacy of the patients. The research analyzed the privacy-preserving techniques and highlighted the different research that aligned with the regulatory standards such as HIPAA and GDPR. The research contributed to providing a method for improved adoption of AI in healthcare with a particular focus on protecting the patient data as well (Khalid et al., 2023).

Another framework that is not designed primarily to work with EHR records, but it can be used to improve the privacy model for the data records was proposed by Bita Rouhani called DeepSecure in which machine learning is used followed by secure computation methods such as Yao's Garbled Circuit protocol. DeepSecure enables the execution of state-of-the-art Deep Learning (DL) models in a privacy-preserving setting. This means that neither the cloud servers holding the DL model parameters nor the delegating clients who own the data need to reveal their information. In the context of EHR privacy, DeepSecure could be used to perform processing tasks without compromising the confidentiality of patient data. For instance, healthcare institutions could deploy DeepSecure to securely analyze EHR data for predictive modeling, disease diagnosis, or treatment recommendations. The framework allows for accurate and scalable DL analysis of data generated by distributed clients, ensuring that sensitive information remains private throughout the computation process. By adopting this framework, healthcare providers can maintain compliance with regulations such as HIPAA and GDPR while still using the processed data obtained from machine learning models. The pre-processing techniques introduced in DeepSecure, data projection and DL network distillation can reduce the communication overhead that may be introduced with Deep learning. This optimization not only

improves the computation to be private, but it also minimizes the risk of accidental exposure of sensitive patient information that may be present during the execution of machine learning tasks (Rouhani et al., 2018).

A federated learning framework was introduced that used concepts of differential privacy mechanisms to improve the privacy of healthcare records by Olivia Choudhury et al. in which several challenges linked with the usage of EHR such as privacy concerns, point of failure were highlighted and a method was introduced to mitigate the concerns. Federated learning allows models to be trained locally at different sites without sharing raw data, and differential privacy adds noise to the optimization process to further protect privacy. This approach ensures patient information is kept private while the collaborative model that is used for generating data driven insights is being used. The aggregation of locally trained models to create a global model with improved utility is also possible in the method endorsed in this research. In total 2 separate layers were introduced for improving the privacy of sensitive data, in the first layer, the data was encrypted and anonymized by making it ready for training using the local models, the raw data was kept intact, and it used a DP strategy for protecting the raw and processed data from accidental breaches by adding noise to it prior to sharing. The performance of this DP model in using federated learning resulted in a considerable loss due to the introduction of noise but it can be optimized to add scheduled artifacts for implementing their designed framework in real world settings (Choudhury et al., 2019).

In another research conducted by J Lee et al, a federated learning model was introduced that focused on the privacy of the patients. The research explores methods to construct feature vectors for patients in electronic health records that include heterogeneous data types like nominal, continuous, and time-evolving features. It introduces hashing techniques to transform

patient data into low-dimensional representations, to enable similarity-based comparisons while preserving data privacy. The proposed approach aims to optimize hash codes for patient data while considering pairwise relationships and label information, enabling efficient and privacy-preserving patient similarity learning. By using temporal sequences in patient data, the research further enhances the representation of patient features, allowing the proposed method to predict target diseases while maintaining patient privacy in EHRs. Experiments were conducted on real data to validate the proposed method, comparing it with open and closed systems in terms of disease prediction accuracy. From the analysis it was observed that the method was effective in predicting disease incidence based on multi-hash codes that outperformed both open and closed systems. The temporal sequence that was introduced extracted relevant events from the patient data that enabled the creation of time-decayed vector representations for improved predictive performance (Lee et al., 2018).

Om Kumar et al. conducted research on privacy-preserving healthcare data sharing using Federated Learning. The primary aim of their research was to address the challenges of privacy preservation and communication rounds. In their paper, they presented a framework called FL+DQRE-Scnet, which combines FL with deep Q reinforcement learning and spectral clustering. A Deep Convolutional Neural Network for local data learning in hospitals was designed where patient data is trained locally to obtain weights. These weights are then encrypted using Homomorphic weight encryption to ensure patient privacy during data sharing. To optimize data aggregation and minimize communication rounds, the researchers employed deep Q reinforcement learning with spectral clustering. This allowed for selecting relevant local users based on their request clustering them as per their needs. The clustering that was introduced helped in minimizing the communication between stakeholders which maintained the

confidentiality of EMRs. The methodology discussed in this paper provided a comprehensive analysis of their proposed approach for preserving privacy in healthcare data sharing (Kumar et al., 2023).

In independent research conducted by Tao Hai, a blockchain based methodology was introduced that used federated learning to address the challenges of privacy protection in Electronic Health Records management. The proposed framework can be broken down into two main components, in the first system, a blockchain based storage service is set up using Hyperledger Fabric and Interplanetary File System to securely store EHR data from different healthcare providers. Allowing the records to be immutable, and by using a system of validators, the changes and updates to the database can be tracked. Secondly, collaborative learning techniques, including LightGBM and N-Gram models, are used to analyze the EHR data for generated remedial measures for the patients. The models that were introduced used LLMs and sentiment analysis to extract insights from the patient records and recommend personalized treatment plans. The primary reason for introducing a blockchain network was to prioritize privacy protection in the process is to enable encryption techniques for safeguarding patient data and Federated Deep Learning was introduced to ensure that learning occurs locally at individual hospitals and clinics without the need for centralized data aggregation. The decentralized nature preserved patient privacy but also enhanced data security by minimizing the exposure of sensitive information. Overall, the framework had the potential for providing effective treatment recommendations while upholding patient confidentiality and data integrity (Tao et al., 2022).

Summary of Researches

Researcher(s)	Framework/Model	Key Techniques/Approaches	Main Contributions
Jinsung Yoon	EHR-Safe	Sequential encoder-decoder networks, GANs	- Synthesizing realistic EHR data - Addressing privacy concerns - Enabling safe sharing of EHR data
Raza Nowrozy et al.	Privacy model for EHR	Privacy ontology, Machine Learning (BERT)	- Identifying inadequate privacy policies in EHR systems - Enhancing healthcare information privacy
N. Khalid et al.	Privacy-preserving AI applications	Federated Learning, Hybrid Techniques	- Implementing privacy-preserving techniques in AI-based healthcare applications - Upholding regulatory standards
Bita Rouhani	DeepSecure	Machine Learning, Secure computation (Yao's GC)	- Performing privacy-preserving DL analysis of EHR data - Maintaining compliance with regulations
Olivia Choudhury et al.	Federated learning framework	Differential privacy mechanisms	- Mitigating privacy concerns in healthcare records - Improving utility through collaborative model training
J Lee et al.	Federated learning model	Hashing techniques, Temporal sequences	- Preserving patient privacy in EHRs - Effective disease prediction based on privacy-preserving patient similarity learning
Om Kumar et al.	FL+DQRE-Scnet	Federated Learning, Deep Q Reinforcement Learning	- Preserving patient privacy during data sharing - Minimizing communication rounds in healthcare data sharing
Tao Hai	Blockchain-based methodology	Blockchain, Federated Learning, Encryption	- Prioritizing privacy protection in EHR management - Enhancing data security

			through decentralized learning

Other methods for Improving Privacy in EHR.

Keeping patient information safe in Electronic Health Records is really important. While machine learning helps with this, there are other ways too. The following section would provide an overview of other techniques besides the use of machine learning that can be used for hiding the patient data. Things like using ICD codes to hide information, controlling who can see what, and special blockchain methods, along with data protection laws can be placed to ensure that privacy is upheld while processing medical data. Understanding all these ways is key to making sure patient records stay private and trustworthy.

A framework was proposed by Jayneel Vora in which blockchain technology was used to preserve the privacy of the EHR. The method introduced efficient storage and management of EHRs, with a focus on ensuring secure and accessible medical data for patients, providers, and third parties to maintain the confidentiality of the records. The framework's goals include analyzing its efficacy in meeting the needs of stakeholders and addressing privacy and security concerns within the healthcare landscape. The paper outlines the execution modes and communication methods of the proposed framework, along with its constituent nodes and IT components, with the help of smart contracts to perform the data processing functions, the vital functions were performed on the blockchain to address the scalability issues that may incur. The study presented a concept where the data being stored was encrypted and the updates and changes that were performed on the database were tracked using a decentralized ledger to preserve the data integrity (Vora et al., 2018).

In another approach to maintain the privacy of patient data, de-identification of medical records was performed by Zhao Tain-shu et al. The process involves designing a de-identification system for Protected Health Information in free-text medical records, to stick to privacy protection methods. The researchers reference HIPAA regulations and leverage techniques such as dictionary and regular expressions pattern matching to de-identify PHI, as the study was performed using datasets from China and Japan, the guidelines for secure storage of patient data according to HIPAA were not followed. The de-identification method proposed, ensures that personal information is obscured from third-party translators or unauthorized individuals, mitigating the risk of data breaches or privacy violations. By de-identifying PHI, the research addresses privacy protection differences between China and Japan and the secure exchange of health information between institutions in both countries. The experiment conducted demonstrates the effectiveness of the de-identification process, yielding a PHI identified ratio of 92.37%, surpassing the identified ratio of a single individual and proving to be more efficient and cost-effective than manual de-identification methods. Overall, the de identification of medical records enhances patient privacy by safeguarding sensitive information from unauthorized access or disclosure during health information exchange processes (Meystre et al., 2010).

Another method that featured a secure EHR can be included in this survey in which Suhair Alshehri et al, developed a secure cloud policy that made use of ciphering to ensure privacy of patient data in which privacy in the EHR system described is maintained through the use of Ciphertext-Policy Attribute-Based Encryption, which ensures secure access control to electronic health records stored in the cloud. With CP-ABE, EHRs are encrypted based on access policies derived from the attributes of healthcare providers, allowing decryption only by those possessing

the corresponding set of attributes necessary for access. This fine-grained access control mechanism ensures that healthcare providers can access only relevant EHRs aligned with their roles and responsibilities, thereby minimizing the risk of unauthorized data disclosure. The system also introduced a Key management server, that included secure key generation, revocation mechanisms, and emergency key escrow, to safeguard against unauthorized access and ensure data security in cloud-based environments. The integration of CP-ABE and key management protocols in the cloud-based EHR system preserves patient privacy by limiting access to sensitive health information to authorized users only. The critical concerns for managing the data records were addressed in their work (Suhair Alshehri, 2023).

Impact of Privacy Preserving in EHR

Healthcare records have improved the current system for managing patient data, and with the help of recent advancements and the abundant availability of data online, the concerns of managing the privacy of patient data persists. There are several privacy preservation techniques that are used globally that offer solutions to safeguard patient data. The following section would provide a brief overview of important privacy preserving techniques and how they can be embedded into current systems to help them with complying with global standards and how Privacy protected data mining can be performed on the patient data.

Research conducted by Anil Kumar and Ravinder Kumar discussed the different methods in safeguarding patient data and what tools can be used to maintain the privacy of user records. In their paper, they presented different existing solutions available to perform a comparative analysis of the pros and cons of each method. From the research it was highlighted that using data mining powered by AI and ML algorithms was a safe approach to balance the privacy of the

records. Moreover, the existing encryption and transfer protocols in PPDM implementations need to be improved so that minimal cost overhead is observed while dealing with patient records. Sectors like WSN, WBAN, and Smart Grids demand heightened attention to privacy and security measures. Current privacy standards, challenges, and regulations in EHRs, required the need for comprehensive privacy preservation policies and future research to ensure secure management and sharing of medical data. By evaluating different privacy models, and legal frameworks, the chapter aims to provide how privacy protection in the existing EHR systems can be maintained globally. Mechanisms such as minimal data disclosure, task-based access control, and secure database were being used to uphold patient privacy in healthcare domains. The paper presented a literature survey that evaluated the previous frameworks with the basics of AIC triad of privacy for healthcare records (A. Kumar & Kumar, 2020).

The research conducted by Shekha Chenthara et al. aimed to comprehensively review the security and privacy challenges in e-health solutions, and how healthcare records can be made safer in cloud computing environments. The primary objective of this research was to identify existing privacy-preserving approaches, analyze their effectiveness, and highlight research directions for building a robust security model for EHRs. The researchers conducted an extensive study spanning various academic databases and platforms. Their investigation covered papers published between 2000 and 2018, and how privacy requirements were followed in different frameworks with the help of secure cloud architecture and cryptographic techniques in preserving privacy. From the findings of the research, it was revealed that there were numerous internal threats in a healthcare environment that needed to be mitigated. Frameworks were analyzed that included encryption techniques such as Attribute Based Encryption but due to inefficiencies and computational limitations of current solutions, most of them lacked in

providing adequate privacy. The research analyzed the impact of privacy-preserving measures in EHRs as it is needed to maintain the trust of the patients. To prevent unauthorized access to sensitive health information and uphold individuals' fundamental right to privacy. The research also highlighted the need for more robust and comprehensive security measures to safeguard patient privacy and ensure the integrity and confidentiality of EHRs. By identifying research areas and the shortcomings of existing approaches, the study provided valuable insights in preserving privacy of e-health systems (Chenthara et al., 2019).

Keith Marsolo et al. implemented a privacy-preserving record linkage solution across PCORnet, to quantify patient overlap and analyze the demographic and clinical characteristics within the patient records. They found 81% of patient records were unique, with variable overlap between partners. From linking the records, it was indicated that disease prevalence was increased, moreover, the study further demonstrated the PPRL's potential and the need for careful consideration in linking decisions. They proposed a framework that used tokens based on demographic data for linkage and successfully created a de-duplicated summary. However, the study had limitations, including the lack of validation for the linkage algorithm and governance challenges limiting data scope. By successfully linking patient records in different healthcare systems to maintain confidentiality of data, the study showcases a method for improving the existing dataset without compromising privacy. This research highlights the potential of PPRL to enable data sharing and analysis in healthcare research to protect sensitive patient information to develop a solution for preserving the privacy for EHR (Marsolo et al., 2022).

Impact of Machine Learning on Privacy of EHR

As stated above, machine learning has made a significant impact in providing utility within the healthcare domain and the management and analysis of healthcare records is also a field where Machine learning models are being used actively. As ML algorithms are now being integrated into EHR systems, concerns about patient privacy have emerged. To address these concerns and assess the impact of ML on the privacy of EHR, several frameworks have been developed. In this literature review, the existing framework that have been placed to protect the patient records would be identified, by examining these frameworks, challenges and opportunities for research focusing on the implementation of ML in EHR systems to uphold patient rights would be evaluated. The evolving field of privacy and how ML can be used for improvement will be discussed in this section.

In a framework proposed by Aziz A. Boxwala et al. a method to detect suspicious accesses to Electronic Health Records using statistical and machine learning techniques was presented. Their motivation was to enhance privacy protection by identifying malicious access to EHRs. They extracted features from EHR access logs and organizational data to train logistic regression and support vector machines. These models were evaluated based on their ability to classify accesses as suspicious or appropriate. The research found that both LR and SVM models performed well in identifying suspicious accesses, with SVM showing slightly higher sensitivity. By applying these models to a dataset of labeled events, they achieved promising results which showed potential of using statistical models to uphold the privacy principles in healthcare records. Their method involved constructing an Event Data Mart from EHR access logs, to define useful features for detecting suspicious accesses, and creating a training set labeled by privacy officers. Through iterative refinement and oversampling techniques, a model which was fit for local use

was developed. The research provided significant advancements in the field of detecting malicious attacks, to enhance privacy protection in healthcare settings. Using their framework as a base model, better statistical models can be created that are used in clinical environments to improve the overall privacy of EHRs (Boxwala et al., 2011).

R. Venugopal et al. conducted research to address privacy concerns in Electronic Health Records by proposing a privacy-preserving Generative Adversarial Network. Their motivation was to generate synthetic data from EHR while preserving privacy and statistical properties. pGAN was evaluated on two datasets, one for classification and the other for regression tasks. They developed a pGAN architecture employing Multi-Layer Perceptrons for the Generator and Discriminator. The generator utilized Batch Normalization and activation functions to model complex data, while the discriminator incorporated dropout to prevent mode collapse. The study compared pGAN with other GAN models and demonstrated its ability to maintain privacy and utility. Results showed that pGAN produced synthetic data similar to the original while achieving high privacy scores. Machine learning models trained on synthetic data from pGAN performed comparably to those trained on original data, indicating high quality and utility of the synthetic data. The research highlighted the potential of synthetic data to replace actual health records for training machine learning models effectively to improve the privacy of actual healthcare records. By enabling hospitals to share synthetic data without compromising privacy or quality to reduce the privacy concerns of sharing healthcare records (Venugopal et al., 2022).

Summary of the Works Examined

Researcher(s)	Framework/Model	Key Techniques/Approaches	Main Contributions and Impacts
Jayneel Vora	Blockchain-based framework for securing Electronic Health Records (EHRs)	Efficient storage and management of EHRs, utilization of smart contracts, decentralized ledger for data integrity, encryption for data security	Ensuring secure and accessible medical data, addressing privacy and security concerns, providing a framework for analyzing efficacy and meeting stakeholder needs
Zhao Tain-shu et al.	De-identification system for Protected Health Information (PHI) in free-text medical records	Utilization of dictionary and regular expressions pattern matching, adherence to HIPAA regulations, de-identification of PHI, mitigation of data breach risks	Addressing privacy protection differences between countries, enhancing patient privacy, efficient de-identification process, facilitating secure health information exchange
Suhair Alshehri et al.	Secure cloud policy for Electronic Health Records (EHRs)	Ciphertext-Policy Attribute-Based Encryption (CP-ABE) for access control, Key management server for secure key generation and revocation mechanisms	Ensuring secure access control, limiting data disclosure to authorized users, safeguarding against unauthorized access, addressing critical concerns in data management
Anil Kumar and Ravinder Kumar	Comparative analysis of privacy preservation methods, utilization of AI and ML algorithms for data mining, evaluation of existing encryption and transfer protocols in PPDM implementations	Balancing privacy and utility with data mining, identifying the need for improved encryption protocols, highlighting privacy preservation policies and future research	Comprehensive overview of privacy preservation methods, highlighting the need for improved privacy standards, providing insights for secure management of medical data

Shekha Chenthara et al.	Review of security and privacy challenges in e-health solutions, analysis of existing privacy-preserving approaches, examination of cryptographic techniques for privacy preservation	Evaluation of internal threats in healthcare, analysis of encryption techniques, identification of research directions for robust security models	Identifying challenges in privacy preservation, analyzing effectiveness of existing approaches, highlighting the need for comprehensive security measures in healthcare systems
Keith Marsolo et al.	Privacy-preserving record linkage solution across PCORnet	Utilization of tokens for linkage based on demographic data, creation of a de-duplicated summary	Enabling data sharing and analysis while maintaining confidentiality, showcasing a method for improving datasets without compromising privacy
Aziz A. Boxwala et al.	Method to detect suspicious accesses to Electronic Health Records (EHRs)	Extraction of features from EHR access logs, training of logistic regression and support vector machines	Enhancing privacy protection by identifying malicious access, development of statistical models for detecting suspicious accesses
R. Venugopal et al.	Privacy-preserving Generative Adversarial Network (pGAN) for generating synthetic data from EHRs	Utilization of Multi-Layer Perceptrons for Generator and Discriminator, Batch Normalization, Dropout, comparison with other GAN models	Enabling the generation of synthetic data while preserving privacy, demonstrating utility and quality of synthetic data, potential for replacing actual health records

References

Boxwala, A. A., Kim, J., Grillo, J. M., & Ohno-Machado, L. (2011). Using statistical and machine learning to help institutions detect suspicious access to electronic health records. *Journal of the American Medical Informatics Association*, *18*(4), 498–505. https://doi.org/10.1136/amiajnl-2011-000217

Briguglio, W., Moghaddam, P., Yousef, W. A., Traoré, I., & Mamun, M. (2021). Machine learning in precision medicine to preserve privacy via encryption. *Pattern Recognition Letters*, *151*, 148–154. https://doi.org/10.1016/j.patrec.2021.07.004

Chenthara, S., Ahmed, K., Wang, H., & Whittaker, F. (2019). Security and Privacy-Preserving challenges of e-Health solutions in cloud Computing. *IEEE Access*, *7*, 74361–74382. https://doi.org/10.1109/access.2019.2919982

Choudhury, O., Gkoulalas-Divanis, A., Salonidis, T., Sylla, I., Park, Y., Hsu, G., & Das, A. K. (2019). Differential privacy-enabled federated learning for sensitive health data. *arXiv (Cornell University)*. https://arxiv.org/pdf/1910.02578.pdf

Khalid, N., Qayyum, A., Qayyum, A., Al-Fuqaha, A., & Qadir, J. (2023). Privacy-preserving artificial intelligence in healthcare: Techniques and applications. *Computers in Biology and Medicine*, *158*, 106848. https://doi.org/10.1016/j.compbiomed.2023.106848

Kumar, A., & Kumar, R. (2020). Privacy Preservation of Electronic Health Record: Current status and future direction. In *Springer eBooks* (pp. 715–739). https://doi.org/10.1007/978-3-030-22277-2_28

Kumar, C. U. O., Sudhakar, G., RM, B., Suguna, M., & Krithiga, R. (2023). EHR privacy preservation using federated learning with DQRE-Scnet for healthcare application domains. *Knowledge-Based Systems*, *275*, 110638. https://doi.org/10.1016/j.knosys.2023.110638

Lee, J., Sun, J., Wang, F., Wang, S., Jun, C., & Jiang, X. (2018). Privacy-Preserving Patient Similarity learning in a Federated Environment: Development and analysis. *JMIR Medical Informatics*, *6*(2), e20. https://doi.org/10.2196/medinform.7744

Marsolo, K., Kiernan, D., Toh, S., Phua, J., Louzao, D., Haynes, K., Weiner, M. G., Angulo, F., Bailey, L. C., Bian, J., Fort, D., Grannis, S. J., Krishnamurthy, A., Nair, V., Rivera, P., Silverstein, J. C., Zirkle, M., & Carton, T. (2022). Assessing the impact of privacy-preserving record linkage on record overlap and patient demographic and clinical characteristics in PCORnet®, the National Patient-Centered Clinical Research Network. *Journal of the American Medical Informatics Association*, *30*(3), 447–455. https://doi.org/10.1093/jamia/ocac229

Meystre, S. M., Friedlin, F. J., South, B. R., Shen, S., & Samore, M. H. (2010). Automatic de-identification of textual documents in the electronic health record: a review of recent research. *BMC Medical Research Methodology*, *10*(1). https://doi.org/10.1186/1471-2288-10-70

Nowrozy, R., Ahmed, K., Wang, H., & McIntosh, T. R. (2023). Towards a universal privacy model for Electronic Health record Systems: An Ontology and Machine Learning approach. *Informatics (Basel)*, *10*(3), 60. https://doi.org/10.3390/informatics10030060

Rezaeibagha, F., Win, K. T., & Susilo, W. (2015). A Systematic Literature Review on Security and Privacy of Electronic Health Record Systems: Technical Perspectives. *Health Information Management Journal*, *44*(3), 23–38. https://doi.org/10.1177/183335831504400304

Rouhani, B. D., Riazi, M. S., & Koushanfar, F. (2018). DeepSecure: Scalable Provably-Secure Deep Learning. *2018 55th ACM/ESDA/IEEE Design Automation Conference (DAC)*. https://doi.org/10.1109/dac.2018.8465894

Suhair Alshehri, S. R. (2023). *Designing a Secure Cloud-Based EHR System using Ciphertext-Policy Attribute-Based Encryption.* New York: Rochester Institute of Technology.

Tao, H., Zhou, J., Srividhya, S., Jain, S., Young, P., & Agrawal, S. (2022). BVFLEMR: an integrated federated learning and blockchain technology for cloud-based medical records recommendation system. *Journal of Cloud Computing*, *11*(1). https://doi.org/10.1186/s13677-022-00294-6

Venugopal, R., Shafqat, N., Venugopal, I., Tillbury, B. M. J., Stafford, H. D., & Bourazeri, A. (2022). Privacy preserving Generative Adversarial Networks to model Electronic Health Records. *Neural Networks*, *153*, 339–348. https://doi.org/10.1016/j.neunet.2022.06.022

Vora, J., Nayyar, A., Tanwar, S., Tyagi, S., Kumar, N., Obaidat, M. S., & Rodrigues, J. J. P. C. (2018). BHEEM: A Blockchain-Based Framework for Securing Electronic Health Records. *IEEE Xplore*. https://doi.org/10.1109/glocomw.2018.8644088

Yoon, J., Mizrahi, M. J., Ghalaty, N. F., Jarvinen, T., Ravi, A. S., Brune, P., Kong, F., Anderson, D., Lee, G., Meir, A., Bandukwala, F., Kanal, E., Arık, S. Ö., & Pfister, T. (2023). EHR-Safe: generating high-fidelity and privacy-preserving synthetic electronic health records. *Npj Digital Medicine*, *6*(1). https://doi.org/10.1038/s41746-023-00888-7