Design of Cross-Domain Medical Data Privacy-Preserving Classification Algorithm Based on Federated Learning

Yang bo Gong yichen

Mongolian National University ， Ulaanbaatar city Bayangol District 11th Khoroo National University of Mongolia，16060；

Abstract： Against the backdrop of rapid medical informatization, the sharing and utilization of cross-domain medical data have become key to improving the accuracy and efficiency of medical diagnosis, but the risk of data privacy leakage has also intensified. Traditional centralized machine learning algorithms require aggregating multi-domain medical data onto a unified platform, making it difficult to meet privacy protection requirements. Federated learning, as a distributed machine learning technique, enables collaborative model training without sharing raw data, providing a new solution for cross-domain medical data privacy protection. This paper designs a cross-domain medical data privacy-preserving classification algorithm based on federated learning. It first analyzes the characteristics of cross-domain medical data and privacy protection requirements, then designs an algorithmic framework combining federated learning architecture, including key steps such as local model training, encrypted transmission of model parameters, and global model aggregation and updating. Finally, it discusses the challenges and optimization directions of the algorithm in practical applications, aiming to provide technical support for the secure utilization of data and classification tasks in cross-domain medical scenarios, and to promote the compliant development of medical artificial intelligence.

Keywords: Privacy Protection; Classification Algorithm; Distributed Training

DOI:10.69979/3041-0843.25.03.058

1 Research Background and Significance

1.1 Research Background

With the deep penetration of artificial intelligence technology in the medical field, machine learning models based on multi-source medical data have shown significant advantages in tasks such as disease diagnosis and risk prediction. Cross-domain medical data includes patient diagnosis and treatment records, imaging data, test results, and other information from different hospitals, regions, or even countries. This data is rich in dimensions and large in sample size, effectively enhancing the generalization ability of models. However, medical data contains a large amount of sensitive personal information, such as patient identifiers, medical history, and genetic data. Constrained by laws and regulations such as the "Personal Information Protection Law" and "Medical Data Security Guidelines," data owners find it difficult to directly share raw data with third-party institutions or other medical domains, leading to prominent "data silo" problems. Traditional centralized classification algorithms rely on centralized data storage and processing, facing not only privacy leakage risks during data transmission but also difficulties in practical application due to issues like data ownership and data quality differences. Therefore, there is an urgent need to explore new technical paths that balance data privacy protection and cross-domain model training.

1.2 Research Significance

From a theoretical perspective, the cross-domain medical data privacy-preserving classification algorithm designed in this paper enriches the application research of federated learning in the medical field and provides new theoretical ideas for

solving the problem of balancing privacy and utility in cross-domain data sharing. From a practical perspective, this algorithm can achieve collaborative modeling of multi-domain medical data without leaking raw medical data, improve the accuracy of disease classification and diagnosis, facilitate the optimal allocation of medical resources, and promote the compliant development of smart healthcare.

1.3 Domestic and International Research Status

Internationally, Google first proposed the concept of federated learning in 2016 and applied it to model training on mobile devices. Subsequently, institutions such as Stanford University and MIT combined federated learning with medical data processing, designing federated classification models for Electronic Health Records (EHRs). However, these models mostly focus on local optimization of single-domain data and lack adaptation to cross-domain data heterogeneity. Domestically, universities like Tsinghua University and Shanghai Jiao Tong University have conducted research on the application of federated learning in medical privacy protection, proposing federated model optimization schemes based on homomorphic encryption. However, there are still shortcomings in balancing cross-domain model aggregation efficiency and classification accuracy. In summary, current research still has room for improvement in adapting to the heterogeneity of cross-domain medical data and collaboratively optimizing model training efficiency and privacy protection strength.

2 Related Technical Foundations

2.1 Federated Learning Technology

Federated learning is a distributed machine learning paradigm. Its core idea is "data does not move, models do." That is, multiple data owners (clients) train models locally and only upload intermediate results such as model parameters or gradients to a central server (coordinator). The coordinator aggregates these parameters to generate a global model, then distributes the global model parameters back to each client for iterative training until the model converges. Based on data distribution characteristics, federated learning can be divided into Horizontal Federated Learning (high overlap in sample features, low overlap in sample identifiers), Vertical Federated Learning (high overlap in sample identifiers, low overlap in sample features), and Federated Transfer Learning (low overlap in both sample identifiers and features). Among these, Horizontal Federated Learning and Federated Transfer Learning are more suitable for cross-domain medical data scenarios, as they can handle data from different medical domains with similar features but different patient populations, or data where both features and patient populations differ.

2.2 Privacy Protection Technology

2.2.1 Encryption Technology

Encryption technology is a core means of ensuring the security of parameter transmission and storage in federated learning. Commonly used techniques include Homomorphic Encryption (HE), Differential Privacy (DP), and Secure Multi-Party Computation (SMPC). Homomorphic encryption allows computation directly on encrypted data without decryption, enabling encrypted aggregation of model parameters, but it has high computational complexity, which can easily lead to reduced training efficiency. Differential privacy adds noise to data or model parameters, preventing attackers from inferring raw data from parameters. It offers strong privacy protection strength and has a controllable impact on model performance. Secure multi-party computation decomposes computational tasks among multiple participants, so that each participant only obtains their own computation results and cannot access the data of other participants. It is suitable for multi-domain collaborative computing scenarios but requires strict control over the number of participants and communication overhead.

2.2.2 Model Compression and Optimization Technology

To reduce the transmission volume of model parameters in cross-domain federated learning and improve training efficiency, model compression and optimization technologies are indispensable. Common methods include parameter pruning (removing redundant parameters from the model), quantization (converting high-precision parameters to low-precision parameters), and knowledge distillation (using a small model to learn the knowledge of a large model). These technologies can reduce data transmission overhead and lower the risk of privacy leakage (less parameter transmission

means lower probability of attack) while ensuring model classification accuracy.

2.3 Medical Data Classification Algorithms

Medical data classification tasks require selecting suitable algorithms based on the data type. Commonly used algorithms include Logistic Regression (LR), Support Vector Machine (SVM), Convolutional Neural Networks (CNN), and Transformer. Logistic Regression and SVM are suitable for classifying structured medical data (such as test indicators, medical history records), with simple model structures and strong interpretability. Convolutional Neural Networks perform well in classifying medical imaging data (such as CT, MRI images), capable of automatically extracting feature information from images. Transformer, leveraging its attention mechanism, has advantages in classification tasks for long-sequence medical data (such as ECG signals, multi-modal diagnostic records). In cross-domain federated learning scenarios, it is necessary to select appropriate basic classification algorithms based on the medical data type and cross-domain heterogeneity characteristics, and adapt them for distributed settings.

3 Design of Cross-Domain Medical Data Privacy-Preserving Classification Algorithm Based on Federated Learning

3.1 Algorithm Design Objectives

The core objectives of the algorithm are threefold: First, privacy protection, ensuring that cross-domain training does not leak raw medical data, and that parameter transmission and aggregation comply with privacy regulations. Second, cross-domain adaptation, handling heterogeneity in sample distribution, feature dimensions, and data quality across different medical domains to prevent model performance degradation. Third, efficiency and practicality, reducing training and parameter transmission overhead while ensuring privacy and accuracy to meet the real-time needs of medical applications.

3.2 Algorithm Framework Design

The algorithm is based on a fused architecture of Horizontal Federated Learning and Federated Transfer Learning, divided into two layers, "Client-Coordinator," containing five core modules:

Data Preprocessing: Clients clean local data (remove missing values, outliers), standardize features, and split training and validation sets. To address cross-domain heterogeneity, feature mapping is used to unify non-overlapping feature spaces, and sample weighting strategies are employed to increase the weight of special samples.

Local Model Training: Select basic models based on data type (Logistic Regression/SVM for structured data, lightweight CNN for image data, simplified Transformer for long-sequence data). Optimize using mini-batch stochastic gradient descent, dynamically adjust the learning rate, train with cross-entropy loss function, add regularization to prevent overfitting, and stop ineffective iterations based on validation set performance.

Parameter Encryption and Transmission: Only model parameters are transmitted. A "Differential Privacy + Homomorphic Encryption" approach is used: first, add Laplace noise (controlled by privacy budget for strength), then encrypt using a homomorphic encryption algorithm, and upload to the coordinator via a secure protocol.

Global Model Aggregation: The coordinator verifies parameter integrity, then performs dynamically weighted aggregation based on client data volume and model performance, generates encrypted global parameters, and distributes them.

Model Iterative Optimization: Clients decrypt the global parameters and restart local training, iterating repeatedly until the model converges. The final global model is used for local prediction by clients.

3.3 Algorithm Security and Complexity Analysis

Security: "Differential Privacy + Homomorphic Encryption" achieves end-to-end protection, preventing parameter inversion to raw data and single-point leakage; raw data is stored locally, complying with the principle of data ownership; TLS 1.3 and hash verification ensure communication security.

Complexity: Local training uses low-complexity models adapted to medical hardware; simplified homomorphic encryption reduces computation time; only parameters are transmitted and they can be compressed, resulting in low

communication overhead, meeting the needs of cross-domain medical applications.

4 Algorithm Application Challenges and Optimization Directions

4.1 Application Challenges

4.1.1 Cross-Domain Data Heterogeneity Exacerbates Model Bias

Differences in feature distribution, label definition, and data quality across cross-domain medical data may lead to model bias. For example, different hospitals may have varying diagnostic criteria for the same disease, leading to label inconsistency; the equipment precision in primary hospitals versus tertiary hospitals may differ, leading to variations in imaging data quality. Such heterogeneity can reduce the classification accuracy of the global model, or even render the model inapplicable in some medical domains.

4.1.2 The Trade-off Between Privacy Protection and Model Accuracy

There is a certain contradiction between privacy protection strength and model accuracy: higher noise intensity in differential privacy and higher levels of homomorphic encryption provide stronger privacy protection but lead to model parameter distortion and reduced classification accuracy; conversely, reducing privacy protection strength can improve accuracy but increases privacy leakage risk. How to dynamically adjust privacy strategies based on the privacy requirements of the medical scenario (e.g., genetic data requires high-strength protection, while ordinary test data can tolerate lower protection) is a key challenge in algorithm application.

4.1.3 High Difficulty in Multi-Party Collaborative Management

Cross-domain medical scenarios involve multiple participants such as hospitals and regulatory agencies, each with different interests and technical capabilities. Some hospitals may refuse to participate in collaborative training due to concerns about data security or technical costs. Regulatory agencies need to supervise the entire algorithm training process to ensure compliance, but the lack of unified regulatory standards and technical tools makes collaborative management difficult.

4.2 Optimization Directions

4.2.1 Introduce Adaptive Cross-Domain Adaptation Mechanisms

To address data heterogeneity, adaptive feature alignment and distribution calibration techniques can be introduced: use federated transfer learning to transfer knowledge from the source domain (medical domains with high data quality) to the target domain (medical domains with low data quality or small sample sizes) to improve the target domain's model performance; adopt adaptive weight adjustment strategies to dynamically update aggregation weights based on real-time distribution changes in each medical domain's data, reducing the impact of data heterogeneity on the global model.

4.2.2 Design Dynamic Privacy Control Strategies

Based on the sensitivity level classification of medical data (e.g., classifying data as high, medium, or low sensitivity), design dynamic privacy control strategies: for highly sensitive data (e.g., genetic data), use "high-noise differential privacy + fully homomorphic encryption"; for medium-sensitivity data (e.g., imaging data), use "medium-noise differential privacy + partially homomorphic encryption"; for low-sensitivity data (e.g., ordinary test data), use "low-noise differential privacy + symmetric encryption." This aims to maximize model accuracy while meeting different privacy requirements.

4.2.3 Build a Multi-Party Collaborative Management Platform

Collaborate with medical regulatory agencies, technology providers, and participants from various medical domains to build a cross-domain medical federated learning collaborative management platform. The platform needs functions such as participant identity authentication, training process monitoring, and privacy compliance auditing. Establish a benefit distribution mechanism to provide technical support or data sharing revenue compensation to medical domains participating in collaborative training, enhancing participation enthusiasm. Develop unified cross-domain medical federated learning standards to standardize operational processes for data preprocessing, model training, parameter transmission, and other steps, ensuring the compliance and scalability of algorithm applications.

5 Conclusion and Outlook

5.1 Conclusion

This paper designed a cross-domain medical data privacy-preserving classification algorithm based on federated learning. By integrating Horizontal Federated Learning and Federated Transfer Learning, a two-layer "client-coordinator" algorithm framework was constructed, achieving privacy protection and collaborative classification for cross-domain medical data. The algorithm addressed cross-domain data heterogeneity through feature alignment and sample weighting in the data preprocessing stage; ensured privacy and security during model parameter transmission and aggregation through a hybrid encryption strategy of "Differential Privacy + Homomorphic Encryption"; and improved the classification accuracy and training efficiency of the global model through dynamic weight aggregation and iterative optimization. Security and complexity analysis showed that the algorithm can achieve efficient cross-domain model training while meeting the privacy protection requirements of medical data, providing a feasible technical solution for the secure utilization of cross-domain medical data.

5.2 Outlook

Future research can proceed in three directions: First, explore collaborative modeling of multi-modal cross-domain medical data, integrating multi-modal data such as text (medical history records), images (imaging data), and numerical values (test indicators) into the federated learning framework to further enhance the model's classification capability. Second, combine blockchain technology to achieve decentralization and traceability of the model training process, enhancing the algorithm's trustworthiness and anti-attack capability. Third, conduct clinical empirical research on the algorithm, verifying its performance and practicality in real medical scenarios, promoting the transition of the algorithm from theoretical design to clinical application, and providing more powerful technical support for the development of smart healthcare.

References

[1] Sun Changyong. Research and Implementation of Cross-Domain Task Scheduling Algorithm Based on Federated Deep Reinforcement Learning [D]. Beijing University of Posts and Telecommunications, 2023.
[2] Zhang Shenyilang. Research on Cross-Domain sEMG Gesture Recognition Based on Transfer Learning [D]. Hangzhou Dianzi University, 2023.
[3] Zhou Chuanxin. Research on Key Technologies for Data Security Sharing in Federated Learning [D]. Strategic Support Force Information Engineering University, 2022.
[4] Pu Liyuan. Research on Federated Learning Load-Aware Intelligent Management of RAN Slicing [D]. Beijing University of Posts and Telecommunications, 2023.
[5] Software Engineering. Research on Cross-Domain Recommendation Method Based on Transfer Federated Learning [D]. 2024.