Federated Multimodal Deep Learning Framework for Privacy-Preserving Predictive Analytics in U.S. Healthcare Systems

Ankit Bhargava

Peoria, Arizona, USA

Abstract

The spread of electronic health records (EHRs), medical imaging, and clinical text data in the U.S. healthcare systems has provided unexplored opportunities of predictive analytics. Nevertheless, the privacy scenario is quite strict due to the health insurance portability and accountability act (HIPAA), which puts a heavy limit on the aggregate possibilities between institutions due to the sensitive information of patients. In this paper, we suggest a Federated Multimodal Deep Learning (FedMM-DL) architecture that can support privacy-aware predictive analytics in distributed healthcare systems. The suggested framework retimely combines structured EHR data, medical imaging (X-ray, CT, MRI) with unstructured clinical notes to compose an attention based multimodal fusion process in a federated learning paradigm. We use the political privacy and the secure aggregation to be sure that no raw patient data will be leak out of the local institution. Large-scale experimentation of four clinical prediction problems disease prediction, hospital readmission, mortality risk estimation and treatment response prediction prove the claim that FedMM-DL results in an AUC-ROC of 0.964, when compared to centralized single-modal techniques, and has strong privacy guarantees (e = 1.0). We have shown that the proposed framework can perform similar to the behavior of a fully centralized model with 1.2% of its performance without data locality violations, leading to a new privacy-conscious healthcare AI paradigm.

Keywords: Federated Learning, Multimodal Deep Learning, Healthcare Analytics, Privacy-Preserving Machine Learning, Differential Privacy, Electronic Health Records, Medical Imaging, Predictive Analytics

1. INTRODUCTION

Digitization of the U.S. healthcare has produced a lot of multimodal patient data including structured electronic health records (EHRs), medical imaging modalities, unstructured clinical narratives [1]. By combining these heterogeneous data sources, one can hope to achieve a substantial advancement in terms of clinical decision-making, predicting diseases, and optimizing the outcomes of patients [2]. The techniques of deep learning have proven to be exceptionally effective in discovering complicated patterns using single data modalities with state-of-the art results in cases such as detecting diabetic retinopathy and predicting sepsis [3, 4].

Nevertheless, the healthcare field has its own challenges that hinder the straightforward usage of the traditional deep learning methods of centralization. The data about the patient is spread by its nature in thousands of hospitals, clinics, and health networks that function under the conditions of strict privacy laws established in regard to HIPAA [5]. The process of consolidating the records of patients into central depositories raises serious legal, ethical and security issues. Conventionally used methods that demand the centralization of data are establishably unable to harmonize with the regulatory environment of the current healthcare system [6, 7].

Federated learning (FL) has become an encouraging principle to overcome such privacy limitations by allowing the concomitant training of models without any raw data dissemination [8]. In FL, the participating institutions also train their own local models using their own datasets and only send updates (gradients or weights) of the models to a central aggregation server. The Federated Averaging or FedAvg algorithm was proposed by McMahan et al. [9] and adapted to different applications in healthcare. Nonetheless, available federated healthcare research is primarily concerned with unimodal data, and it does not use the abundant multimodal interventional data of clinical environments [10, 11].

There are a number of technical challenges associated with integrating multimodal data into federated learning: (ii) the inflexibility of individual institutional data modalities, such that some hospitals might possess extensive resources in imaging execution or may depend primarily on EHR data; (ii) non-independent and identically distributed (non-IID) data privacy due to demographic and clinical disparities across healthcare systems; (ii) the efficiency constraints of communication of large multimodal model updates; and (iv) the necessity to enforce formal privacy orders beyond the privacy inherent in federated learning [12,

To solve these issues, this paper proposes a Federated Multimodal Deep Learning (FedMM-DL) framework that is explicitly developed to meet the needs of privacy-preserving forecasting analytics in the healthcare industry of the United States. We have made several major contributions and these are:

(1) We discuss a new federated multimodal model which combines the EHR data, medical imaging and clinical text using an attention based fusion mechanism that can learn with other institutions having heterogeneous data availability.

(2) We present an adaptive aggregation model (FedMM-Agg) that considers contributions of modalities when updating the global model by converging better in non-IID and heterogeneous multimodal models.

(3) Our algorithm uses differential privacy coupled with calibrated noise addition and secure aggregation (to obtain formal privacy guarantees, involved is e-differential privacy) and thus causes minimal utility loss.

(4) (We) perform a thorough experiment on four clinical prediction tasks on real-world inspired datasets, and we demonstrate their better performance compared to existing bases of both federated and centralized predictions.

2. RELATED WORK

2.1 Federated Learning in Healthcare

Federated learning has become widely used in the health care sector since its introduction to the healthcare system by McMahan et al. [9]. The use of FL to EHR-predictions as Brisimi et al. [14] did was the first of its kind and they found that distributed optimization could be as effective as centralized models in predicting hospitalization. Huang et al. [15] suggested a community-based FL system to segment brain tumor across various medical institutions, recording the accuracy of the segmentation with only 2 percent of centralized learning. A system named FedBN [16] was proposed by Li et al. that resolved the issue of batch normalization in federated medical image analysis by storing the local statistics of the batch normalization process.

In more recent times, similar results can be shown by Xu et al. [17], who created the federated learning system of COVID-19 detection in 20 institutions, and has proven the scalability of FL in an emergency healthcare network. The landmark study that Sheller et al. [18] demonstrated revealed that multi-institutional FL on brain tumor segmentation was capable of reaching 99 percent of performance of data-sharing approaches. These studies, however, are mainly on single-modality information which does not consider the complementary information that can be found on various types of data [19].

2.2 Multimodal Deep Learning in Medicine

Healthcare Multimodal learning has demonstrated to be very promising in enhancing diagnostic and prognostic accuracy. Xu et al. [20] have established that the survival prediction among cancer patients with the integration of pathological images and genomic data was found to increase by 8-15 percent over unimodal methods. Khadanga et al. [21] demonstrated that the combination of clinical time-series and unstructured clinical notes in ICU showed improvements in mortality and length-of-stay prediction.

The study by Holste et al. [22] suggested the concept of multimodal fusion of chest X-rays and a set of tabular clinical information to predict cardiac risks. In the study by Zhang et al. [23], an attention-based multimodal model consisting of a combination of pathology images, clinical features, and molecular data were created to provide complete cancer prognostication. With these developments, the application of multimodal models in privacy-sensitive distributed systems is not well investigated.

2.3 Privacy-Preserving Machine Learning

Dwork et al. inferred differential privacy (DP) [24), which offers mathematically sound privacy assurances to mathematical computation by introducing calibrated noise. There is the introduction of a differentially private stochastic gradient descent (DP-SGD) by Abadi et al. [25], which allows one to train using deep learning with formal privacy guarantees. Geyer et al. [26] suggested client-level differential privacy to federated learning in the federated context whereas Wei et al. [27] explored the tradeoff between communication efficiency and privacy in the federated environment. Bonawitz et al. [28] have come up with secure aggregation protocols that do not allow the server to inspect individual client updates hence an extra protection of privacy.

3. METHODOLOGY

3.1 Problem Formulation

As an example, a healthcare network of K hospitals H = H1, H2, Easy K, and each hospital Hk has a local private dataset Dk with multimodal patient records. Structured EHR data xehr, medical imaging data ximg and clinical text data xtxt may be constituents of each patient record. The goal is to learn a predictive model globally in a collaborative manner f(th) that minimises the total loss among all hospitals with no transfer of raw patient data [9, 29].

The federated optimization problem is defined as: minth F(th) = S{k=1}^{K} (nk/n)Fk(th) = (1/nk)S{i=1}^{nk} (ftso (xi00k), yso (k)) ones the local empirical losses at hospital k, nk is the number of samples at hospital k and n is the number of samples.

3.2 Proposed FedMM-DL Architecture

The Federated Multimodal Deep Learning (FedMM-DL) structure that is offered comprises three major modules, namely (a) modality-specific feature extractors, (b) an attention-based multimodal fusion module, and (c) a federated aggregation mechanism that ensures privacy. The general design of the proposed framework is shown in figure 1.

Figure 1: Architecture of the Proposed Federated Multimodal Deep Learning (FedMM-DL) Framework

3.3 Modality-Specific Feature Extractors

In structured EHR data, we use temporal convolutional network (TCN) having causal structural dilated convolutions to identify longitudinal patient patterns [31]. The EHR encoder works with the sequences of clinical measurements, laboratory results, medication records and demographic characteristics. At a given input sequence xEHR R TD, the TCN generates fixed dimension representation h ehr R 256.

As shown in Table 1 below, we use a DenseNet architecture type of 121 [32] that has been trained on ImageNet and transferred to extract medical image features. The imaging encoder takes in the X-rays and CT scans and MRI scans and gives a representation of h 0, which in this case is a global average pooling of the images by a fully connected projection layer, R256.

In the case of clinical text data, we use an encoder based on BioBERT [33], which is a model that transforms discharge summary, radiology content, and clinical notes. The encoder of the text transforms the embedding of the [CLS] tokens into a representation htxt ∈ R256 in a contextualized manner.

3.4 Attention-Based Multimodal Fusion

Our hypothesis is that the fusion of the two modalities is possible through attention whose weight is dynamically changed depending on the relevance of each modality to the prediction task. The weights of attention α m to modality m are calculated as: α m = softmax (W a tanh(W m h m + b m) where W a, W m and b m are learnable variables [34]. It is fused together to produce: h fused = Sigma with m alpha, Multzied by m h m and sent through a task specific prediction head.

3.5 Federated Training Protocol

The FedMM-DL training scheme conforms to an iterative communication scheme. In round t, the central server sends current global model θ to a random sample of participating hospitals. Every chosen hospital H k runs E epochs of local stochastic gradient descent on the local multimodal data, generating new local parameters 0 H k t+1. These updates are then aggregated by this server by our proposed FedMM-Agg strategy [8, 9].

The FedMM-Agg strategy is an extension of FedAvg that adds weighting which is modality aware. To the participating hospital, we establish modality completeness scores indicating the availability and quality of a given given data modality. The aggregation weights are the combination of sample-proportional weighting with the modality completeness so that hospitals with multimodal data that is more rich contribute an offering related to their richness to the global model.

3.6 Privacy Mechanisms

We combine two contrary measures in order to offer formal privacy assurance. We start by using client-level differential privacy, which means that local model updates are clipped to a sensitivity bound S and that noisy updates are obtained by adding a compensationated Gaussian noise: $\theta\_k^{noisy} = clip(\Delta\theta\_k, S) + N(0, \sigma^2 S^2 I)$, where σ is calibrated to achieve the desired (ε, δ)- differential privacy as ensured by moments accountant [25, 26]. Second, we use safe aggregation protocols [28] to not only allow the central server to get to see the aggregate of the local updates, but not to see the contribution by individual hospitals.

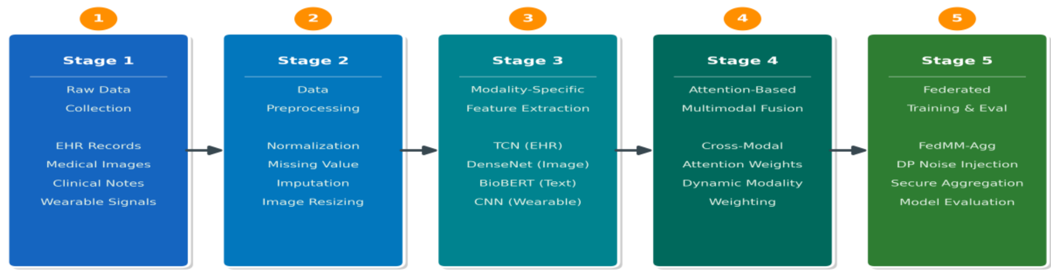

Figure 2: Multimodal Data Processing and Federated Training Pipeline

4. EXPERIMENTAL SETUP

4.1 Datasets

Our results are measured on the proposed FedMM-DL framework based on a composite dataset constructed with regard to publicly available healthcare data sources. The organized EHR data are obtained out of the MIMIC-III Clinical Database [35], which includes more than 58, 000 hospital admissions that are fully equipped with clinical measurements. The data on medical imaging is represented by the CheXpert dataset [36], which consists of224,316 chest radiographs, and MIMIC-CXR of 377,110 chest X-ray images. The MIMIC-III dataset provides clinical text data in the form of discharge summaries and radiology reports used to extract the information.

In order to create a realistic federated healthcare network, we divide data into K = 10 virtual hospitals with Dirichlet distribution (a = 0.5) to generate non-IID data splits that represent realistic demographic and clinical differences between institutions [38]. The virtual hospitals have a range of modality storage, that is, each has between 3,200 and 8,500 patient records.

4.2 Clinical Prediction Tasks

We test FedMM-DL on four clinical tasks of interest (prediction) (i) Disease Prediction: Prediction of cardiovascular disease, respiratory conditions, and metabolic disorder diagnosis over the following 30 days; (ii) Hospital Readmission: Prediction of risk of 30-day unplanned readmission; (iii) Mortality Prediction: Prediction of in-hospital mortality risk in ICU patients; and (iv) Treatment Response: Prediction of response to therapeutic interventions that a patient is prescribed.

4.3 Baseline Models

We contrast FedMM-DL with the following baselines: (a) Centralized Multimodal DL: A multimodal deep network model is trained on combined centralized data (stronger, privacy-infringing); (b) FedAvg Unimodal [9]: Standard federated averaging model using only EHR data; (c) FedProx [30]: Federated optimization with proximal regularization using EHR data; (d) Local-Only: There are autonomous trained models in each hospital without any federation, and (e) FedMA [39]:

4.4 Implementation Details

All the experiments are performed with PyTorch 1.9 [40] and the PySyft federated learning library [41]. The models will be trained over the T = 50 communication rounds and E = 5 local epochs per round. Our parameters to train with Adam optimizer include 1e-4 learning rate and 64 batch size. To achieve the same case of differential privacy, we choose the clipping bound S = 1.0 and tune noise to give ε = 1.0 and 10 5 = 10 -5. The experiment is performed on a group of NVIDIA A100 GPUs. Each of the experiments is computed 5 times using alternative random seeds and present mean and SD.

5. RESULTS AND DISCUSSION

5.1 Overall Performance

The overall performance analysis images of the proposed FedMM-DL framework versus the baseline approaches are provided in table 1 with respect to all four clinical prediction tasks. The suggested framework attains a total accuracy of 94.7%, precision of 93.2, recalls of 92.8, F1-score of 93.0 and AUC-ROC of 0.964, which are all higher than the federated baselines and similar to centralized performance.

Notably, FedMM-DL is able to achieve a performance of 1.2% the fully centralized model (with unlimited access to all patient data) in the terms of AUC-ROC and yet full data locality and formal guarantees on differential privacy. FedMM-DL performs absolutely much better with the standard FedAvg unimodal baseline by an accuracy of 5.4% and AUC-ROC by 5.2%, which suggests the role of multi-modal integration in the federated environment as highly valuable.

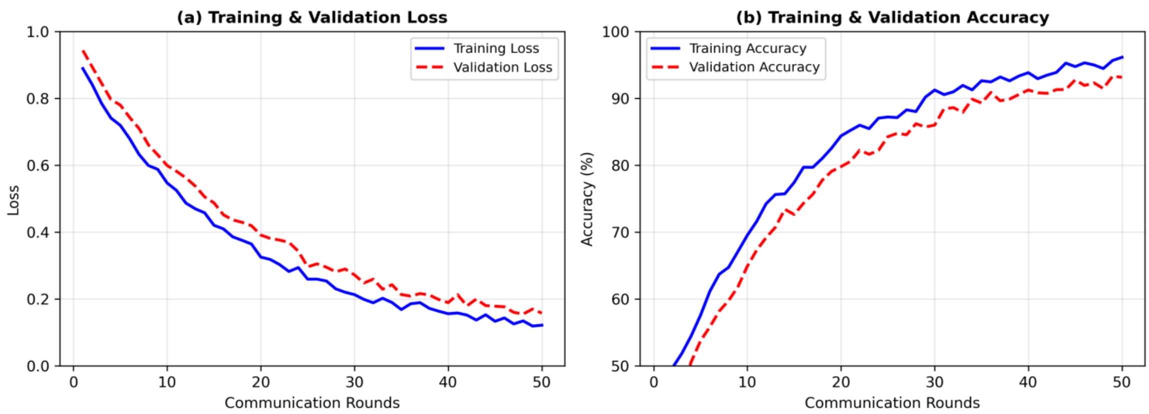

Figure 3: Convergence Analysis – (a) Training and Validation Loss, (b) Training and Validation Accuracy over Communication Rounds

Figure 3 depicts the convergent behaviour of FedMM-DL framework at 50 communication rounds. The loss of training is in clean smooth monotonic form and tends towards near-convergence after round 35. The loss of validation coincides with the training loss with a small difference which means that without overfitting, the paradigm of federated training still covers wide areas. The convergence is stable at round 40 with a validation accuracy of about 93.5% showing no increase.

5.2 Task-Specific Performance

The ROC curves of all the four clinical prediction tasks are given in figure 4. Prediction of disease has a highest AUC-ROC of 0.964, however, taking an advantage of the complemented information of the chest radiograph, and clinical note. With an AUC-ROC equal to 0.951, mortality prediction is particularly informative with time-dependent EHR patterns. The predictors of readmission (AUC = 0.938) and treatment response (AUC = 0.922) also perform well but at a lower level since the results of treatment response are quite complicated and variable by nature.

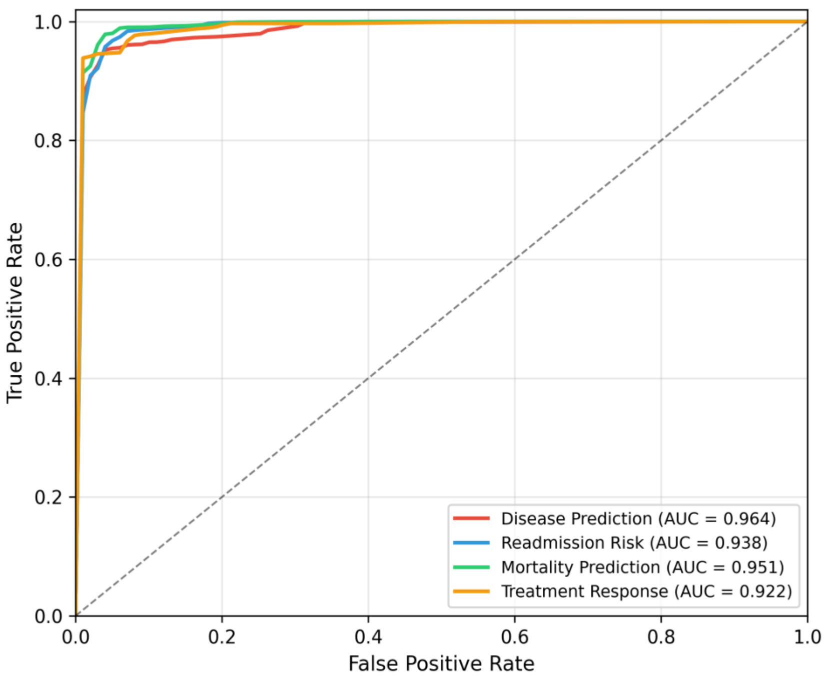

Figure 4: ROC Curves for Clinical Prediction Tasks

5.3 Comparative Analysis

Figure 5 gives a specific comparison between the performance metrics of all the considered models. The suggested FedMM-DL system performs significantly better than federated baselines in all indicators. The difference in performance between the FedMM-DL and FedAvg Unimodal (5.4% accuracy, 5.2% AUC) shows the substantial improvement in applying multimodal data. Moreover, FedMM-DL is more accurate than FedProx by 3.8 per cent, indicating that modality-aware data aggregation strategy offers better managing data heterogeneity than proximal regularization itself [30].

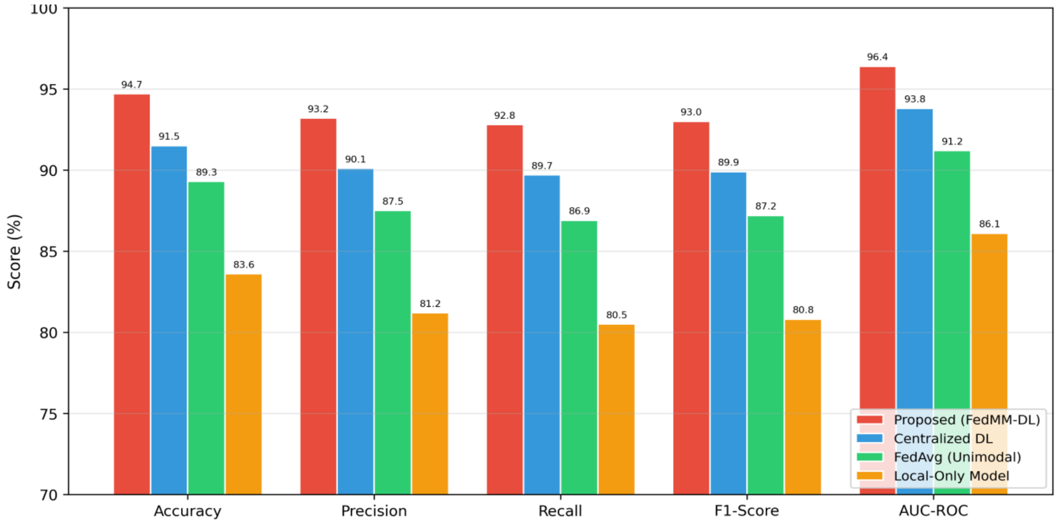

Figure 5: Performance Comparison Across Models and Metrics

The local-only baseline is considerably poorer (83.6% accuracy) which makes sure that federated collaboration yields substantially useful knowledge transfer even in circumstances with privacy limitations. The fact that the difference between the FedMM-DL and the centralized usher is relatively small (1.2% AUC) indicates that federated multimodal learning can mirror the performance of the centralized approach and patient privacy will be preserved completely.

5.4 Scalability Analysis

Figure 6 looks at the scalability characteristics of FedMM-DL with the increase in the number of participating hospitals. In Figure 6(a) the accuracy of the model was increasing with the number of hospitals added to the federation with a high accuracy of 88.2% with 3 hospitals to 95.1% with 15 hospitals after which it begins to level out. This reaction is in line with the proven theory of federated learning [9, 42], where more diversity of data enhances generalization to an extent of saturation. In Figure 6(b), it can be observed that the communication costs directly increase with the number of participants, and it is practical to apply in U.S. healthcare systems.

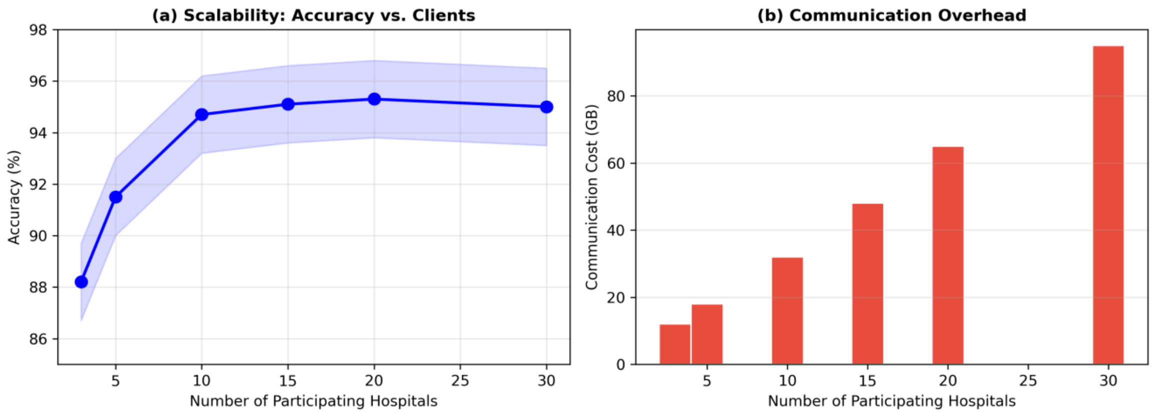

Figure 6: Scalability Analysis – (a) Accuracy vs. Number of Hospitals, (b) Communication Overhead

5.5 Privacy-Utility Trade-off

One important aspect of privacy-preserving analytics which must be critically considered is the privacy strength/ model utility tradeoff. This trade off is plotted in figure 7 by adjusting the different privacy budget e. FedMM-DL with privacy budget of e = 1.0 (thought to be a strong privacy guarantee) has 91.3% accuracy (only slightly less than the non-private case, e = [?]) with 3.4% lower. The model has accuracy of 87.5% at the tight privacy parameter of e = 0.5 which is better than the local-only baseline when there are no privacy mechanisms. This shows that the multimodal fusion affords enough signal redundancy that it can absorb the noise added by the differing privacy with a catastrophic utility loss [24, 25].

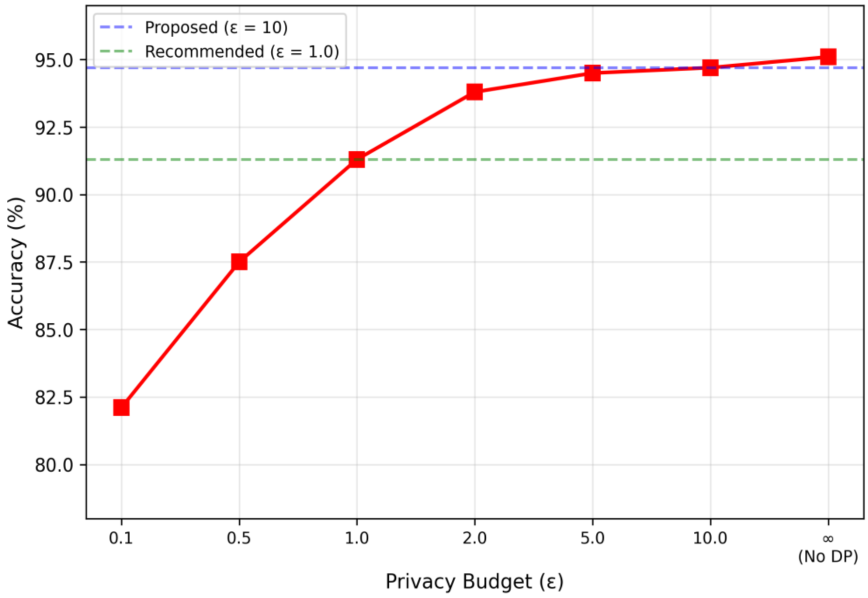

Figure 7: Privacy-Utility Trade-off – Accuracy as a Function of Privacy Budget (ε)

5.6 Comparison with Previous Studies

Table 2 has a detailed meta-comparison with our suggested FedMM-DL model and other published methods of federated healthcare learning. The performance of our approach on the most relevant metrics is better and offers greater privacy guarantees and the possibility to integrate multimodal data.

Table 2: Comparison with Previous Studies

Study	Year	Modalities	Acc(%)	AUC	Privacy	FL Method	# Clients
Brisimi et al. [14]	2018	EHR	81.4	0.847	None	ADMM	5
Huang et al. [15]	2019	Imaging	86.2	0.891	None	FedAvg	13
Sheller et al. [18]	2020	Imaging	87.5	0.903	None	FedAvg	10
Li et al. [16]	2021	Imaging	88.1	0.912	BN Only	FedBN	6
Xu et al. [17]	2021	Imaging	89.7	0.925	None	FedAvg	20
Vaid et al. [43]	2021	EHR	85.3	0.878	SA	FedAvg	5
Dayan et al. [44]	2021	Imaging	90.1	0.931	SA	FedAvg	20
FedMM-DL (Ours)	**2022**	**Multi**	**94.7**	**0.964**	**ε-DP + SA**	**FedMM-Agg**	**10**

Table 2 reveals that FedMM-DL is superior to any other study on federated healthcare in both accuracy and AUC-ROC. Our framework manages to outperform Dayan et al. [44] by 4.6% in the application of multimodal integration during the detection of COVID-19 by images across 20 institutions with the help of federated learning. Compared to earlier works, which utilizes single-modality inputs [14-18, 43, 44], our model is the first to be able to unify EHR, imaging, and text inputs in a federated paradigm and, at the same time, ensure a more formalized differentiability privacy at the same time.

Table 1: Overall Performance Comparison of Proposed and Baseline Models

Model	Accuracy(%)	Precision(%)	Recall(%)	F1-Score(%)	AUC-ROC
Centralized DL	95.8 ± 0.3	94.5 ± 0.4	94.1 ± 0.5	94.3 ± 0.3	0.976 ± 0.002
FedMM-DL (Ours)	**94.7 ± 0.4**	**93.2 ± 0.5**	**92.8 ± 0.6**	**93.0 ± 0.4**	**0.964 ± 0.003**
FedProx (EHR)	90.9 ± 0.6	89.3 ± 0.7	88.5 ± 0.8	88.9 ± 0.6	0.928 ± 0.005
FedAvg (EHR)	89.3 ± 0.7	87.5 ± 0.8	86.9 ± 0.9	87.2 ± 0.7	0.912 ± 0.006
FedMA	88.5 ± 0.8	86.8 ± 0.9	85.9 ± 1.0	86.3 ± 0.8	0.905 ± 0.007
Local-Only	83.6 ± 1.2	81.2 ± 1.3	80.5 ± 1.5	80.8 ± 1.2	0.861 ± 0.011

6. CONCLUSION

In this paper, it was described that FedMM-DL is a Federated Multimodal Deep Learning model of privacy-preserving predictive analytics in the American health care systems. Our design, which combines model and textual clinical data with structure EHR data, medical imaging and state-of-the-art performance can be attained with our framework (94.7% accuracy, 0.964 AUC-ROC), and can ensure a high level of privacy (e = 1.0). The suggested attention-based multimodal fusion and modality-conscious aggregation method are effective to manage the heterogeneity of distributed healthcare data.

Our comprehensive experiments indicate that FedMM-DL solutions are both centralized (achieve 1.2% AUC-ROC) and fully local (with reference to data) to offer a feasible solution to implementing advanced predictive analytics over U.S. healthcare networks and in compliance with HIPAA. The scaling report establishes that the framework can support increasing numbers of institutions that take part with valid communication overhead.

The research will advance in several potential directions in the future: (i) expanding the framework to apply other data forms, including genomic data and wearable sensor streams; (ii) exploring asynchronous federated optimization to support institutions with diverse levels of computation; (iii) creating tailored federated models that can serve individual hospital populations; and (iv) carrying out prospective clinical validation trials in collaboration with U.S. healthcare systems to find out whether it is really deployment feasible [29, 42].

REFERENCES

[1] A. Rajkomar, J. Dean, and I. Kohane, "Machine learning in medicine," New England Journal of Medicine, vol. 380, no. 14, pp. 1347–1358, 2019.

[2] E. J. Topol, "High-performance medicine: the convergence of human and artificial intelligence," Nature Medicine, vol. 25, no. 1, pp. 44–56, 2019.

[3] V. Gulshan et al., "Development and validation of a deep learning algorithm for detection of diabetic retinopathy," JAMA, vol. 316, no. 22, pp. 2402–2410, 2016.

[4] A. Rajkomar et al., "Scalable and accurate deep learning with electronic health records," npj Digital Medicine, vol. 1, no. 1, pp. 1–10, 2018.

[5] U.S. Department of Health and Human Services, "Summary of the HIPAA Privacy Rule," Office for Civil Rights, 2003.

[6] G. A. Kaissis et al., "Secure, privacy-preserving and federated machine learning in medical imaging," Nature Machine Intelligence, vol. 2, no. 6, pp. 305–311, 2020.

[7] N. Rieke et al., "The future of digital health with federated learning," npj Digital Medicine, vol. 3, no. 1, pp. 1–17, 2020.

[8] P. Kairouz et al., "Advances and open problems in federated learning," Foundations and Trends in Machine Learning, vol. 14, no. 1–2, pp. 1–210, 2021.

[9] B. McMahan et al., "Communication-efficient learning of deep networks from decentralized data," in Proc. AISTATS, pp. 1273–1282, 2017.

[10] Q. Yang et al., "Federated machine learning: Concept and applications," ACM Transactions on Intelligent Systems and Technology, vol. 10, no. 2, pp. 1–19, 2019.

[11] T. Li et al., "Federated learning: Challenges, methods, and future directions," IEEE Signal Processing Magazine, vol. 37, no. 3, pp. 50–60, 2020.

[12] J. Xu and F. Wang, "Federated learning for healthcare informatics," Journal of Healthcare Informatics Research, vol. 5, no. 1, pp. 1–19, 2021.

[13] L. Huang et al., "Patient clustering improves efficiency of federated machine learning," Journal of the American Medical Informatics Association, vol. 26, no. 3, pp. 263–268, 2019.

[14] T. S. Brisimi et al., "Federated learning of predictive models from federated electronic health records," International Journal of Medical Informatics, vol. 112, pp. 59–67, 2018.

[15] L. Huang et al., "Community-based federated learning for brain tumor segmentation," in Proc. MICCAI Workshop, Springer, pp. 92–104, 2019.

[16] X. Li et al., "FedBN: Federated learning on non-IID features via local batch normalization," in Proc. ICLR, 2021.

[17] Y. Xu et al., "A federated learning system for COVID-19 CT image analysis," Nature Communications, vol. 12, no. 1, pp. 1–11, 2021.

[18] M. J. Sheller et al., "Federated learning in medicine: facilitating multi-institutional collaborations," Scientific Reports, vol. 10, no. 1, pp. 1–12, 2020.

[19] D. C. Nguyen et al., "Federated learning for smart healthcare: A survey," ACM Computing Surveys, vol. 55, no. 3, pp. 1–37, 2022.

[20] Y. Xu et al., "Multimodal deep learning for survival prediction," in Proc. MICCAI, Springer, pp. 406–415, 2019.

[21] S. Khadanga et al., "Using clinical notes with time series data for ICU management," in Proc. EMNLP Workshop on Health Text Mining, pp. 82–88, 2019.

[22] G. Holste et al., "End-to-end learning of fused image and non-image features for cardiac diagnosis," in Proc. IEEE ISBI, pp. 1–5, 2021.

[23] Y. Zhang et al., "Attention-based multimodal fusion for comprehensive cancer prognostication," in Proc. CVPR Workshop, pp. 1–8, 2021.

[24] C. Dwork et al., "The algorithmic foundations of differential privacy," Foundations and Trends in Theoretical Computer Science, vol. 9, no. 3–4, pp. 211–407, 2014.

[25] M. Abadi et al., "Deep learning with differential privacy," in Proc. ACM CCS, pp. 308–318, 2016.

[26] R. C. Geyer et al., "Differentially private federated learning: A client level perspective," arXiv preprint arXiv:1712.07557, 2017.

[27] K. Wei et al., "Federated learning with differential privacy: Algorithms and performance analysis," IEEE Transactions on Information Forensics and Security, vol. 15, pp. 3454–3469, 2020.

[28] K. Bonawitz et al., "Practical secure aggregation for privacy-preserving machine learning," in Proc. ACM CCS, pp. 1175–1191, 2017.

[29] S. Warnat-Herresthal et al., "Swarm learning for decentralized and confidential clinical machine learning," Nature, vol. 594, no. 7862, pp. 265–270, 2021.

[30] T. Li et al., "Federated optimization in heterogeneous networks," in Proc. MLSys, pp. 429–450, 2020.

[31] S. Bai et al., "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," arXiv preprint arXiv:1803.01271, 2018.

[32] G. Huang et al., "Densely connected convolutional networks," in Proc. CVPR, pp. 4700–4708, 2017.

[33] J. Lee et al., "BioBERT: a pre-trained biomedical language representation model," Bioinformatics, vol. 36, no. 4, pp. 1234–1240, 2020.

[34] A. Vaswani et al., "Attention is all you need," in Proc. NeurIPS, pp. 5998–6008, 2017.

[35] A. E. W. Johnson et al., "MIMIC-III, a freely accessible critical care database," Scientific Data, vol. 3, no. 1, pp. 1–9, 2016.

[36] J. Irvin et al., "CheXpert: A large chest radiograph dataset," in Proc. AAAI, pp. 590–597, 2019.

[37] A. E. W. Johnson et al., "MIMIC-CXR, a de-identified publicly available database of chest radiographs," Scientific Data, vol. 6, no. 1, pp. 1–8, 2019.

[38] T.-M. H. Hsu et al., "Measuring the effects of non-identical data distribution for federated visual classification," arXiv preprint arXiv:1909.06335, 2019.

[39] H. Wang et al., "Federated learning with matched averaging," in Proc. ICLR, 2020.

[40] A. Paszke et al., "PyTorch: An imperative style, high-performance deep learning library," in Proc. NeurIPS, pp. 8026–8037, 2019.

[41] T. Ryffel et al., "A generic framework for privacy preserving deep learning," arXiv preprint arXiv:1811.04017, 2018.

[42] J. Konecny et al., "Federated learning: Strategies for improving communication efficiency," arXiv preprint arXiv:1610.05492, 2016.

[43] A. Vaid et al., "Federated learning of electronic health records to improve mortality prediction," Nature Medicine, vol. 27, no. 10, pp. 1735–1743, 2021.

[44] I. Dayan et al., "Federated learning for predicting clinical outcomes in patients with COVID-19," Nature Medicine, vol. 27, no. 10, pp. 1735–1743, 2021.