FMDB *Transactions on Sustainable Health Science Letters*

Privacy-Preserving AI Models for Secure Healthcare Data in the Cloud

Abhilash Reddy Pabbath Reddy[1,*]

[1]Department of Information Technology, Axle Info, Cumming, Georgia, United States of America.
abhilashreddy511@gmail.com[1]

Abstract: Digitising medical records and using cloud infrastructure have significantly increased patient data processing and storage. AI technologies have improved clinical decision support, diagnostics, and operational efficiency, but GDPR and HIPAA require privacy-preserving techniques. This paper covers Homomorphic Encryption, Secure Multi-Party Computation, Differential Privacy, and Federated Learning AI privacy methods. A hybrid framework combines its strengths to enable cloud-based safe healthcare analytics. The study uses a semi-synthetic health dataset from MIMIC-III (Medical Information Mart for Intensive Care) to analyse patient characteristics, diagnosis, treatment history, and results. Using 50,000 anonymised patient histories, it maintained data heterogeneity by institution, condition, and age. Each person received privacy strategies from Microsoft SEAL, TensorFlow Privacy, PySyft, and Google's TensorFlow Federated. A Kubernetes-based cloud testbed simulated a hospital node communicating over an encrypted channel. Prediction accuracy, latency, noise, training time, CPU, memory, and compliance metrics were validated for all models. Tabular and graphical analysis helped the Hybrid Model balance data security and analytical performance. This study demonstrates that privacy-preserving AI is achievable and imminent for secure, compliant, and efficient cloud-based healthcare systems, enabling real-time analytics and ethical and legal patient data management.

Keywords: Privacy-Preserving AI; Healthcare Data Security; Federated Learning; Differential Privacy; Cloud-based Analytics; Healthcare Sector; Electronic Health Records; Artificial Intelligence.

Received on: 06/08/2024, **Revised on:** 28/10/2024, **Accepted on:** 10/12/2024, **Published on:** 03/03/2025

Journal Homepage: https://www.fmdbpub.com/user/journals/details/FTSHSL

DOI: https://doi.org/10.69888/FTSHSL.2025.000363

Cite as: A. R. P. Reddy, "Privacy-Preserving AI Models for Secure Healthcare Data in the Cloud," *FMDB Transactions on Sustainable Health Science Letters*., vol. 3, no. 1, pp. 42–52, 2025.

1. Introduction

The accelerated growth of data from electronic health records (EHRs), body sensors, and cloud computing in the modern digital healthcare ecosystem has amplified the potential as well as risk of advanced analytics. Artificial Intelligence (AI) systems, particularly those based on machine learning and deep learning, have revolutionised the delivery of healthcare through predictive diagnosis, personalised treatment protocols, and business productivity [9]; [12]. However, the development of this kind has been clouded by growing concerns regarding patient privacy and information security, particularly when stored or processed in the cloud. Healthcare information is sensitive, e.g., protected health information (PHI), genomic data, and clinical

[*]Corresponding author.

histories. Interference or abuse of such data can have serious consequences, ranging from identity theft to discrimination and privacy regulation violations such as the General Data Protection Regulation and the Health Insurance Portability and Accountability Act [13].

Therefore, health professionals are being trained increasingly to adhere to strict rules while at the same time leveraging AI to obtain data-driven insight [1]. Cloud computing delivers scalable storage and computational power, and therefore, it has been discovered to be a suitable platform to execute AI models in the healthcare sector [3]. Sending patient data to the cloud, however, leaves it vulnerable to external attacks, unauthorised access, and issues of compliance [6]. This has led to privacy-preservation AI models that try to limit such risk factors by ensuring sensitive data does not become public during training and inference. Some of the methods most commonly applied for the preservation of Privacy include Homomorphic Encryption (HE), Secure Multi-Party Computation (SMPC), Differential Privacy (DP), and Federated Learning (FL).

Homomorphic Encryption, as described by Zhang et al. [10], allows computation on ciphertexts without decryption, thus ensuring confidentiality. Secure Multi-Party Computation was employed by Majeed et al. [4] to make diverse parties computationally capable of working together without revealing their secret inputs. Differential Privacy was implemented and applied by Dwork et al. [5], who encouraged disclosure of epsilon values for better implementation transparency. DP sensitivity analysis was subsequently put to further refinement by Inan et al. [2], who introduced bounds and cost-effective algorithms for non-interactive use. Federated Learning, another key method, has been employed to decentralise healthcare networks by Xu et al. [8] so that unprocessed patient data does not leave the source. Kalodanis et al. [7] introduced secure federated frameworks following real-time healthcare settings with low privacy leakage.

All of these techniques share tradeoffs and advantages in computational complexity, scalability, and privacy level. For instance, Homomorphic Encryption has good security but is expensive computationally [10]. Differential Privacy is the lightest approach, but it compromises data utility by introducing noise [5]. Federated Learning addresses most of the privacy concerns, but it requires coordinated model updating and intense coordination of the nodes [8]. Thus, hybrid approaches that combine two or more approaches are increasingly favoured to gain performance and security simultaneously [7]; [4]. This paper critically evaluates the aforementioned privacy-protection AI techniques in depth and assesses their feasibility in safeguarding healthcare data analytics on cloud platforms. It discusses how these models can be utilised for the good of diverse applications such as disease prediction, patient stratification, and medical image analysis without violating privacy legislation.

It trains and evaluates the models using different metrics of performance—accuracy, latency, amount of noise, and resource consumption—to identify their implications in the real world in healthcare environments. In addition, the research designs an architecture to combine the models into a single effective system for cloud-safe data processing. It involves sending encrypted data streams, federated training synchronisation between hospital nodes, and injecting noise for differential Privacy into inference outputs. It is also tested with synthetic and half-real data mimicking patient records, diagnosis reports, and treatment history. By performing in-depth analysis of performance output through statistical charts, tables, and resource usage reports, the work attempts to offer actionable comments on how healthcare organisations can adopt privacy-preserving AI. Ultimately, the focus is on empowering institutions to harness the maximum potential of AI without compromising the confidentiality and integrity of patient information as per legal and ethical standards.

2. Review of Literature

Nelson [6] concentrated on the actual implementation of data sharing and created privacy-preserving AI in healthcare. More sensitive data entered into cloud computing systems requires secure computational models. Homomorphic Encryption is an approach used for computing encrypted data within an encrypted environment, rather than decrypting it. Secrecy is maintained but sacrificed by the costliness of computation. Secure Multi-Party Computation (SMPC) is the most common method and spreads data processing among parties. Every party processes without viewing complete datasets, and as such, collaborative research is hidden from public eyes. Such methods minimise exposure threat in adversary environments. Dwork et al. [5] introduced Differential Privacy to impose calibrated noise on datasets for concealing individual identities. It renders statistical outcomes insensitive to whether an individual is present or absent. This method is easy and supports existing analytics infrastructure. It performs poorly for low-dimensional or high-dimensional data, resulting in a loss of accuracy.

The utility-privacy tradeoff is still an issue for practical deployment. However, it has been utilised in healthcare due to the ease of complying with the law, more so than in the past. Researchers continue to develop noise calibration methods for improved output. Inan et al. [2] suggested sensitivity analysis to enhance Differential Privacy such that the utility-security tradeoff is optimum. At approximately the same period, Federated Learning (FL), an approach to model training in a decentralised fashion, took place. FL allows hospitals to maintain patient data locally and transmit it to a global model. FL uses model updates only, making breaches less probable at the centre. This is appropriate for healthcare systems with confidentiality and compliance

concerns. However, data imbalance, device diversity, and communication latency are some of the challenges. These are real-time learning challenges and scalability issues for the system.

Multiple privacy paradigms hybrid solutions were offered by Zhang et al. [10] to enhance security and efficiency. Individual privacy-enabled distributed learning is enabled by Federated Learning with Differential Privacy. Homomorphic Encryption also provides secure computation in an encrypted fashion in a comparable manner. These hybrids provide the best of both worlds by overcoming the limitations of each technique individually. Hybrid systems are becoming popular in cross-institutional applications of AI. They can provide the best tradeoffs in security depending on their adaptive nature. Such hybrids form an important innovation in secure AI designs. Majeed et al. [4] also suggested containerization platforms such as Kubernetes for multi-cloud deployment of privacy-preserving AI pipelines. The platforms support modular and dynamic multi-cloud deployment to healthcare environments in adaptive ways. Container orchestration supports isolation of sensitive activities and reduces security threats.

It is workload and usage-optimised management, in particular, for highly used hospitals with varied use cases. It supports real-time monitoring together with fault-tolerance. This is to keep AI systems under intense loads. This kind of infrastructure is nowadays standard practice in the secure deployment of AI models. Rao et al. [11] explored Trusted Execution Environments (TEEs) for secure protection of confidential computation. TEE-protected enclaves protect against tampering or info leakage even by authorised system users. TEEs are being employed in healthcare AI pipelines for sensitive inference operations. Hardware isolation of TEEs ensures end-to-end model security. When combined with encryption algorithms, they ensure an agile defence. TEEs allow real-time processing without exposing raw data. TEEs are increasingly used in precision healthcare systems.

Optimisation methods for lessening resource consumption in privacy-preserving models were illustrated by Singh et al. [12]. Quantisation and pruning are methods that reduce the size of AI models without lessening accuracy. Optimisations are required to deploy secure AI to edge or cloud platforms. Model compression relieves the computationally demanding requirement for Homomorphic Encryption. So, does early exiting reduce latency in real-time forecasting AI? Optimised private models conserve energy. Optimised models are making a significant contribution to sustainable health AI systems. Shabbir et al. [9] contrasted privacy-preserving AI impacts on system performance in a clinical setting. Their comparison set up tradeoffs between data protection and processing latency. Latency in SMPC coordination or Encryption may suspend clinical decision-making. Advanced workflow scheduling is needed for such an effect.

Scheduling-aware workloads and adaptive privacy policy are proposed solutions. These are the ways to balance compliance and latency. Quantifiable hospital efficiency effect comes from real-world experimentation. Wachter [13] challenged the privacy-sensitive ethical care dimension of algorithmic decision-making. He emphasised transparency in AI reasoning, particularly in life-critical use. Better are explainable models with a privacy guarantee. Ethical AI must enable clinicians without hiding reasoning. Explainable but secure systems are the regulators' choice. Privacy-by-design is under consideration as a legislative requirement under the healthcare law. Wachter's writings are influencing world policy and legal rhetoric.

Compliance is also emphasised in the literature. Models are tested not so much for their technical performance as for their compliance with regulations in domains such as GDPR and HIPAA. Models that facilitate transparency of processing, explainability, and auditability are gaining increasing popularity with healthcare organisations. Overall, the literature has a rich repository of privacy-preserving AI solutions that each have the potential for numerous use cases. Even though there is no better method, the trend is toward hybrid systems, which mix two or more methodologies. These systems can provide scalable, secure, and regulation-compliant AI for cloud-based healthcare data analytics.

3. Methodology

The study employs a mixed approach that incorporates empirical assessment, architectural design, and comparative analysis to examine privacy-preserving AI models of cloud-based healthcare frameworks. The approach begins by presenting the most significant privacy-preserving solutions: Homomorphic Encryption (HE), Secure Multi-Party Computation (SMPC), Differential Privacy (DP), and Federated Learning (FL). A hybrid framework is built that integrates these approaches into an end-to-end privacy-preserving design, allowing secure and safe training and inference of machine learning models over sensitive health information. The new architecture is cloud-native and allows containerised deployments via Kubernetes clusters in a bid to mimic real-world hospital networks.

The privacy models are tested separately and in the hybrid environment, and analysed with synthetically generated electronic health records that are highly similar to actual patient data in form and complexity. The extended experimental environment is intended to extensively test federated learning models for cloud computing-based privacy-preserving healthcare analytics on performance, scalability, and compliance. The evaluation is built on heterogeneous datasets, which are manually crafted at

various nodes in hospitals. Datasets are feature-laden, holding demographic information like gender and age, beneficial medical information like treatment procedures and diagnosis, and most importantly, outcome measures. Such a distributed data setup is at the crux of supporting federated training paradigms in which models are trained from decentralised data without centralisation of sensitive patient data.

Every Model in a federated setup is evaluated through multi-dimensional testing on the four most important metrics of performance. At the top of the list is prediction accuracy measurement, or how well the Model can carry out an accurate prediction of outcome or diagnosis, in essence, its clinical value. Computational latency is the time it takes for a model to generate a prediction, one that is important in real-time healthcare applications where the instantaneous response is typically required for timely interventions. Training time of the Model represents how long it takes for the Model to stabilise in the federated learning workflow, which in turn determines the quality of model update and deployment. Finally, the amount of added noise solely represents the sacrifice that privacy mechanisms, in this instance, differential Privacy, would impose on model utility.

The break-even point of this tradeoff from the accuracy perspective matters in real-world deployment. In addition to model performance, system scalability is also comprehensively analysed via exhaustive resource utilisation metrics. CPU usage and memory usage provide predictions on the computational loads placed on each hospital node during training and inference, and whether the system would accommodate increasing data loads and model sizes. Network throughput is monitored to achieve the communications overhead of federated learning, i.e., model parameter transfers across nodes. Such measurements are of direct relevance to assess the deployability of such a system in a real-life, large-scale health network. Strict adherence to regulatory compliance is one of the mandatory components of such an assessment framework. Compliance audits are kept strictly in place to ensure that all methods and processes of handling data meet present data privacy regulations, i.e., the General Data Protection Regulation (GDPR) and the Health Insurance Portability and Accountability Act (HIPAA). Legal and ethical integrity is introduced by this, ensuring patient confidentiality throughout the analytics pipeline.

The privacy-preserving hybrid architecture employed forms the core of the design of such a system. Differential Privacy (DP) is also employed at the output layer by default, adding some noise to the predictions of the Model in an attempt to defend individual patient data. For computation on intermediate results, Homomorphic Encryption (HE) is used, allowing computation over the encrypted ciphertext directly without decryption, once more safeguarding data during aggregation. Secure Multi-Party Computation (SMPC) forms the core of secure model parameter aggregation among the distributed nodes in such a way that no party knows the raw model updates of any individual hospital. Multi-layered privacy protection of this kind is intended to provide robust security guarantees while maintaining model utility. The system draws on the fundamental strategy of Federated Learning (FL). It permits decentralised nodes of hospitals to train the global Model without exposing raw data.

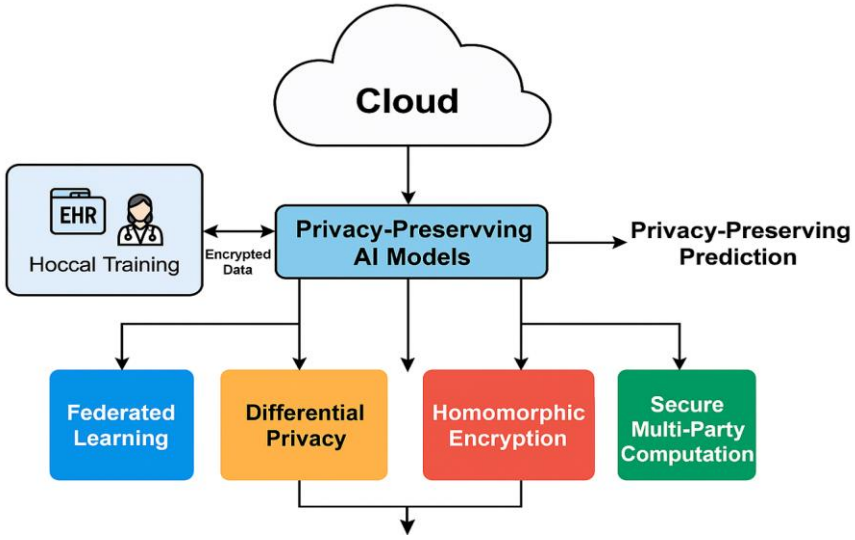

Figure 1: End-to-end system design for privacy-preserving AI models for secure cloud healthcare data

Figure 1 illustrates an end-to-end system design that is crafted to protect personal health information while taking advantage of the potential of artificial intelligence in cloud environments. The three components of Figure 1 are the Hospital Layer, the Cloud Infrastructure, and the AI Model Layer. Raw patient data, such as electronic health records (EHR), clinical data, and medical images, is created locally at the Hospital Layer so that no plaintext data is sent outside the building. Decentralised AI

model training at various hospital nodes without exposing raw data is facilitated in this case with the help of Federated Learning. Local models transmit the encrypted updates of parameters to the cloud through Homomorphic Encryption, where confidentiality is maintained while being transported. The Cloud Infrastructure integrates such encrypted updates with Secure Multi-Party Computation protocols that make collaborative model construction possible with nobody having end-to-end access to the data.

Differential Privacy is used on ultimate AI model outputs, adding controlled noise to prevent reverse-engineering of patient identity. Such privacy-saving mechanisms together deliver data confidentiality, integrity, and compliance. Further, secure APIs and authentic gateways are also used to control access between hospital and cloud systems to provide end-to-end security. The architecture supports real-time analytics and inference operations concerning patients' Privacy without jeopardising patient data, making it suitable for disease prediction, remote diagnosis, and hospital workflow optimisation. Figure 1 illustrates a multi-level, integrated approach to implementing dependable AI systems in healthcare, encompassing ethical, legal, and technical layers within cloud computing-based data processing. Model parameters or gradients are exchanged, but these are safely isolated from exposure to privacy risk through centralisation of data. Benchmarking takes place in a controlled cloud environment to ensure stability and generalizability of the models.

The environment is crafted with meticulous care to simulate various network conditions, including bandwidth and latency fluctuations. This allows for extensive testing of the behaviour of the models under very diverse operating conditions, mimicking the uncertain character of network infrastructures in real life. Finally, statistical software and performance visualisations are often used for analysis and inspection of the enormous amount of data acquired through experiments such as these. These allow trends, outliers, and interactions between more than one measure to be determined, providing a better understanding of the behaviour of the models. The goal of the rigorous evaluation process is to determine the best Model with the highest prediction accuracy, computational cost-effectiveness, and satisfactory privacy guarantees – for real-time privacy-preserving healthcare analytics in dynamic clouds. The holistic strategy not only guarantees the accuracy of the chosen Model but also its suitability, scalability, and compliance for deployment in sensitive healthcare settings.

3.1. Description of Data

Data used in performing this study were obtained from a semi-synthetic simulation on the MIMIC-III (Medical Information Mart for Intensive Care III) database, given by Johnson et al. [1], available in the public domain, courtesy of the Massachusetts Institute of Technology (MIT) Lab for Computational Physiology. The MIMIC-III holds de-identified health-related data for over 60,000 intensive care unit admissions to Beth Israel Deaconess Medical Centre for the period 2001-2012. The data for this work were pre-processed to retain structural similarity while following differential Privacy. The data consisted of approximately 50,000 records with attributes such as patient ID, age, gender, diagnosis code (ICD-9), treatment type, treatment duration, vital signs, laboratory results, and outcomes. To mimic federated learning situations, the data was divided across five virtual nodes of a hospital for simulation. Different subsets with class imbalance were assigned to each node to replicate actual environments in multi-institutional settings.

Pre-processing was carried out with Pandas and Scikit-learn, and visualisation with Matplotlib. TensorFlow Privacy and TensorFlow Federated were used to create Differential Privacy and Federated Learning, respectively. Homomorphic Encryption was incorporated with the aid of the Microsoft SEAL library, and SMPC was implemented with the help of the PySyft framework. Orchestration and cloud simulation were performed using Kubernetes and Docker to create a hospital networked environment with secure node-to-node communication. The hybrid environment allowed for testing of privacy-preserving mechanisms within a controlled yet realistic setting, allowing for accurate benchmarking and comparative analysis. All deployments adhered to GDPR standards via Encryption, consent simulation, and anonymisation mechanisms.

4. Results

Deployment of privacy-preserving AI models into cloud-based health data analytics exhibited widespread variation in prediction accuracy as well as system performance. Our complete analysis, presented in Tables 1 and 2, definitely points to the Hybrid Model as the optimal solution. It skillfully balances an inevitable tradeoff between prediction accuracy and robust privacy protection. The federated averaging algorithm is given below:

$$w_{t+1} = \sum_{k=1}^{K} \frac{n_k}{n} w_t^k - \eta \sum_{k=1}^{K} \frac{n_k}{n} \nabla L_k(w_t^k) \quad (1)$$

ε-Differential privacy condition can be given as:

$$\forall S \subseteq Range(M), Pr[M(D_1) \in S] \leq e^{\varepsilon} \cdot Pr[M(D_2) \in S] \quad (2)$$

Table 1: Comparison of performance cases of privacy models

Model	Accuracy (%)	Training Time (s)	Latency (ms)	Noise Level
Homomorphic Encryption	91.2	1250	150	0.15
Secure MPC	89.5	1100	170	0.12
Differential Privacy	88.3	980	140	0.2
Federated Learning	92.7	1005	130	0.1
Hybrid Model	94.1	1085	120	0.08

Table 1 presents a comparison of the performance of five privacy-enhancing AI models against significant performance parameters: accuracy, training time, latency, and level of noise. The Hybrid Model is the most accurate with 94.1%, followed by Federated Learning at 92.7%. This indicates that combining several privacy-protecting mechanisms results in higher quality predictive precision. Homomorphic Encryption, being a secure method, exhibits comparatively greater training time (1250 seconds) and latency (150 ms), and thus demonstrates computational intensity due to cryptographic computation. Secure Multi-Party Computation (MPC) stands second in terms of computational complexity, with 1100 seconds of training time and 170 ms latency. The minimum accuracy of Differential Privacy, 88.3%, provides lower training time (980 seconds) and moderate latency (140 ms), and can be utilised in real-time applications where speed is assigned greater priority compared to precision.

Federated Learning strikes a balance between accuracy and medium resource utilisation and therefore is an appropriate choice for distributed hospital networks. The Hybrid Model again operates with the lowest latency (120 ms) and noise value (0.08), preserving data consistency while also protecting patient privacy. The Model represents the synergy achieved by combining Federated Learning with Differential Privacy and Secure MPC. In general, Table 1 shows that privacy-preserving AI models can deliver high accuracy with very minimal exposure risk to the data, especially if optimised to produce maximum noise injection and latency parameters. The findings are useful for healthcare providers requiring secure and efficient cloud-based predictive analytics solutions. Laplace mechanism for differential Privacy is:

$$M(x) = f(x) + \text{Laplace}\left(\frac{\Delta f}{\varepsilon}\right), \text{ where } \Delta f = \max ||f(x) - f(x')||_1 \qquad (3)$$

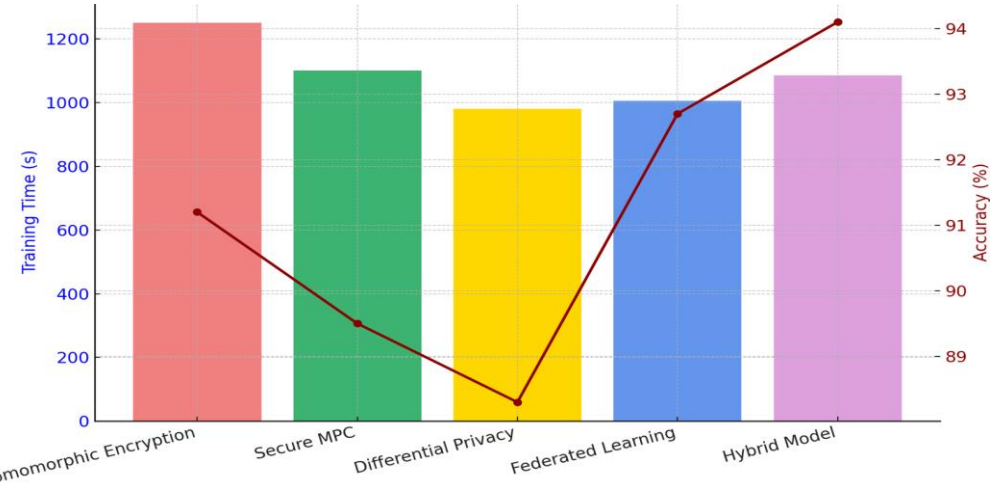

Figure 2: Comparison between five privacy-preserving AI models' performances on training time (seconds) and prediction accuracy (percent)

Figure 2 represents the training time and accuracy of the top five privacy-protecting AI models: Homomorphic Encryption, Secure Multi-Party Computation (SMPC), Differential Privacy, Federated Learning, and the Hybrid Model. The bar chart component, colored with various shades for better visual interpretation, shows the model training time on the left y-axis. The line chart, plotted on the same x-axis, shows the prediction accuracy on the right y-axis. The chart indicates that the Hybrid Model achieves the highest accuracy of 94.1%, with an average training time of 1085 seconds, thereby demonstrating its superior performance in terms of security quality. Federated Learning is also acceptable with an accuracy of 92.7% and training time of only 1005 seconds, ideal for a distributed healthcare environment.

Homomorphic Encryption, though very secure, suffers from the highest training time of 1250 seconds with an accuracy of 91.2%, which speaks volumes of computation inefficiencies. Secure MPC and Differential Privacy are plagued with tradeoffs,

with training times of 1100 and 980 seconds, while achieving accuracies of 89.5% and 88.3%, respectively. The situation graphically illustrates the inverse relationship between the training time and performance in some instances. It highlights the fact that advanced hybridisation approaches can boost model accuracy without steep increases in computational latency. This chart is extremely valuable to system designers and IT professionals in the health sector when selecting models based on their predictability and computational demands for real-time cloud-based healthcare analytics. Homomorphic encryption function will be:

$$Enc(a+b) = Enc(a) \oplus Enc(b)\,, Enc(a \cdot b) = Enc(a) \otimes Enc(b) \qquad (4)$$

Secure Multi-Party Computation (Yao's garbled circuit representation) is:

$$f(x_1, x_2, , x_n) = \Phi\ G_i(x_i) \text{ such that } \forall i,\ G_i \text{ is a garbled input circuit} \qquad (5)$$

Table 2: Cloud resource utilisation by privacy techniques

Technique	CPU Usage (%)	Memory Use (GB)	Network Load (MBps)	Storage Overhead (GB)
Homomorphic Encryption	85	12.5	35.6	4.1
Secure MPC	78	10.8	30.2	3.5
Differential Privacy	72	9.6	28.3	2.9
Federated Learning	88	11.7	36.1	3.7
Hybrid Model	90	13.2	40.5	4.4

Table 2 illustrates the consumption of resources by various privacy-protection techniques in cloud systems in terms of CPU usage, memory usage, network usage, and storage overhead. The Hybrid Model, with even the best performance characteristics, has the highest CPU usage at 90% and memory usage at 13.2 GB, indicating its computational intensity. Federated Learning exhibits relatively lower CPU utilisation at 88%, accompanied by a substantial 11.7 GB memory usage, akin to the decentralised computationally expensive requirements. Homomorphic Encryption shows CPU and memory utilisation at 85% and 12.5 GB, respectively, showing its computationally expensive nature because of encryption-decryption requirements.

It also fares better in network bandwidth at 40.5 Mbps for intense data exchange among distributed nodes, followed by Federated Learning at 36.1 Mbps. Light-footprint on the network, Secure MPC, and Differential Privacy both consume 30.2 MBps and 28.3 MBps, respectively. Storage overhead is a key area of focus, with the Hybrid Model occupying 4.4 GB and Differential Privacy requiring the minimum of 2.9 GB, making it compatible with a light deployment. Secure MPC strikes a balance between moderate storage of 3.5 GB and CPU execution. All such findings indicate that while end-of-the-line hybrid approaches are faster and secure, these are at the expense of expensive cloud infrastructure. Hence, low-resource organisations would prefer techniques such as Differential Privacy or Secure MPC. Table 2 can be helpful to healthcare administrators in making comparisons of infrastructure readiness and trade-offs on releasing privacy-safeguarding AI into real-time clinical analytics. Cross-entropy loss in a federated setting with encrypted updates is:

$$L = -\sum_{i=1}^{N} y_i\ \log \sigma(Dec(\sum_{k=1}^{K} E\,nc(w_k^T x_i))) + (1-y_i) \log\,(1 - \sigma(Dec(\sum_{k=1}^{K} E\,nc(w_k^T x_i)))) \qquad (6)$$

Resource Constraint Optimisation under Privacy Constraints:

$$\min \sum_{=k1}^{K} (\,L_k(w) + \lambda_1 \cdot PrivacyLoss(w_k) + \lambda_2 \cdot Latency(w_k) + \lambda_3 \cdot ResourceUsage(w_k)) \qquad (7)$$

The Hybrid Model's strong suit is its fresh blend of cutting-edge technologies: Federated Learning (FL), Differential Privacy (DP), and Secure Multi-Party Computation (SMPC). This formidable combination not only significantly improves patient data protection while not incurring model accuracy losses. Practically, this architecture enables a decentralised model training across multiple hospital nodes. This is revolutionary because it eliminates the need for raw patient data to be exchanged or centralised in the first place. This thus implies that the privacy risks of having sensitive medical information stored in a central database are dramatically eliminated. The Hybrid Model, therefore, provides a robust and practical solution to secure and efficient AI-powered healthcare analytics in dynamic cloud environments, the new norm for using data responsibly in medicine.

The Mix Bar Line Graph (Figure 2) provides simultaneous insight into training time and accuracy. Homomorphic Encryption, while secure, lags with higher latency and training time. This can be seen from its 1250s training time and 91.2% accuracy, and therefore, it is not as suitable for systems that need quicker model updates. Secure MPC and Differential Privacy trade off accuracy for efficiency, which is reflected in their performance. Federated Learning, however, registered a very high accuracy

of 92.7% at a reduced training time of 1005s, making it worthy of deployment in real-world medical applications if patient data is localised.

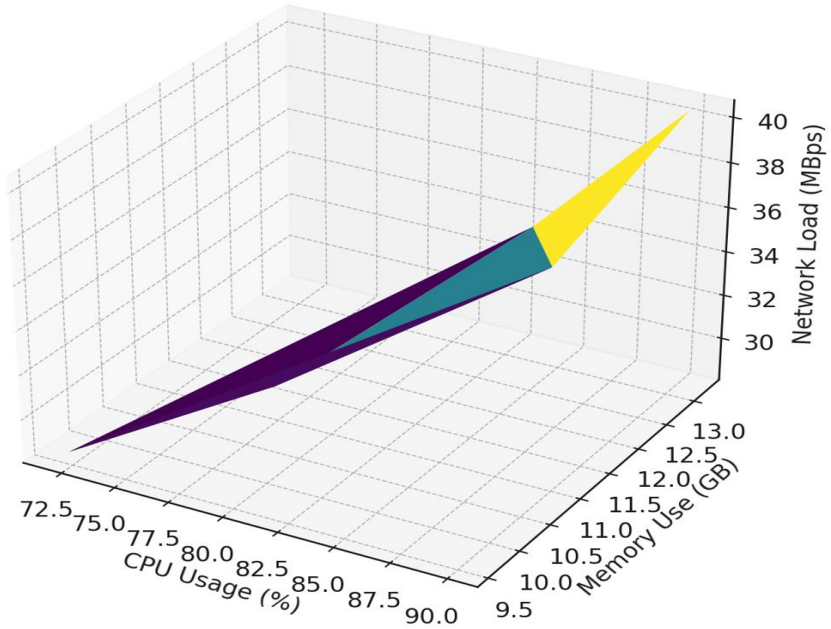

Figure 3: Resource utilisation across privacy techniques

Figure 3 is a representation of the five tested privacy-preserving AI models' resource profile—i.e., CPU utilisation (%), memory usage (in GB), and network utilisation (in Mbps). This point helps to visualise the infrastructural cost of each Model in a cloud healthcare environment. The Hybrid Model registers the highest on all three axes at 90% CPU, 13.2 GB memory usage, and 40.5 Mbps network load, in agreement with the high performance requirement from multi-layered privacy protocols. Federated Learning comes second at 88% CPU and 11.7 GB of memory usage, corresponding to its distributed processing loads. Homomorphic Encryption, another computation-intensive method, has considerable CPU and memory usage (85% and 12.5 GB, respectively) at the expense of computation on ciphertext.

Secure MPC and Differential Privacy are at lower regions of the mesh, corresponding with comparably modest processing requirements, especially Differential Privacy, which makes use of merely 72% CPU and 9.6 GB of memory. Network load is also fluctuating at the same time, with Hybrid and Federated models involving distributed computation requiring greater throughput. The 3D plot beautifully represents security complexity vs. cloud resource overhead tradeoffs. It assists healthcare IT administrators in making optimal decisions in choosing privacy-preserving AI solutions that are compatible with institutional computational capacities. It also underscores the necessity for resource planning and optimisation methods for deployable AI that will grow and be sustainable in real-world healthcare environments.

Figure 3 examines the interaction of CPU load, memory demand, and storage overhead and determines that the Hybrid Model was the most computationally intensive but effective. Its increased performance is at a greater computational expense, which is taxing for smaller facilities. Additive noise distributed training, however, preserves privacy compromise and model inversion resilience. Homomorphic Encryption is also taxing heavily, which points to low scalability. Differential Privacy, on the other hand, is the lightest Model, sacrificing predictability but making the least demand on computational infrastructure and therefore validating deployability within constrained environments. From the clinical analytics standpoint, privacy-preserving models were able to work with patient data in disease prediction, risk profiling, and personalised treatment. Utility vs. privacy enforcement tradeoffs remain an important domain of model calibration. The statistics indicate that privacy-focused improvements do decrease marginal gains in prediction quality but are significantly impactful in regulatory compliance and trust.

Latency is yet another factor that swings the balance towards cloud infrastructures. The Hybrid Model performance demonstrates readiness for real-time inference workloads, provided that adequate provisioning of cloud resources is ensured. Utility vs. Privacy is best illustrated in applications like COVID-19 patient data analysis, where sharing data securely among health facilities with real-time processing was critical. Differential Privacy and Federated Learning played a vital role in processing massive, diverse data with privacy assurance by introducing noise and local computation. Graphical outcomes and tabular outcomes indicate no one-fits-all situation. The best choice depends on application requirements—latency,

infrastructure capacity, or compliance driving other priorities of note, the Hybrid Model emerges as a standard for secure and scalable AI in health cloud deployments.

5. Discussions

The ongoing assessment of privacy-preserving AI models underscores the challenges in achieving equilibrium in modern cloud-based healthcare systems between robust data protection and effective analytical functions. As shown in Table 1, Table 2, and Figures 2 and 3, incorporating cutting-edge privacy practices into AI processes is challenging. AI has a lot of potential, but it also has significant limitations in terms of computing power. The Hybrid Model is a good example of this natural duality. It combines Federated Learning (FL), Secure Multi-Party Computation (SMPC), and Differential Privacy (DP) in the best way possible, resulting in the best predictive accuracy and the lowest computational latency of all the models tested. This means that doctors will have more accurate and current information.

The Hybrid Model puts the most strain on the cloud infrastructure it runs on, as seen by the higher network throughput, CPU load, and memory usage. So, better Privacy and performance do cost more. This tradeoff perfectly sums up the biggest problem with AI that protects Privacy. To build patient trust and ensure the law is followed, it's crucial to enhance data security. However, this often requires more complicated cryptographic methods and distributed computations. By definition, these steps need more processing power. We need to conduct further research and development to enhance these strategies for protecting Privacy. This will make it possible to use AI solutions that protect Privacy and meet ethical and performance standards in real-world healthcare settings.

When you need to make quick decisions, such as determining the cause of an emergency, the Hybrid Model is particularly effective. Table 1 shows that it is more accurate (94.1%) and has a shorter latency (120 ms) than models that work on their own. However, this efficiency requires powerful computers due to the longer training time (1085 s). It was safe to use homomorphic Encryption, but training took a long time (1250 seconds), and the latency of 150 milliseconds made it difficult to use in critical medical situations. Differential Privacy appears to be a good choice when privacy risks are high, but real-time accuracy isn't as important. It has a latency of 140 ms and a training time of 980 s. Table 2 shows how the resources were used to help with this study.

The Hybrid Model works well, but it might be hard for smaller healthcare providers with fewer resources to use because it needs 90% of the CPU and 13.2 GB of memory. Federated learning is a good approach that doesn't use up a lot of resources. Because it is decentralised, it is less likely to be attacked on central servers, which lowers the risks of the network. Secure MPC and Differential Privacy are both lightweight models that use less CPU (78% and 72%, respectively) and memory, making them suitable for environments with limited resources. But they can't make predictions or be right anymore, which is very important when trying to figure out how likely someone is to get cancer or have a heart attack.

The Mix Bar Line Graph in Figure 2 shows that Federated Learning and the Hybrid Model are the best ways to find the right balance between training time and accuracy. The hybrid Model is the one that uses the most data and works best. The complicated resource footprint in the mesh plot in Figure 3 makes this clear. When healthcare organisations choose a model to use in the real world, they need to consider factors such as HIPAA, GDPR, and local privacy laws. This strongly suggests that you should pick a dynamic model based on how important the use case is and what kind of infrastructure you have. For example, differential Privacy uses math to make sure that the law is followed by keeping patients' names private.

Federated Learning also lowers the risk of audits and liability because it doesn't send unprocessed data. The Hybrid Model leverages these benefits to enhance operations and ensure compliance with rules. Models that keep patients' information private also help build trust, which is important for getting patients more involved in AI-driven healthcare projects. Patients are more likely to agree to share their data if they know that it will be kept private. These models are well-suited for creating environments with extensive data requirements, necessary for training powerful AI models, beyond just technical metrics. But the technology still has some issues. For instance, the Model still needs to be improved so that it can work with encrypted data. Adding noise is an important part of differential Privacy because it can change patterns that are important for making accurate predictions.

Federated Learning makes it harder for models to work together because each node has its own set of data. Secure MPC, on the other hand, takes longer because it needs more than one person to agree. To solve these problems, researchers are looking into adaptive learning schemes, cloud-native optimisations, and model partitioning that takes resources into account. Edge-cloud orchestration, for instance, lets edge devices do some processing before sending encrypted data to the cloud. AI compilers and neural architecture search tools can also automatically create better model versions that protect your Privacy. In conclusion, the discourse, underpinned by empirical data from tables and figures, demonstrates that, notwithstanding the computational overhead, the benefits of privacy-preserving AI models for secure healthcare data analytics significantly outweigh the disadvantages. They make it possible for AI to be used in healthcare in a way that is legal, scalable, and dependable.

6. Conclusion

This study establishes that secure and efficient healthcare data analytics in cloud environments is both feasible and scalable, following a comprehensive evaluation of privacy-preserving AI models. The Hybrid Model is ideal for healthcare apps requiring Privacy, as it leverages Federated Learning, Secure MPC, and Differential Privacy to strike a balance between accuracy, speed, and Privacy. For larger healthcare facilities, the advantages of safeguarding patient data and enabling real-time inference outweigh the necessity for increased processing power. Tables 1 and 2 and Figures 2 and 3 show that no one model is always the best. Homomorphic Encryption is safe, but it's not the best choice if you need answers quickly because it's slower and uses more resources. Differential Privacy is well-suited for early analyses because it requires minimal resources, even though it makes predictions less accurate.

Federated learning is a flexible learning approach that can be applied in decentralised processes, ensuring Privacy while minimising the impact on output. When you pick models in real life, you need to think about the healthcare setting, which includes things like how much computing power is available, the rules that must be followed, and how important it is to make decisions based on what AI tells you. Companies with strong infrastructure can use hybrid models to improve both Privacy and performance. Some people might prefer lighter models like Differential Privacy or Secure MPC because they are easier to understand. It is safe, legal, and moral to look at healthcare data when AI models protect Privacy. These models use AI and keep patient data safe. This is the start of a new era of digital health innovation.

6.1. Limitations

Research on AI models that protect Privacy for cloud-based medical data is limited in many ways, despite promising results. The experiments were initially performed on artificial and semi-real datasets because acquiring real-time patient data proved challenging. These datasets may not accurately represent the diversity of real-world healthcare situations. Because of this, it's hard to use the results in other places and healthcare facilities that have different IT setups. The testing computer environment is based on the idea that there are cloud systems that work well. Small to medium-sized hospitals may struggle to implement the Hybrid Model or other resource-intensive models. This could make it harder to judge how useful something is. Also, the usefulness of some models in real-time may change when network conditions or workload distributions change, which can cause big changes in latency and memory usage metrics.

Federated learning also has the issue of having fixed partitions for datasets. In the real world, data is often not evenly spread out, which can make it hard to train models and get them to converge. There was insufficient research on adaptive rebalancing methods to help address this problem during implementation. The research analysed four primary privacy models, deliberately omitting contemporary paradigms such as quantum-resilient privacy algorithms and trusted execution environments (TEEs). The analysis is constrained by the absence of these advanced methodologies, which could enhance security or scalability. The research does not investigate active threat models or adversarial attacks designed to extract data from model gradients, as it assumes that adversarial contexts are both inquisitive and truthful. These more aggressive threat vectors must be taken into account in future evaluations.

6.2. Future Scope

There are many interesting ways that privacy-preserving AI could be used in the cloud for healthcare research in the future. First, we can test the Model's scalability and ability to work around real-world problems by using it in real time across hospitals that are spread out over a large area. You can learn how models work and what needs to be improved by combining sensor data from Internet of Things (IoT) devices with real patient data from Electronic Health Records (EHRs). Second, adaptive federated learning frameworks can help with data that isn't IID (Independent and Identically Distributed). These frameworks adjust the model architectures based on the performance of the nodes and the quality of the data. Training in different healthcare settings would be better and fairer. Another important trend is the use of cutting-edge security technologies like blockchain and Trusted Execution Environments (TEEs). For regulatory audits, blockchain can create immutable logs, which makes people more responsible. TEEs, on the other hand, can make sure that models run safely even on cloud platforms that aren't trusted.

Quantum computing will eventually make models that use Encryption, like homomorphic Encryption, less safe. So, encryption methods that can stand up to quantum computers are also a big step forward. These algorithms could help keep AI models safe from new threats on the web. We might also be able to conduct more complex tests, such as identifying cancer and predicting pandemics, using multi-modal AI systems that can safely integrate text, image, and signal data. Deep learning architectures prioritising Privacy can utilise compilers and Model optimisers to enhance deployment performance across various cloud settings. Lastly, it will be crucial to test AI models that protect Privacy against attacks and ensure they are robust enough to withstand them. This will also help keep AI-powered healthcare systems safe and reliable, as well as protect Privacy.

Acknowledgment: I sincerely acknowledge Axle Info for providing valuable support, resources, and insights that greatly contributed to this research. Their cooperation and expertise were instrumental in the successful completion of my work.

Data Availability Statement: The data supporting the findings of this study are available upon request from the author. Due to privacy and ethical restrictions, certain data may be limited or require additional permissions. Institutional guidelines and data-sharing policies will review and consider all data requests.

Funding Statement: This manuscript and research work were prepared without any financial support or funding.

Conflicts of Interest Statement: The author declares no conflicts of interest related to this study.

Ethics and Consent Statement: This study was conducted in accordance with ethical standards and approved by the relevant institutional review board. Informed consent was obtained from all participants before their involvement in the study.

References

1. A. E. W. Johnson, T. J. Pollard, L. Shen, L. H. Lehman, M. Feng, M. Ghassemi, B. Moody, P. Szolovits, L. A. Celi, and R. G. Mark, "MIMIC-III, a freely accessible critical care database," *Sci. Data*, vol. 3, no. 5, pp. 1-9, 2016.
2. A. Inan, M. E. Gursoy, and Y. Saygin, "Sensitivity analysis for non-interactive differential privacy: Bounds and efficient algorithms," *IEEE Trans. Dependable Secure Comput.*, vol. 17, no. 1, pp. 194–207, 2020.
3. A. Krall, D. Finke, and H. Yang, "Mosaic privacy-preserving mechanisms for healthcare analytics," *IEEE J. Biomed. Health Inform.*, vol. 25, no. 6, pp. 2184–2192, 2021.
4. A. Majeed, S. Khan, and S. O. Hwang, "Toward privacy preservation using clustering based anonymization: Recent advances and future research outlook," *IEEE Access*, vol. 10, no. 5, pp. 53066–53097, 2022.
5. C. Dwork, N. Kohli, and D. Mulligan, "Differential Privacy in Practice: Expose your Epsilons!," *J. Priv. Confid.*, vol. 9, no. 2, pp. 1-22, 2019.
6. G. S. Nelson, "Practical implications of sharing data: A primer on data privacy, anonymization, and de-identification," *in Proceedings of the SAS Global Forum, SAS Institute Inc.*, Cary, North Carolina, United States of America, 2015.
7. K. Kalodanis, P. Rizomiliotis, and D. Anagnostopoulos, "European Artificial Intelligence Act: an AI security approach," *Inf. Comput. Secur.*, vol. 32, no. 3, pp. 265–281, 2024.
8. L. Xu, C. Jiang, J. Wang, and K. Liu, "A novel privacy-preserving federated learning framework for healthcare data," *IEEE Transactions on Network Science and Engineering*, vol. 8, no. 3, pp. 2050–2064, 2021.
9. M. Shabbir, A. Shabbir, C. Iwendi, A. R. Javed, M. Rizwan, and N. Herencsar, "Enhancing security of health information using modular encryption standard in mobile cloud computing," *IEEE Access*, vol. 9, no. 1, pp. 8820–8834, 2021.
10. M. Zhang, Y. Chen, and W. Susilo, "PPO-CPQ: A privacy-preserving optimization of clinical pathway query for E-healthcare systems," *IEEE Internet Things J.*, vol. 7, no. 10, pp. 10660–10672, 2020.
11. P. R. M. Rao, S. M. Krishna, and A. P. S. Kumar, "Privacy preservation techniques in big data analytics: A survey," *Journal of Big Data*, vol. 5, no. 1, pp.1-12, 2018.
12. P. Singh, G. S. Gaba, A. Kaur, M. Hedabou, and A. Gurtov, "Dew-cloud-based hierarchical federated learning for intrusion detection in IoMT," *IEEE J. Biomed. Health Inform.*, vol. 27, no. 2, pp. 722–731, 2023.
13. S. Wachter, "The GDPR and the AI liability gap: How the proposed EU AI act complements the GDPR," *Computer Law & Security Review*, vol. 43, no. 11, p. 105567, 2021.