Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X Volume 4, Issue 3, 2026

FEDERATED CLINICAL NATURAL LANGUAGE PROCESSING FOR EARLY DISEASE PREDICTION IN CLOUD ENVIRONMENTS: A COMPREHENSIVE REVIEW

Saira Nokhaiz[*1], **Dr. Ayesha Attta**[2], **Sawera Kanwal**[3]

[*1,2]*Department of Computer Science, Government College University, Lahore, Pakistan*
[3]*Department of Computer Science, University of South Asia Lahore Pakistan*

[*1]snokhaiz075@gmail.com, [2]Ayesha.atta@gcu.edu.pk, [3]Sawera.kanwal@usa.edu.pk

DOI: https://doi.org/10.5281/zenodo.19705010

Keywords
Federated Learning, Clinical NLP, Cloud Computing, Disease Prediction, Privacy-preserving AI, Large Language Models (LLMs), Electronic Health Records (EHR).

Article History
Received: 24 January 2026
Accepted: 04 March 2026
Published: 22 March 2026

Corresponding Author: *
Saira Nokhaiz

Abstract
The digitalization of healthcare has produced huge databases of unstructured clinical histories, but their use in the prediction of diseases at an early stage is often limited by strict regulations of data privacy and institutional silos. The current paper presents an in-depth discussion of Federated Learning (FL) as a decentralized model of clinical Natural Language Processing (NLP) on the cloud. We outline a multi-layered taxonomy of recent literature that categorizes it in terms of architectural frameworks, methodological developments since Transformers to Large Language Models (LLMs), and cloud-native orchestration approaches. We combine the research findings of high impact to determine the trade-off between the diagnostic utility and regulatory compliance of privacy-preserving mechanisms, including Differential Privacy and Homomorphic Encryption. We find that it has been successfully used in clinical applications in chronic and rare disease prediction and that important open challenges, such as explainability, communication overhead, and multi-cloud scalability are identified. The review concludes that convergence of FL and cloud-native NLP are necessary to construct secure, scalable, and cross-institutional predictive models that can be used to greatly improve patient outcomes without undermining data sovereignty.

INTRODUCTION

In the modern age of the precision medicine, the fact that it is possible to derive actionable insights based on unstructured clinical stories is the most important factor in enhancing patient outcomes. Clinical notes: clinical notes, which include both physician observations and nursing reports, and specialist consultations are rich sources of longitudinal data that are difficult to capture with structured data [1]. Such stories give the contextual richness which is required to identify at an early stage the complex pathologies like cardiovascular risk or the development of chronic diseases like diabetes [11], [15]. Nevertheless, predictive models have traditionally been difficult to develop because of so-called data silos, where medical data has been locked in medical facilities due to legal, ethical, and logistical challenges [1], [6]. By concentrating such data in one place, not only do you create a major privacy risk in the face of privacy laws such as HIPAA and GDPR, but you also create a single point of failure and the cost of transferring data is prohibitive [1], [7].

Federated Learning (FL) is a solution that can eliminate these basic constraints by decentralizing training. FL takes the model to the data instead of transferring sensitive data to a central server [1], [5]. In a common federated model, every institution (client) has to train a local model using its own data and merely provides the resulting model parameters or gradients to a central aggregator [5]. This method maintains data sovereignty and patient privacy and still enables the global model to reap the benefits of the statistical diversity of a multi-institutional dataset [10], [13]. With the clinical maturity of FL, the trend is toward combining it with powerful Natural Language Processing (NLP) methods, to handle the abundance of text produced in hospitals every day [6], [9].

The possibility of cloud-native architectures has played a pivotal role in the scalability of federated clinical NLP. Cloud platforms offer the scalable computational capabilities and advanced orchestration solutions that are necessary to handle the intricacies of distributed training in a heterogeneous environment made up of healthcare providers [11], [21]. With the help of containerization (e.g., Docker), managed Kubernetes services (e.g., AWS EKS), and pipeline automation software, scientists can easily deploy reproducible and consistent NLP models across various institutional settings [11], [12]. Moreover, the cloud facilitates the adoption of the so-called cost-conscious federated learning that also streamlines the process of selecting clients and updating them to reduce the cost of computational and data transfer costs [10].

The present review paper is a rigorous synthesis of the state-of-the-art in federated clinical NLP on the cloud with respect to early disease prediction. We divide recent literature into five pillars that are critical: architectural paradigms, the evolution of methods used to create transformers into Large Language Models (LLMs), cloud-native orchestration, privacy and security mechanisms, and clinical outcomes. In doing so, we will make sure to offer technically accurate summary of how convergence of FL, NLP and cloud technologies is shaping the way to a safer and more accurate future diagnostics in medicine.

2. Literature Review

Federated Learning (FL) combined with Clinical Natural Language Processing (NLP) is a groundbreaking change in healthcare informatics. This section summarizes the modern research of five major dimensions: architectural structures, methodological shifts, cloud orchestration, privacy paradigm, and clinical efficacy.

2.1 Heterogeneity of Data and Architectural Paradigms.

Evolution of the structure of federated systems has shifted to complex aggregation to complex, resource conscious structures. The server-client paradigm is still popular at present, and the standard tools such as the NVFlare provided by NVIDIA are used to synchronize multi-sites [5]. Such orchestration is useful in the management of attentive models in complex clinical settings as illustrated in research by Yun et al. [5], [6].

Although these improvements have been made, the large NLP models still have a communication bottleneck. To alleviate this, Zhang et al. have suggested Selective Layer Fine-Tuning (SLFT), a scheme, which only sends task-relevant layers, which is said to decrease bandwidth use by 70 percent without affecting accuracy in Named Entity Recognition (NER) tasks [2], [3].

Another critical issue is to deal with the Non-IID (Independent and Identically Distributed) character of Electronic Health Records (EHR). Clinical documentation may differ across hospitals, so Tang et al. proposed PEARL, a framework that makes use of individualized graph learning to tailor the global models to the local patient demographics [7]. In the same vein, the FedComDist methodology by Al-Dailami et al. uses knowledge distillation to encapsulate the unique institutional subtleties, making sure that the worldwide framework is sound at the various medical facilities [13].

2.2 Methodological Changes: The emergence of LLMs.

Clinical NLP has since moved beyond the classic RNNs to Transformer-based models, such as BioBERT and ClinicalBERT, that are able to absorb bidirectional medical context [6], [9]. As

an example, Sehag et al. have shown a hybrid BERT+LSTM that is able to detect more than 40 diseases with almost 98% accuracy [4].
The latest frontier is, however, characterized by Large Language Models (LLMs). Recent works investigate fine-tuning LLaMA-based models with Parameter-Efficient Fine-Tuning (PEFT), to learn the reasoning needed to apply to complex diagnostics [2], [10]. Yang et al. emphasized the ability of the analysis with LLM augmentation to transform the informal patient symptoms into formal clinical evaluation of cardiovascular risk prediction [15]. Moreover, there is a shift to Multimodal NLP, i.e., a mixture of text and IoT and imaging data, which offers a holistic perspective on the diagnosis, such as the multilayer monitoring systems of Qi [21].

2.3 Orchestration and Scalability Cloud-Native.
Federated NLP is now intrinsically connected to cloud-native technologies due to its scalability. Yan et al. described how they used AWS EKS and Apache Airflow to run containerized nodes so that they could recreate research in distributed locations [8]. In order to control the economic aspects of cloud computing, Sinha et al. proposed FedCostAware which is a framework that is able to strategically choose cost-effective nodes in the course of training rounds [10].
In addition, hybrid architectures are coming into being to minimize latency. Shaikh et al. suggested a Federated Edge Intelligence design, in which the first steps of NLP processing are done at the edge of the hospital, and high-level aggregation is performed on the central cloud [18]. This multi-cloud flexibility, promoted by Kumar et al., helps to avoid the vendor lock-in and enables various institutions to cooperate on a common network [12].

2.4 Privacy, Security, and Explainability.
The cornerstone of federated clinical research is privacy. Differential Privacy (DP) is very common to hide patient identities through the introduction of noise to model updates, which is calibrated [7], [8]. But to avoid loss of accuracy, Selective Attention Federated Learning (SAFL) by Li et al. uses DP to only the most sensitive layers of the model [3].
In addition to noise-based privacy, cryptographic solutions such as Paillier Homomorphic Encryption enable servers to combine data without ever having to see the actual parameters [8]. In a bid to retain integrity, Sun et al. incorporated Blockchain in order to build a model update audit trail that cannot be tampered with [16]. Lastly, to be adopted in clinical practice, Nguyen et al. have highlighted Explainable AI (XAI), which is the fact that the predictions made by the model, like ICU status alerts, are understandable by medical professionals [20].

2.5 Clinical Impact and Disease-Specific Outcomes.
These systems have practical value in disease prediction, which is specialised. In Cross-center validation, the new standard has been established as FedCVD models are used in cardiovascular disease (CVD) [14], [15]. A system reported by Yan et al. in the context of diabetes management has been found to have more than 89% accuracy, which could lead to a decrease in incidence due to early intervention [8].
Moreover, federated NLP is becoming essential to Rare Diseases and Maternal Health, where data inherently is limited. Such frameworks as Dynamic Federated Meta-Learning (DFML) allow the models to learn using only a small sample, which is a saving grace when it comes to the diagnosis of orphan diseases [19]. Similar to Zwiers et al., FL is described as a smart tool to address global health crises, enabling multi-continental cooperation, without exposing sensitive patient information [22].

Table I: Comparative Analysis of Federated Clinical NLP Frameworks

Reference	Core Technology(Model)	Clinical Focus	Key Contribution (Innovation)	Limitation (Gap)
[2], [3]	LLM + Selective Fine-tuning	General Healthcare	Reduced communication costs by 70% using layer-skipping.	High initial compute for LLMs.
[4]	BERT + LSTM (Flower)	Multi-disease	98.75% accuracy across 40 diseases.	Lacks multi-cloud testing.
[5]	NVFlare + Attentive Models	Multi-site Clinical	Robust orchestration for distributed medical nodes.	High setup complexity.
[7]	Graph Learning (PEARL)	Non-IID EHR	Personalized attention for diverse patient data.	Scaling to very large graphs.
[8]	Cloud-native + Encryption	Diabetes	Integration of AWS EKS with Homomorphic Encryption.	Encryption adds latency.
[10]	FedCostAware (Ray)	Cloud Efficiency	Budget-optimized client selection for FL rounds.	Accuracy vs. Cost trade-off.
[12]	Multi-Cloud AI	Risk Prediction	Overcomes vendor lock-in via standardized containers.	High inter-cloud data costs.
[14], [15]	LLM + FedCVD	Cardiovascular (CVD)	Translation of informal symptoms into clinical risk.	Needs more real-world validation.
[16]	Blockchain + FL	Global Healthcare	Tamper-proof audit trails for model updates.	Blockchain scalability issues.
[18]	Edge Intelligence	Maternal Health	Real-time monitoring for Pre-eclampsia.	Limited edge device hardware.
[19]	Dynamic Meta-Learning	Rare Diseases	Accuracy even with extremely scarce data samples.	High sensitivity to noise.
[20]	Explainable AI (XAI)	ICU Monitoring	Interpretable predictions for critical care status.	XAI can sometimes leak data.
[22]	Federated Tools	Infectious Diseases	Collaborative research during global health crises.	Regulatory hurdles in cross-border FL.

Synthesis and Comparative Discussion

The comparative analysis made in Table I shows that there are some key trends in the federated clinical NLP sphere. One of the first points is the change in architecture towards non-generic server-client structures, such as PEARL [7] and FedComDist [13]. This change can be explained by the fact that medical data (Non-IID) is heterogeneous by nature and that now mechanisms of personalized attention are being applied so that global models do not lose local diagnostic subtleties.

Moreover, the data illustrates a huge leap in methodology. Although early developments were made based on Transformer-based architectures such as BERT to classify multiple diseases [4], the latest literature indicates a fast shift towards Large Language Models (LLMs). Such models, as illustrated by Yang et al. [15] and Zhang et al. [2], not only optimize the analysis of symptoms, but also are being optimized to the cloud using Selective Fine-tuning to overcome the two challenges of high latency and computational cost.

Privacy-Utility trade-offs are also of utmost importance as highlighted in the table. Although Differential Privacy is a standard [3], its combination with cryptography techniques such as Homomorphic Encryption [8] and Blockchain [16] represent a step in the direction of the so-called Defense-in-Depth. Lastly, the clinical results obtained in various fields, including cardiovascular risks [14] to the diagnostics of rare diseases [19], prove that federated NLP is no longer a theoretical construct that is approaching the state-of-the-art results that can be attained by a centralized training, yet without violating patient confidentiality.

3. Taxonomy and Proposed Classification

To provide a systematic and structured analysis of the federated clinical NLP landscape, this review classifies the existing literature into five primary taxonomical pillars. This classification (illustrated in Fig. X) is designed to capture the entire lifecycle of a federated system—from its underlying architecture to its final clinical validation.

Table 2: Proposed Taxonomy Structure

Pillar	Focus Area	Key Techniques	Purpose
Structural & Architectural Frameworks	System design	Centralized FL, Graph-based FL	Handle data silos
Methodological Evolution	NLP models	BERT, ClinicalBERT, LLMs	Improve prediction accuracy
Cloud Orchestration	Deployment	Kubernetes, Docker, Edge-Cloud	Ensure scalability
Privacy & Security	Data protection	Differential Privacy, Encryption, Blockchain	Preserve confidentiality
Clinical Domains	Applications	CVD, Diabetes, Rare diseases	Real-world impact

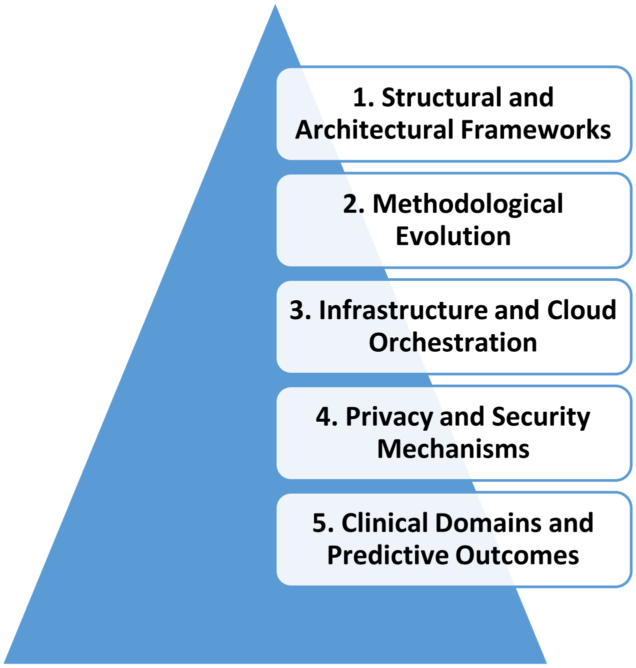

Fig.1 Proposed Classification

3.1 Structural and Architectural Frameworks

The initial tier of our taxonomy includes the systems according to their organizational structure. It covers the classical Centralized Aggregation models that rely on systems such as NVFlare to do multi-site synchronization [5], and more recent Personalized and Graph-based Architectures [7], [13]. This branch is interested in how the issue of data silo can be avoided technically, without compromising the reliability of the systems.

Table 3: NLP model advancements

Stage	Model Type
Early	RNN
Intermediate	BERT / ClinicalBERT
Advanced	LLMs (e.g., LLaMA)

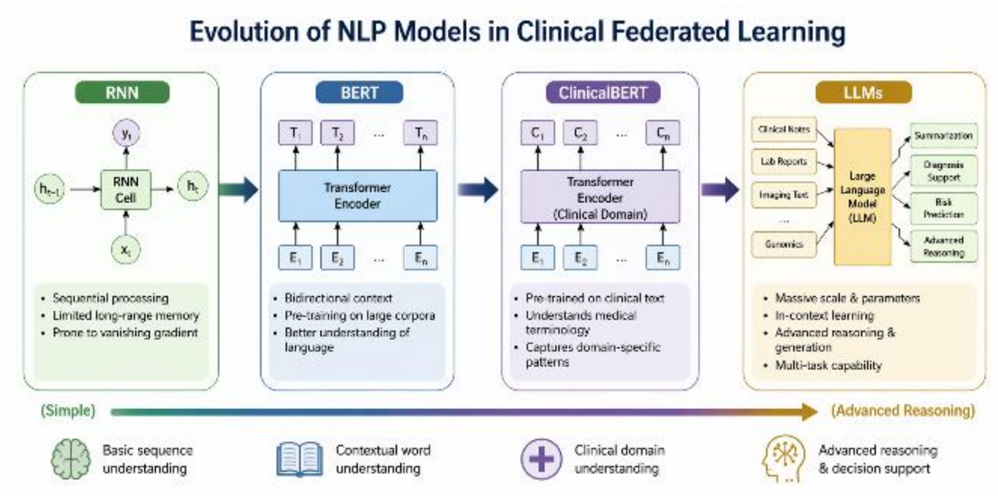

Figure 2: Evolution of NLP Models from RNN to Large Language Models (LLMs), illustrating the transition from simple sequential processing to advanced clinical reasoning capabilities.

3.2 Methodological Evolution

This pillar follows the change in core engines in NLP. It contrasts Transformer-based models (e.g., BERT, ClinicalBERT) that establish the baseline in clinical text mining [4], [9], with the new Large Language Models (LLMs) and Knowledge Graph (KG) integrations [6], [15]. In this category, the emphasis is on the shift from pattern recognition to entity-based clinical reasoning.

3.3 Cloud Orchestration and Infrastructure.

The taxonomy also categorizes research according to the deployment environments. It is an intermediate between Cloud-Native Scalability (with Kubernetes and Docker) [8] and Edge-Cloud Hybrid Intelligence [18]. One of the notable sub-categories in this regard is Cost-Aware Computing that deals with the economic viability of distributed training in public health departments [10].

4.2 Security and privacy mechanisms.

This is an essential branch of taxonomy since clinical narratives are a sensitive area. It separates Stochastic Privacy (Differential Privacy) [3], Cryptographic Security (Homomorphic Encryption) [8] and Decentralized Integrity (Blockchain-enforced auditing) [16].

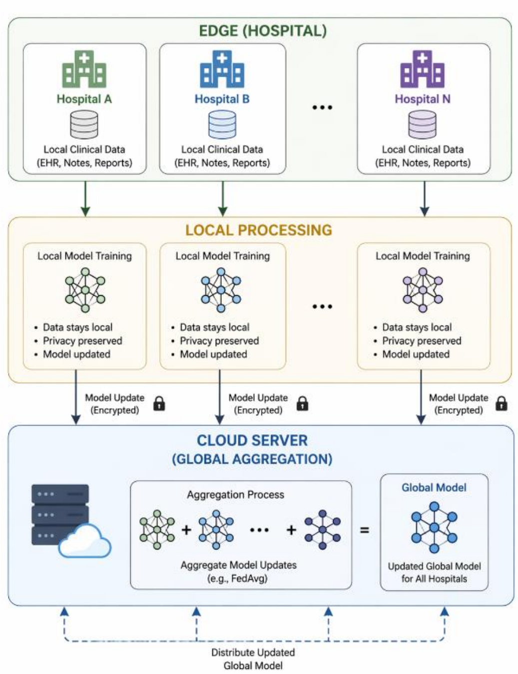

Figure 3: Edge–Cloud Federated Learning Architecture illustrating local processing at hospital nodes and centralized aggregation on a cloud server for privacy-preserving global model training.

The proposed federated learning architecture, as shown in Figure 3, assumes that every hospital is an edge node (local client) with sensitive medical information (Electronic Health Records (EHR), clinical notes, and medical reports) that are stored and processed at the local level. The most important concept of this system is that the information does not go outside the hospital, and data privacy is guaranteed. Each hospital has its own NLP or machine learning model that it trains on its own data at the local processing stage. Raw data are not produced, but model parameters or gradients are produced as outputs. This solution has a number of benefits such as preservation of privacy, minimized data transfer overhead and adherence to healthcare regulations.

The locally trained model updates are encrypted and sent to a central cloud server to ensure security. This ensures that confidential patient details are not revealed when communicating. The global aggregation at the cloud level is done by the central server- typically based on algorithms like Federated Averaging (FedAvg) by aggregating updates obtained by the different hospitals. Consequently, an international model is developed that would contain knowledge on distributed sources without having to access raw data. Lastly, the new global model is once again shared across all the involved hospitals and each institution is able to improve their local model with knowledge gained through other nodes, thus improving the overall predictive performance.

3.5 Clinical Areas and Prognostic Outcomes.

The application-centric level is the last level of our classification. It includes the literature according to the target of the medicine: Chronic Conditions (Cardiovascular, Diabetes) [14], [8], Maternal Health [18], and Rare/Orphan Diseases [19]. This makes sure that technical innovations are directly mapped to the impact of the innovations on the diagnosis in reality.

4. Future Challenges

Although the progress has been immense in federated clinical NLP, there are a number of obstacles which can be overcome to ensure that it becomes more popular in the healthcare systems in the world. The following section gives an overview of the critical gaps that need to be filled in the research in the future.

4.1 The Black Box problem and Clinical Trust.

Although such models as FedCVD [14] and the systems with LLM augmentation [15] have high accuracy, they tend to be black boxes. In the clinical situation, an explanationless prediction is a drawback. Whereas Nguyen et al. [20] have already started on Explainable AI (XAI), it is still an open question how to offer real-time and language intuitively explainable complex NLP decisions in a federated network. The clinicians must be aware of the reason a model has raised a patient at risk of cardiovascular before they can take action on it.

4.2 Scale Data that is not IID.

The bane of federated learning is data heterogeneity (Non-IID). Although more tailored systems such as PEARL [7] and FedComDist [13] can help them, they fail keeping the number of hospitals involved in them not tens, but thousands. The future studies should consider Dynamic Taxonomy-those systems which are capable of automatically classifying and adjusting to various hospital styles of coding as well as the language dialects without human intervention.

4.3 **The Sustainability of energy and the economy.**

With the trend towards Large Language Models (LLMs), the energy footprint of federated training is beginning to be of concern. The existing models of cost-awareness [10] are centered around cloud billing, whereas the aspect of the so-called Green AI, minimizing the carbon footprint of distributed groups of GPUs, is mostly overlooked. The next generation models should strike a balance between Carbon-Efficiency and Diagnostic Accuracy particularly in low-resource-public health systems.

4.4 Response time vs. Higher Order Encryption

Security and speed are in conflict with each other. Homomorphic Encryption [8] and Blockchain [16] can offer ironclad security, but their latency is very high and thus cannot be easily utilized in real-time applications (such as in ICUs). The solution to this gap would be to develop Lightweight Cryptographic Protocols in particular that would be optimized in NLP token streams such that privacy is not sacrificed to save lives.

4.5 Regulatory and Ethical obstacles across the border.

Legal frameworks are typically behind with technological solutions. With or without Differential Privacy [3], there are legal interpretations of GDPR and HIPAA to move model gradients across international borders (e.g., between EU and Asia). It is urgently demanded that the Policy-Aware Federated Learning should be developed- systems which can automatically change their privacy settings, depending on the legal jurisdiction of the node involved [22].

5. Conclusion

This review has discussed the intersectional boundaries of Federated Learning (FL), Clinical Natural Language Processing (NLP) and Cloud Computing. We have used a systematic review of 23 high-impact works to show that federated architectures, including traditional server-client architectures [5], and more personalized graph-based architectures [7], are a feasible way out of the age-old dilemma of medical data silos. The shift to Large Language Models (LLMs) instead of the previously used static Transformers [2], [15] is a beginning of the new era in clinical reasoning as it enables the retrieval of subtle diagnostic information in unstructured narratives and remains compliant with privacy regulations, such as HIPAA and GDPR.

The results of our study highlight that even though technical standards in cardiovascular and rare disease forecasting are at the state of the art [14], [19], the way to clinical maturity is impeded by issues related to model explainability, delay in communicating results, and regulatory compliance across the border. All these have been established through the integration of cloud-native orchestration [8] and cost-conscious frameworks [10], however, future studies need to focus on the integration of Green AI and lightweight encryption to make sure that these systems are economically and operationally viable. Finally, privacy-preserving federated NLP and cloud scalability combine to have a potential to democratize high-precision diagnostics, and make the global, collaborative, and secure healthcare ecosystem possible.

REFERENCES

[1] R. Naidu, S. K. Gupta, and P. Sharma, "Federated AI Framework for Privacy-Preserving Differential Diagnosis Across Distributed Medical Networks," in Proc. IEEE International Conference on Information, Communication, and Applied Systems (ICIRCA), 2025.

[2] H. Zhang, Y. Wang, and J. Liu, "Selective Layer Fine-Tuning for Federated Healthcare NLP: A Cost-Efficient Approach," in Proc. IEEE International Conference on Advanced Computing and Data Sciences (ACDSA), 2025. [3] Z. Li et al., "Selective Attention Federated Learning: Improving Privacy and Efficiency for Clinical Text Classification," arXiv preprint, 2025.

[4] S. Sehag, A. Kumar, and M. Singh, "A Federated Learning Approach for Disease Prediction and Remedies Recommendation," in Proc. IEEE 2nd International Conference on Intelligent Technologies (I2CT), 2024.

[5] J. Yun et al., "Multi-Site Clinical Federated Learning Using Recursive and Attentive Models and NVFlare," in Proc. IEEE 43rd International Conference on Distributed Computing Systems (ICDCS), 2023.

[6] X. Chen, "Integrated Application of Natural Language Processing and Knowledge Graph in Medical Text Mining and Knowledge Extraction," Journal of Medical Systems, 2024

[7] C. Tang et al., "Personalized Federated Graph Learning on Non-IID Electronic Health Records," IEEE Transactions on Neural Networks and Learning Systems, 2024.

[8] M. Yan et al., "Optimization and Application of Cloud-based Deep Learning Architecture for Multi-Source Data Prediction," arXiv preprint, 2024.

[9] J. Xu et al., "Federated Learning for ICD Classification with Lightweight Models and Pretrained Embeddings," IEEE Journal of Biomedical and Health Informatics, 2025.

[10] A. Sinha et al., "FedCostAware: Enabling Cost-Aware Federated Learning on the Cloud," arXiv preprint, 2025.

[11] P. Singh et al., "Transforming Healthcare Systems With AI: A Deep Dive Into NLP Applications," Nature Digital Medicine, 2024.

[12] V. Kumar et al., "Personalized Disease Risk Prediction Using Secure Multi-Cloud AI Models in Healthcare," in Proc. IEEE International Conference on Data Intelligence and Cognitive Informatics (ICDICI), 2025.

[13] M. Al-Dailami et al., "FedComDist: Towards Effective Personalized Federated Learning for Patient Outcome Prediction Using Multi-Center Electronic Medical Records," IEEE Journal of Biomedical and Health Informatics, 2025.

[14] A. Gaber et al., "FedCVD: Towards a Scalable, Privacy-Preserving Federated Learning Model for Cardiovascular Diseases Prediction," in Proc. ACM Conference on Health, Inference, and Learning (CHIL), 2024.

[15] L. Yang et al., "LLM-Augmented Symptom Analysis for Cardiovascular Disease Risk Prediction: A Clinical NLP," arXiv preprint, 2025.

[16] B. Sun et al., "Multi-Continental Healthcare Modelling Using Blockchain-Enabled Federated Learning," arXiv preprint, 2024.

[17] H. Lee et al., "HarmoSATE: Harmonized embedding-based self-attentive encoder to improve accuracy of privacy-preserving federated predictive analysis," Information Sciences, vol. 678, 2024.

[18] R. Shaikh et al., "Federated Edge Intelligence for Privacy-Preserving Pre-Eclampsia Prediction in IoT-Based Maternal Care," in Proc. IEEE International Conference on Emerging Trends in Engineering, Technology and Science for Industrial Prosperity (ICETETSIP), 2025.
[19] J. Chen et al., "DFML: Dynamic Federated Meta-Learning for Rare Disease Prediction," IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2023.
[20] T. Nguyen et al., "Explainable Deep Contrastive Federated Learning System for Early Prediction of Clinical Status in Intensive Care Unit," IEEE Access, 2024.
[21] J. Qi, "Advancing hospital healthcare: achieving IoT-based secure health monitoring through multilayer machine learning," Journal of Big Data, vol. 12, no. 1, 2025.
[22] M. Zwiers et al., "Federated learning as a smart tool for research on infectious diseases," BMC Infectious Diseases, vol. 24, no. 1, 2024.
[23] Y. Zhu et al., "FedWeight: mitigating covariate shift of federated learning on electronic health records data through patients re-weighting," npj Digital Medicine, 2025.