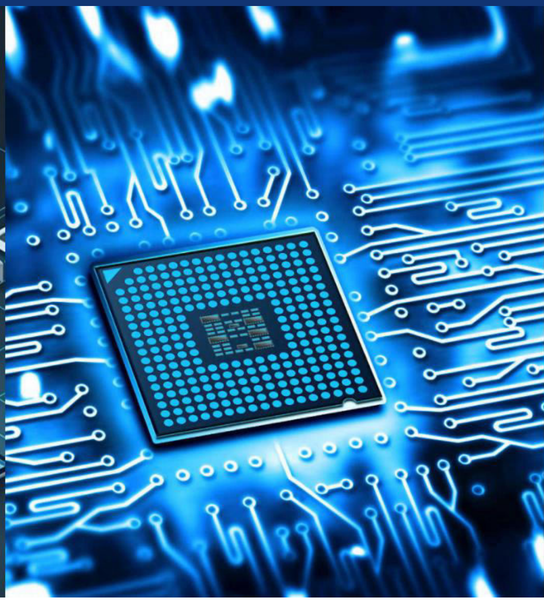
IJIRCCE

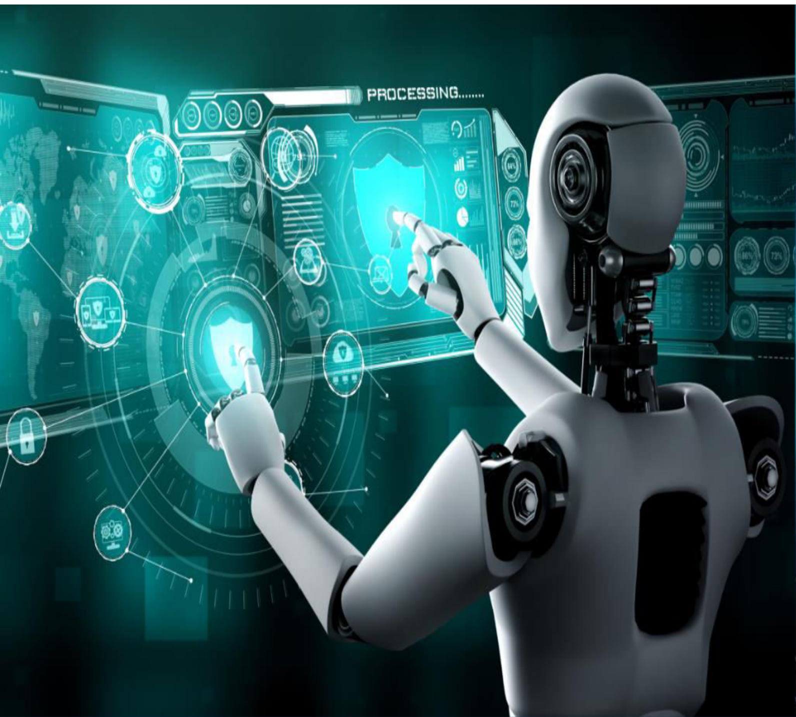
ISSN(O): 2320-9801

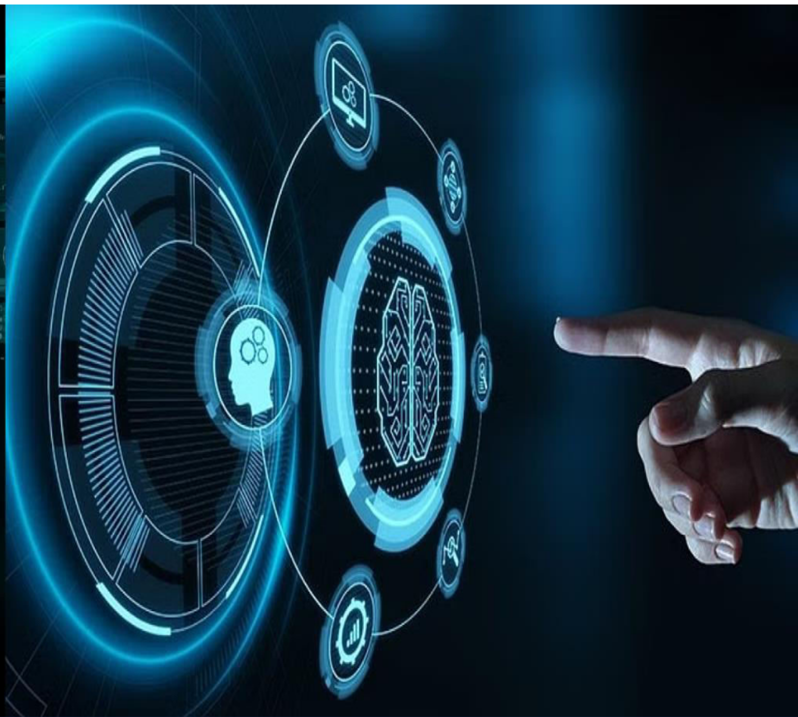
ISSN(P): 2320-9798



International Journal of Innovative Research in Computer and Communication Engineering

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)





Impact Factor: 9.927

Volume 14, Issue 8, August 2026

www.ijircce.com ijircce@gmail.com +91-9940572462 +91 63819 07438

© 2026 IJIRCCE | Volume 14, Issue 8, August 2026 DOI: 10.15680/IJIRCCE.2026.1408009

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

AI for Healthcare: How Can Hospitals Share Data to Build Better Models?

Sejal Gupta

Student, United World Academy, Bengaluru, Karnataka, India

ABSTRACT: The study has explored the importance of sharing data in hospitals for training Artificial Intelligence (AI) models. It has been observed that sharing data in the model training faces challenges such as ethical and privacy concerns, and the revelation of sensitive data of the patients. AI has been used in the detection of skin diseases and in the prediction of cancer from X-Ray images. However, limited data within a single hospital poses challenges and raises data privacy concerns. Investigation on this topic was found to be effective for improving the patient survival and recovery rates. The study has outlined several techniques to address the challenge of data privacy concerns while sharing data for AI training. This includes Federated Learning techniques, Homomorphic encryption, synthetic data generation and differential privacy.

KEYWORDS: Artificial Intelligence (AI), Health Care, Federated Learning, Patients

List of important terms
ADSGAN- Anonymized Data Synthesis using Generative Adversarial Networks (A data generation technique)
AUC- Area Under the Curve (High AUC indicates better classification performance)
C-Index-Concordance Index (Predictive Accuracy used in healthcare and survival analysis)
Brier Score- Evaluate the accuracy of probabilistic prediction.

I. INTRODUCTION

1. 1 Background

Artificial Intelligence (AI) is a branch of science which emphasises building intelligent machines and parts that support performing multiple tasks using human intelligence. Mijwil *et al.,* (2022) highlighted that in the healthcare sector, AI is used to detect, predict and diagnose disease in the areas of neurology, cancer and cardiology. There is a large amount of data in the healthcare sector, which can support decision-making in radiology, ophthalmology, and pathology. Thomason (2024) reported that the healthcare industry generates around 30% of the data volume. This vast amount of data can be helpful in the discovery and understanding of new diseases. In this regard, Enshaei and Naderkhani (2024) noted that AI models are evolving continuously, providing a foundation for decision-making in real-world situations. The quality of data is important for ensuring precise diagnostic, effective treatment plans, and improving patient outcomes. In the health care sector, this is vital for precise diagnostic and effective treatment plans. However, challenges such as ethical dilemmas, regulatory gaps regarding the use of AI, data security concerns, and algorithmic partiality persist in this regard.

1.2 Problem Statement

According to the reports of North (2025), 4.5 million people lack access to necessary healthcare services, and AI can be helpful to close this gap. Sharing data across the hospital to train an AI model is significant for early detection of diseases and obtaining the past disease history of the patients. The UK universities trained the AI software on a dataset of 800 brain scans of stroke patients and then trialled it on 2000 patients. The results were astonishing, with the model achieving high accuracy in identifying the timescale of the stroke. These insights were useful for the treatment of patients so that healthcare professionals can start treatment as early as possible. Despite these benefits, hospitals face major barriers regarding ensuring the privacy of the patients, limited legal compliance and a gap in interoperability between healthcare systems. For instance, the regulations such as the ‘Health Insurance Portability and Accountability Act (HIPAA)’ and the ‘General Data Protection Regulation’ (GDPR) impose strict rules, making healthcare organisations hesitant to share sensitive patient information freely. Norori *et al.,* (2021) argued that training of the AI models on data with limited variation fails to generalise to real-world scenarios. The research is going to investigate the strategic approaches which hospital can use to securely share healthcare data to develop better AI models while

© 2026 IJIRCCE | Volume 14, Issue 8, August 2026 DOI: 10.15680/IJIRCCE.2026.1408009
www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|
International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)
(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

maintaining security, privacy and ethical standards. Additionally, the research will explore the role of modern solutions to eliminate the challenge of data sharing in the healthcare sector.

1.3 Research aims and Objectives

- To evaluate the role of AI in improving healthcare diagnosis, treatment prediction and patient management.
- To explore the main challenges hospitals, face in sharing healthcare data for the training of AI healthcare models.
- To suggest strategies and best practices that hospitals can adopt for safe and collaborative AI-driven healthcare data sharing.

II. REVIEW OF LITERATURE

2.1 Role of AI in improving the healthcare diagnosis, treatment prediction and patient management

Devi *et al.* (2023) argued that the traditional diagnosis therapy by doctors is constrained by factors such as experience, subjectivity and weariness. However, AI has the power to overcome these disadvantages by analysing the vast health records available on Electronic Health Records. Diagnosis of diseases in healthcare is dependent upon medical imaging and in this regard, AI has shown significant potential. AI systems can be trained using different algorithms such as supervised learning, deep learning and time series analysis. For instance, organisations such as Google have used this model across millions of devices without even collecting the raw data of the users on its servers. The advantage of a federated learning model is data privacy. Reduced latency and successful training on diverse datasets. Sun (2024) studied the large language model in the context of medical fields and revealed that the general Large Language Model (LLM) model often fails to give an accurate response. On the contrary, Abdel-Jaber *et al.* (2022) Different AI Models to Improve Healthcare Diagnosis studied the significance of deep learning algorithms to draw conclusions from large data sets. For instance, the model mimics the human brain and automatically learns feature from large datasets. Deep learning model is a convenient tool for diagnosing COVID-19 in patients and its risks. This model takes less than 10 seconds to produce prognostic and diagnostic predictions for the chest CT scans of patients. The rapid development of Artificial Intelligence (AI) in the domain of medical image analysis has helped in combating COVID-19.

2.2 Challenges hospitals face in sharing healthcare data for the training of the AI healthcare models

Youssef *et al.* (2023) stated that the large data sharing initiatives are often performed by a large academic institution from a few geographic locations with access to funding and technical expertise, and funding to share and transform well-curated health data sets responsibly. This underrepresentation in the health data precludes effective generalisation of algorithms to huge populations. However, Alanazi (2023) highlighted the short-term and long-term challenges of AI in the healthcare sector. The short-term challenges in implementing AI systems in healthcare involve data quality concerns, ethical and privacy challenges, limited trust and transparency bias. The long-term challenge involves cybersecurity concerns, cost and resource allocation. Similarly, Li *et al.* (2024) expressed that a deep learning model of AI automatically schedules the appointment for the patients based on their preferences, schedules and medical history. However, compliance of the new AI-driven system with regulations is crucial to improve the trust of the patients and avoid regulatory fines. Additionally, the willingness of the patients to share their data with AI based tool is also crucial in this regard.

2.3 Best practices hospitals can adopt for safe and collaborative AI-driven healthcare data sharing.

Reddy *et al.* (2020) stated that the adoption of strong governance framework is one of the best practices hospitals can use to establish clear policies on the way date of the patients can be stored, collected, shared and accessed among healthcare institutions. The four important components of the governance model are fairness, transparency, trustworthiness and accountability. In addition, the formulation of a governance committee involving the manager, clinicians, patient group representative, and ethical and technical experts to support the development of the AI model has been considered. Another critical practice is data anonymisation and deidentification, were, before sharing information of the patients, hospitals can remove personally identifiable information to eliminate the risk of privacy breach. The advanced privacy-preserving techniques, such as tokenisation and differential privacy is further beneficial in improving security.

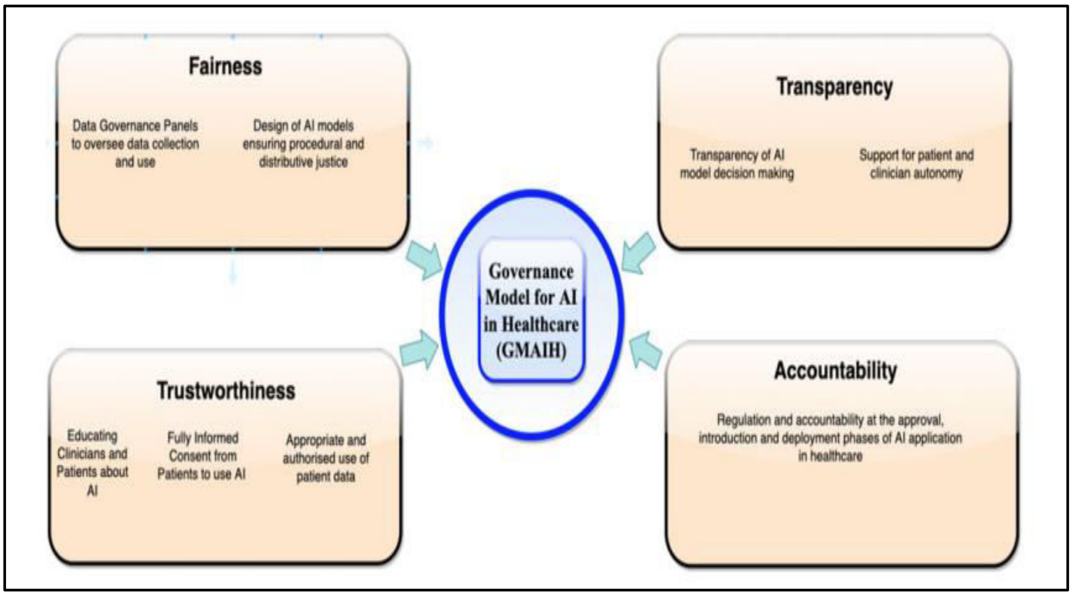


Figure 1: Four principals of AI governance
(Source: Reddy *et al.* 2020)

Karalis (2024) evaluated that achieving clinical practice in AI requires careful planning and adherence to critical principles. This is because a flawed AI model can be responsible for propagating dangerous misinformation

III. METHODOLOGY

3.1 Research Onion

Research onion is a framework used in methodology where data is obtained by opening up the layer from philosophy to data collection. The research onion guides the methodology section of the study through six logical and interconnected stages, such as philosophy, approach, strategy, choice, time horizon and techniques. In this regard, the interpretivism research philosophy was pursued to understand the challenges faced by hospitals in sharing the data for the training of the AI models. William (2024 underscored that interpretivism is a research philosophy that enables humans to gain an understanding of the social realities from the perspective of human beings. Consequently, Ashraf *et al.* (2026) highlighted that an inductive research design is highly flexible in nature and it is a bottom-up approach that emphasises building theories from specific observations. The inductive research design was chosen to analyse the approaches through which data can be shared between hospitals to build better models. The qualitative research strategy was considered to evaluate the best practices hospitals can adopt for sharing data in the training of the AI models. The choice for this research study was a qualitative research choice where data were obtained only from one source to examine the subject matter of interest. As per Lim (2025), the mono qualitative research choice prefers strictly qualitative data to interpret upon subjective behaviour, experience and meaning. Cross sectional qualitative time horizon is effective in obtaining in-depth insight from the available research studies.

3.2 Data Collection

The Study has adopted a systematic Literature Review (SLR) methodology to examine the role of Artificial Intelligence (AI) in the diagnosis of healthcare, treatment prediction, management of patient and healthcare data sharing. Batool *et al.* (2025) opined that SLR methodology enable researchers to systematically collect, explore and synthesise evidence from multiple studies, thereby reducing bias in the literature selection process. In order to search for relevant information, the study will go through three phases such as planning the review, conducting the review and reporting the findings. This method was applied for this research to gather insights in relation to the research questions. This research process was selected to reach the vulnerable populations and gather rich information from historical datasets. This method has been selected to gather data in a short time, thereby enabling researchers to examine available information from a fresh point of view. In this paper, the data have been gathered from credible journals with a DOI

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

number and that are published between 2018 and 2026. Additionally, the information was also obtained from newspapers and credible websites such as BBC, Forbes and The Guardian to provide evidence in support of the claim.

Planning

In the planning phase, the already defined research objectives were translated into Research questions. Hence, the three research questions for this study are as follows:
RQ1: How does AI contribute to healthcare diagnosis, treatment prediction and patient management?
RQ2: What are the challenges do hospital face is sharing healthcare data for AI model development?
RQ3: What strategies and best practices can hospitals adopt to encourage safe and collaborative AI-driven healthcare data sharing?

Search

During the process of finding the answer to the developed research question the search query was performed to retrieve the relevant journal and website articles in Google search Engine. For instance, "Artificial Intelligence in Healthcare", OR "Healthcare Data Sharing", "Privacy preserving machine learning" AND AI diagnosis, "Treatment Prediction OR Patient Outcomes"

Inclusion Criteria:	**Exclusion Criteria**
● Only considering the article, which focuses on healthcare data-sharing mechanisms, privacy and security concerns, and collaborative AI development. ● Papers that are published in English language only.	● Studies unrelated to healthcare and the use of AI models in healthcare were excluded. ● Papers other than English language were excluded.

3.3 Data Analysis

There were 128 research papers found after the application of inclusion and exclusion criteria. After the review only 17 studies were selected based on the review of the title, keywords and abstracts. There is a large body of research relate to AI and its application in the healthcare sector. However, the papers presented in this study were mainly related to the effective AI models for data sharing and its relevance in the healthcare sector. The collected data were analysed using thematic analysis in order to identify recurring patterns and themes. The table below highlights the list of the paper included during the thematic analysis process.

IV. RESULTS

RQ1: Importance of AI models in the betterment of patient health outcomes

Artificial Intelligence (AI) models play a significant role in improving the health outcomes of patients. AI models trained to detect breast cancer, which is not easy to detect easily from open eyes. It has been shown that breast cancer detection improved by 10% with the use of AI. This technological innovation has helped initiate early treatment through screening (Banks, 2026). *Mammography Intelligent Assessment (MIA) was built upon an ensemble of deep learning algorithms* (de Vries *et al.* 2023). In this stance, Wang (2024) mentioned that ensemble algorithms are highly effective in mitigating the concerns associated with model biases. Additionally, the collective decision-making process of the models balances the strengths of the individual algorithms to reduce false-positive and false-negative rates. The integration of AI for breast cancer screening was studied by De Vries *et al.* (2026), who considered 17,421 women attending routine breast cancer screening at NHS UK between February and October 2023. Among them, 10,889 women aged 50-71 years were included in the dataset. The primary AI workflow presented significant improvements in screening performance. The AI-assisted screening system improved the chances of breast cancer detection compared to the standard screening process. Routine screening detected only 9.7 cancers for every 1000 women, while an AI-supported approach detected 10.7 cancers per 1,000 women.

In contrast, Siniosoglou *et al.* (2024) stated that federated learning is a methodology which is aimed at training deep learning (DL) and machine learning algorithms in a decentralised manner so that the key issues commonly found in AI can be resolved. This includes the issue of data privacy and security, model optimisation and resource optimisation. It has been obtained that the *Federated learning algorithm allows data utilisation on local devices.* The deep learning and machine learning models can be trained on dispersed datasets such as medical services without needing centralised

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

collection. For instance, *Kakao Healthcare built a medical data platform using Google Cloud*, where data was standardised in a consistent manner without affecting the existing treatment and work environment of each hospital. It has been reported that the prediction performance for each participant earlier varied from 0.6397 to 0.8362, but the outcome of federated learning increased and reached 0.8482. The model was also tested using the new external data, which AI had never seen before, also known as external validation. On this news, the performance score of 0.7769 was achieved. Although the performance was slightly lower than 0.8482, it still retained 92% of the performance (Google Cloud, 2026). Therefore, the use of the federate learning model was reliable in the context of healthcare diagnosis.

RQ2: Challenges hospitals face in sharing data and training AI models

The study on NHS England in 2023 highlighted that the implementation of AI for the diagnosis of chest health conditions to detect lung cancer in the 66 NHS hospitals met with challenges. The infrastructural challenge limited the successful adoption of this technology inside the hospital premises (News Medical Life Sciences, 2025). The implementation of new technology in the ageing system of the NHS was difficult and met with instances of scepticism and a lack of understanding among healthcare staff. As per the study of Klumpp *et al*. (2024) expressed that application of AI in the existing Electronic Health Record (EHR) systems (EHS) meet with challenges such as interoperability, standardisation and compatibility of healthcare data formats. However, during the pandemic, AI models have helped healthcare institutions with the prediction of infection trends and the spread of the virus. Nevertheless, challenges related to privacy and security in the management of data, specifically in the healthcare domain, still persisted. In a similar stance, outlined that in the field of dentistry and medicine, the clinical judgement-making process is highly personalised. The use and collection of personal information raise concerns regarding privacy, security, and legal responsibility. Training for the AI model, if it lacks compliance with HIPAA security protocols in the US and GDPR regulations in the UK, can violate privacy and integrity (Keith Blankenship, 2025). Besides the issue of data privacy, the other challenge linked to obtaining the data successfully is that old data is not available for training. Models trained on outdated information may fail to answer questions relevant to today's situation (Prescience, 2025). Furthermore, models trained on different languages and variations can lead to inconsistency, specifically for low resource language. Lack of security of AI data models can cause huge loss of data. Only 24% of Gen AI initiatives are secure. Hence, the above discusses the key challenges for hospitals to face in the successful training of AI, including data and privacy challenges, an absence of a governance framework and bias.

RQ3: Strategies to share data

Federate learning is a method where data is shared and trained across multiple institutions without revealing raw data. For instance, K-MELLDODY (Machine Learning Ledger Orchestration for Drug Discovery) is an AI-based drug development system utilised by 10 pharmaceutical companies such as AstraZeneca, Novartis, Bayer, GSK and Johnson & Johnson for the successful discovery of drugs (Melloddy, 2019). This stem has solved the issue of data silos by using Federated Learning (FL). Instead of pulling the confidential chemical datasets into a single server, the system distributes the learning process across the secure infrastructure of the partners. As per the study of Heyndrickx *et al.* (2023) the federated learning algorithm was trained on more than 2.6 billion experimental activity data points over 21 million chemical compounds. In this study the data shared from the ten pharmaceutical companies were utilised. 12.5% of improvement in classification performance was achieved. This demonstrated that federate learning model became significantly better at correctly categorising the compound compared to the model trained by a single company alone. This process helped recognizing the promising drug candidates more precisely thereby making drug discovery more cost effective while keeping proprietary datasets private.

On the other hand, LeewayHertz (2026) explored the concept of *Synthetic Data* for the training of AI models and mentioned it is the most promising privacy preserving approach compared to federate learning. Synthetic data are artificially generated data that mimic the statistical properties and characteristics of real data, while preserving no sensitive or identifiable information. This approach addresses key challenges such as the limited availability of data for training. Increasing importance in the healthcare industry, where privacy is prioritised. The other approach includes implementing differential privacy. *Differential Privacy* is a mathematical framework that ensures the privacy of individuals in the data set is maintained. One of the mechanisms as a part of differential privacy includes adding calibrated noise in the output to ignore the contribution of any possible individual in the data while still preserving the overall accuracy in the analysis. For instance, in a study conducted on 216,714 UK biobank participants who were former or current smokers developed synthetic datasets consisting of 173,371 participants (Qian *et al.* 2024). The results showed that models trained using ADSGAN synthetic data received a C-index of 0.698 compared with 0.823 for

© 2026 IJIRCCE | Volume 14, Issue 8, August 2026 DOI: 10.15680/IJIRCCE.2026.1408009

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

model that were trained on real patient data. The brier score was almost identical highlighting the effectiveness of synthetic patient data for preserving the predictive utility of real healthcare data.

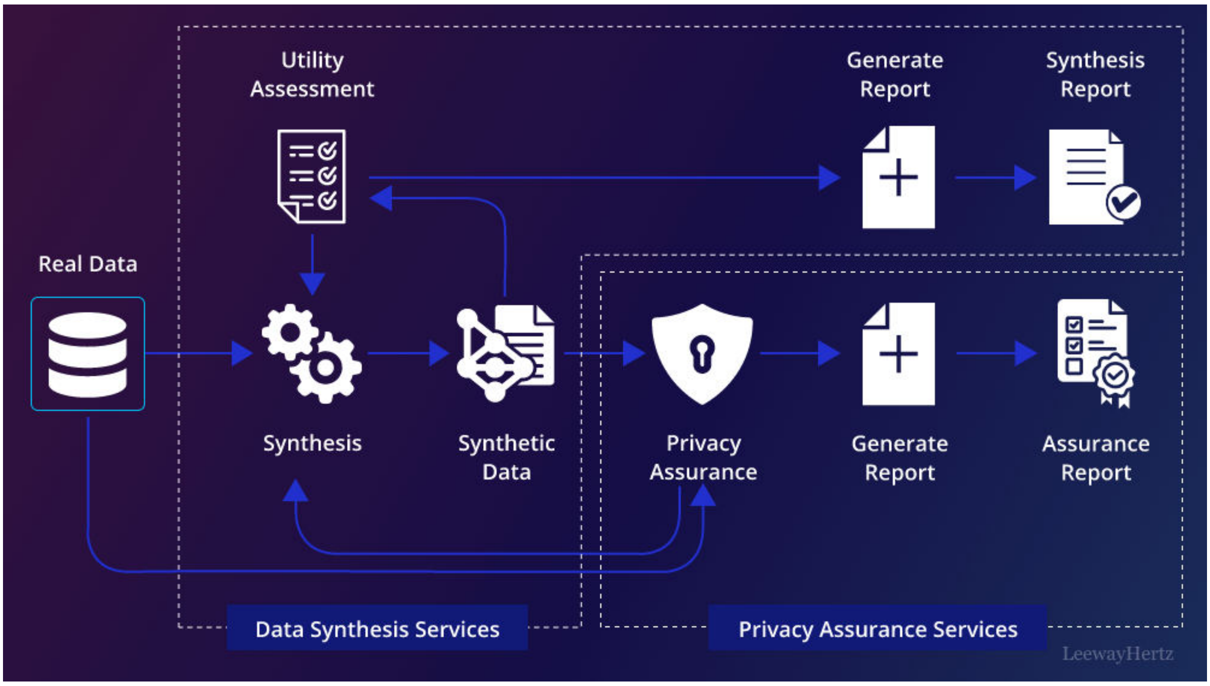


Figure 2: Synthetic Data Technique
(Source: LeewayHertz, 2026)

The other strategic approach is using *Homomorphic Encryption (HE)* in this approach, the algorithm processes data while it still remains completely locked down (Manchester Digital, 2026). In this method, the plain text data is converted into cypher text, which makes the encrypted data hard to read without a proper decryption key. However, efficient key management in this approach is important in the midst of the growing number of datasets. The study has used the data from the Cancer Genome Atlas involving 3,622 samples of patients with a training dataset of 2,713 patients and a testing dataset of 909 patients (Hong *et al.* 2022). The samples have presented 11 cancer types, including bladder, lung, cervical, colon, uterine, ovarian, breast, kidney, liver, skin and stomach. The proposed model has achieved a microAUC score of 0.988 and an overall accuracy of approximately of 85%.

Method	Privacy level	Performance	Cost	Feasibility for hospitals
Federates Learning	High	12.5% in MELLODDY	Moderate	Most Feasible
Synthetic Data	Medium high	Moderate (C-index 0.698 vs 0.823)	Low to moderate	Highly feasible
Homomorphic encryption	Very High	High (microAUC 0.988)	Very high Computational cost	Less feasible

V. DISCUSSION

The findings have revealed that a single hospital might have a limited record of patients, which can limit the successful training of the AI model to predict or diagnose disease. The shortage of workers has made it important to investigate

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

the role of AI technologies in healthcare. Previous literature work by Devi *et al.,* (2023) has supported these findings that traditional diagnosis therapy by doctors is limited by factors such as weariness, subjectivity and experience. Sharing old hospital data for training of AI models meets with several challenges, such as breach of sensitive information of the patients and ethical concerns. In a similar stance, an existing study by Alanazi (2023) argued that ethical, data quality, and privacy concerns come under the short-term challenges of the AI implementation. Nevertheless, the long-term challenges include cybersecurity attacks from hackers, funding and resource allocation constraints. In response to these challenges, findings discussed that techniques such as federated learning can be efficient. The MELLODDY project uses this approach to share the model updates, thereby protecting the privacy of the patients after training the data locally. The findings of other techniques that can help hospitals to share their data include the adoption of the Synthetic data technique, where artificial data is generated to train the model. Furthermore, the differential privacy approach involves introducing noise in the data. However, the existing study by Reddy *et al.* (2020) underscores that the techniques are not enough; integrating AI governance in compliance with government data protection law is important to address the concern of data privacy and ethical concerns linked to the training of AI models. It is implied that the federated learning algorithm currently offers the most effective balance between improvement of model, privacy protection and scalability. Synthetic data and homomorphic encryption further support collaborative AI development by ensuring access to valuable information while mitigating the risk of privacy. From the findings, it was evident better AI model has been possible to build when healthcare organisations share data containing diverse datasets. The evidence highlights that the federated learning algorithm currently offers the most effective balance for model improvement.

VI. CONCLUSION

It can be concluded that the study has identified key barriers such as data privacy and security concerns, cybersecurity risks, lack of infrastructure and governance structure in healthcare data sharing. These insights fulfil objective two of the study by outlining their impact in the successful training of the AI model. Additionally, the objective three of the study was achieved through discussion of several collaborative AI approaches and privacy-preserving techniques. This involves discussing the application of Federated Learning, Synthetic Data, Differential privacy and Homomorphic encryption.

VII. LIMITATION OF THE STUDY

The study has only focused on examining the existing approaches of collecting and sharing data within different hospitals rather than testing a specific intervention experimentally. Furthermore, collecting primary research with hospital and healthcare data would need significant funding, time and institutional support. Hence, secondary qualitative data collection is the limitation of the study.

VIII. CONTRIBUTION OF THE STUDY

The study is contributing to the growing field of the use of AI in healthcare by explaining the way secure data sharing methods can help hospitals to create better AI models by maintaining data privacy and ethical standards.

IX. FUTURE SCOPE

The future scope involves conducting the study on a similar topic through primary research in collaboration with healthcare professionals and hospitals. The future scholars can experimentally test the effectiveness of privacy retaining methods discussed in this study to ensure security, accuracy and effectiveness.

REFERENCES

1. Abdel-Jaber, H., Devassy, D., Al Salam, A., Hidaytallah, L. and El-Amir, M., (2022). A review of deep learning algorithms and their applications in healthcare. *Algorithms*, *15*(2), p.71. Available at: https://www.mdpi.com/1999-4893/15/2/71
2. Alanazi, A., (2023). Clinicians' views on using artificial intelligence in healthcare: opportunities, challenges, and beyond. *Cureus*, *15*(9). Available at: https://pubmed.ncbi.nlm.nih.gov/37842420/

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

3. Ashraf, M.H., Saeed, N., Cullinane, K. and Dresner, M., (2026). Enhancing resilience and sustainability: an inductive approach to innovations in cruise industry. *Transportation Research Part D: Transport and Environment*, *150*, p.105118. https://www.sciencedirect.com/science/article/pii/S1361920925005280

4. Baklola, M., Elmahdi, R.R., Ali, S., Elshenawy, M., Mossad, A.M., Al-Bawah, N. and Mansour, R.M., (2025). Artificial intelligence in disease diagnostics: a comprehensive narrative review of current advances, applications, and future challenges in healthcare. *Annals of Medicine and Surgery*, *87*(7), pp.4237-4245. Available at :https://pubmed.ncbi.nlm.nih.gov/40851938/

5. Banks K. (2026) *My breast cancer was too small to be detected by the human eye - it was found by AI.* Available at: https://www.bbc.com/news/articles/cjd9gn4j7dyo [Accessed on: 13th May 2026]

6. Batool, A., Zowghi, D. and Bano, M., (2025). AI governance: a systematic literature review. *AI and Ethics*, *5*(3), pp.3265-3279. Available at: https://link.springer.com/content/pdf/10.1007/s43681-024-00653-w.pdf

7. de Vries, C.F., Colosimo, S.J., Staff, R.T., Dymiter, J.A., Yearsley, J., Dinneen, D., Boyle, M., Harrison, D.J., Anderson, L.A., Lip, G. and iCAIRD Radiology Collaboration, (2023). Impact of different mammography systems on artificial intelligence performance in breast cancer screening. *Radiology: Artificial Intelligence*, *5*(3), pp.e220146. Available at: https://pubmed.ncbi.nlm.nih.gov/37293340/

8. De Vries, C.F., Lip, G., Staff, R.T., Dymiter, J.A., Tse, B., Ng, A., Fox, G., Oberije, C. and Anderson, L.A., (2026). Prospective evaluation of artificial intelligence integration into breast cancer screening in multiple workflow settings: the GEMINI study. *Nature Cancer*, pp.1-10. Available at: https://www.nature.com/articles/s43018-026-01126-1

9. Devi, K.J., Alghamdi, W., N, D., Alkhayyat, A., Sayyora, A. and Sathish, T., (2023). Artificial intelligence in healthcare: diagnosis, treatment, and prediction. In *E3S web of conferences* (Vol. 399, p. 04043). EDP Sciences. Available at: https://www.researchgate.net/publication/372338887_Artificial_Intelligence_in_Healthcare_Diagnosis_Treatment_and_Prediction

10. Enshaei, N. and Naderkhani, F., (2024). The role of data quality for reliable AI performance in medical applications. *IEEE Reliability Magazine*, *1*(3), pp.24-28. Available at: https://colab.ws/articles/10.1109/mrl.2024.3430192

11. Google Cloud (2026) *kakao healthcare logo.* Available at: https://cloud.google.com/customers/kakao-healthcare-federated-learning [Accessed on: 13th May 2026]

12. Heyndrickx, W., Mervin, L., Morawietz, T., Sturm, N., Friedrich, L., Zalewski, A., Pentina, A., Humbeck, L., Oldenhof, M., Niwayama, R. and Schmidtke, P., (2023). MELLODDY: cross-pharma federated learning at unprecedented scale unlocks benefits in QSAR without compromising proprietary information. *Journal of chemical information and modeling*, *64*(7), pp.2331-2344. Available at: https://pubs.acs.org/doi/pdf/10.1021/acs.jcim.3c00799

13. Hong, S., Park, J.H., Cho, W., Choe, H. and Cheon, J.H., (2022). Secure tumor classification by shallow neural network using homomorphic encryption. *BMC genomics*, *23*(1), p.284. Available at: https://link.springer.com/content/pdf/10.1186/s12864-022-08469-w.pdf

14. Karalis, V.D., (2024). The integration of artificial intelligence into clinical practice. *Applied Biosciences*, *3*(1), pp.14-44. Available at: https://www.mdpi.com/2813-0464/3/1/2

15. Karmitsa, N., Airola, A., Pahikkala, T. and Pitkämäki, T., (2025). A comprehensive guide to differential privacy: From theory to user expectations. *arXiv preprint arXiv:2509.03294*. Available at: https://www.researchgate.net/publication/395244318_A_Comprehensive_Guide_to_Differential_Privacy_From_Theory_to_User_Expectations

16. Keith Blankenship B. (2025) *Regulations and considerations needed to legally use patient data for artificial intelligence (AI) and to align with data protection laws in U.S. and internationally.* Available at: https://www.oralhealthgroup.com/features/regulations-and-considerations-needed-to-legally-use-patient-data-for-artificial-intelligence-ai-and-to-align-with-data-protection-laws-in-u-s-and-internationally/ [Accessed on: 13th May 2026]

17. Klumpp, M., Hintze, M., Immonen, M., Ródenas-Rigla, F., Pilati, F., Aparicio-Martínez, F., Çelebi, D., Liebig, T., Jirstrand, M., Urbann, O. and Hedman, M. (2021). Artificial intelligence for hospital health care: application cases and answers to challenges in European hospitals. In *Healthcare* (Vol. 9, No. 8, p. 961). MDPI. Available at: https://pubmed.ncbi.nlm.nih.gov/34442098/

18. LeewayHertz (2026) *Comprehensive Guide to Synthetic Data: Types, generation, evaluation, use cases and applications.* Available at: https://www.leewayhertz.com/what-is-synthetic-data/#Synthetic-data-generation-process [Accessed on: 13th May 2026]

19. Li, Y.H., Li, Y.L., Wei, M.Y. and Li, G.Y., (2024). Innovation and challenges of artificial intelligence technology in personalized healthcare. *Scientific reports*, *14*(1), p.18994. Available at: https://www.nature.com/articles/s41598-024-70073-7

20. Lim, W.M., (2025). What is qualitative research? An overview and guidelines. *Australasian marketing journal*, *33*(2), pp.199-229. https://journals.sagepub.com/doi/10.1177/14413582241264619

21. Mammen, P.M., (2021). Federated learning: Opportunities and challenges. *arXiv preprint arXiv:2101.05428*. Available at: https://arxiv.org/abs/2101.05428

22. Manchester Digital (2026) *Cryptographic Techniques used in Securing AI Datasets.* Available at: https://www.manchesterdigital.com/post/ve3/cryptographic-techniques-used-in-securing-ai-datasets [Accessed on: 17th May 2026]

23. Melloddy (2019) *New Research Consortium Seeks to Accelerate Drug Discovery Using Machine Learning to Unlock Maximum Potential of Pharma Industry Data.* Available at: https://www.melloddy.eu/new-research-consortium [Accessed on: 13th May 2026]

24. Mijwil, M.M., Faieq, A.K. and Al-Mistarehi, A.H., 2022. The Significance of digitalisation and artificial intelligence in the healthcare sector: A review. *Significance*, *10*(3), pp.25-32. Available at: https://www.researchgate.net/profile/Maad-Mijwil/publication/365114420_The_Significance_of_Digitalisation_and_Artificial_Intelligence_in_The_Healthcare_Sector_A_Review/links/636514ef2f4bca7fd03023f8/The-Significance-of-Digitalisation-and-Artificial-Intelligence-in-The-Healthcare-Sector-A-Review.pdf

25. News Medical Life Sciences (2025) *AI rollout in NHS hospitals faces major challenges.* Available at: https://www.news-medical.net/news/20250910/AI-rollout-in-NHS-hospitals-faces-major-challenges.aspx [Accessed on: 13th May 2026]

26. Norori, N., Hu, Q., Aellen, F.M., Faraci, F.D. and Tzovara, A., (2021). Addressing bias in big data and AI for health care: A call for open science. *Patterns*, *2*(10). Available at: https://pmc.ncbi.nlm.nih.gov/articles/PMC8515002/

27. Ogugua, J.O., Onwumere, C., Arowoogun, J.O., Anyanwu, E.C., Odilibe, I.P. and Akomolafe, O., (2024). Data science in public health: A review of predictive analytics for disease control in the USA and Africa. *World Journal of Advanced Research and Reviews*, *21*(1), pp.2753-2769. Available at: https://pdfs.semanticscholar.org/6375/bb6feec312250cadf2243eb6d0b0c7259402.pdf

28. Prescience (2025) *The Data Problem in Generative AI: Key Challenges and How to Overcome Them.* Available at: https://prescienceds.com/the-data-problem-in-generative-ai-key-challenges-and-how-to-overcome-them/ [Accessed on: 13th May 2026]

29. Qian, Z., Callender, T., Cebere, B., Janes, S.M., Navani, N. and van der Schaar, M., (2024). Synthetic data for privacy-preserving clinical risk prediction. *Scientific Reports*, *14*(1), p.25676. Available at: https://www.nature.com/articles/s41598-024-72894-y

30. Reddy, S., Allan, S., Coghlan, S. and Cooper, P., (2020). A governance model for the application of AI in health care. *Journal of the American medical informatics association*, *27*(3), pp.491-497. Available at: https://pubmed.ncbi.nlm.nih.gov/31682262/

31. Siniosoglou, I., Bibi, S., Kollias, K.F., Fragulis, G., Radoglou-Grammatikis, P., Lagkas, T., Argyriou, V., Vitsas, V. and Sarigiannidis, P., (2024). Federated learning models in decentralized critical infrastructure. *Shaping the Future of IoT with Edge Intelligence*, pp.95-115. Available at: https://pubmed.ncbi.nlm.nih.gov/38564569/

32. Sun, M., (2024). Llamacare: A large medical language model for enhancing healthcare knowledge sharing. *arXiv preprint arXiv:2406.02350*. Available at: https://arxiv.org/abs/2406.02350

33. Thomason, J., (2024). Data, digital worlds, and the avatarization of health care. *Global Health Journal*, *8*(1), pp.1-3. Available at: https://www.sciencedirect.com/science/article/pii/S2414644724000034

34. Wang, L., (2024). Mammography with deep learning for breast cancer detection. *Frontiers in oncology*, *14*, p.1281922. Available at: https://www.frontiersin.org/journals/oncology/articles/10.3389/fonc.2024.1281922/full

35. William, F.K.A., (2024). Interpretivism or constructivism: Navigating research paradigms in social science research. *International Journal of Research Publications*, *143*(1), pp.134-138. https://www.researchgate.net/publication/378396560_Interpretivism_or_Constructivism_Navigating_Research_Paradigms_in_Social_Science_Research

36. Youssef, A., Ng, M.Y., Long, J., Hernandez-Boussard, T., Shah, N., Miner, A., Larson, D. and Langlotz, C.P., (2023). Organizational factors in clinical data sharing for artificial intelligence in health care. *JAMA network open*, *6*(12), p.e2348422. Available at: https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2813016

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi crossref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING