IOWA STATE UNIVERSITY

Ushashi Bhattacharjee
Bioinformatics & Computational Biology
Iowa State University

Tirtho Roy
Data Science
Iowa State University

MICCAI 2025 Daejeon REPUBLIC OF KOREA

Federated In-Context Prompt Selection for Multi-Modal 3D Dental Imaging: A Theoretical Framework with Privacy-Preserving Guarantees

Abstract

Vision-language models show remarkable capabilities in medical imaging analysis, yet their deployment in federated healthcare environments faces key challenges in privacy preservation, data heterogeneity, and adversarial robustness. We present FedDental3D-ICL, a theoretical framework for federated in-context prompt learning that enables privacy-preserving collaboration across healthcare institutions without sharing sensitive patient data or model parameters. Our framework introduces four core algorithmic contributions:Multi-Modal Prompt Space (MMPS) abstraction unifying visual and textual prompt representations across 2D and 3D medical imaging modalities; Cross-Modal Prompt Alignment (CMPA) ensuring semantic consistency through information-theoretic contrastive objectives; Hierarchical Multi-Modal Optimization (HMMO) providing theoretical convergence guarantees for non-convex federated objectives; and Byzantine-Resilient Cross-Modal Aggregation (BRCMA) with differential privacy bounds. Our theoretical analysis suggests potential convergence rates of $O(1/\sqrt{T})$, theoretical communication complexity bounds of $O(K \log |P|)$ compared to traditional $O(K \cdot d)$, and (ε, δ)-differential privacy guarantees with optimal composition bounds. While this work establishes comprehensive mathematical foundations, empirical validation and practical implementation remain important directions for future research.

Framework

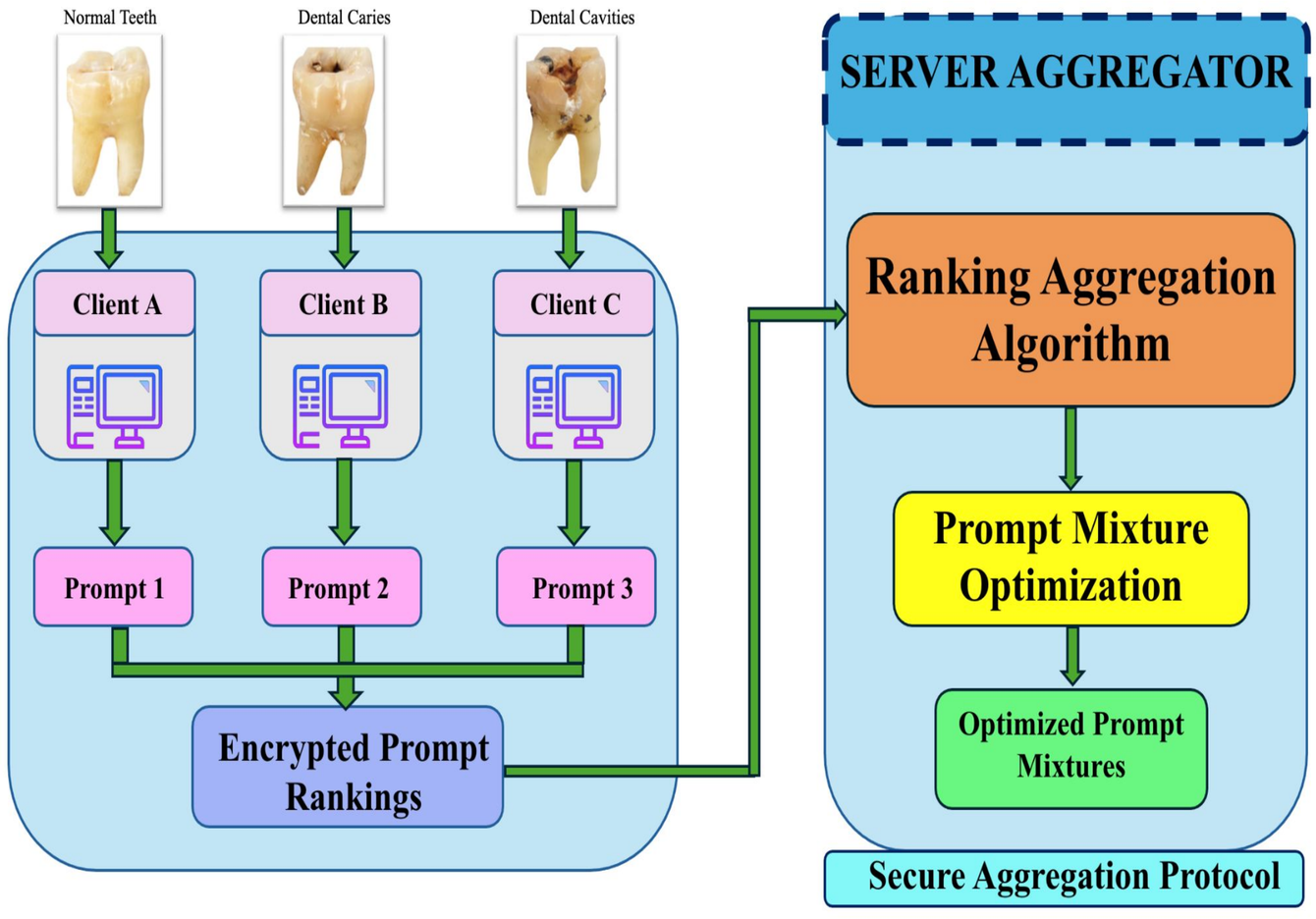

Fig. 1. FedDental3D-ICL System Architecture showing multi-modal data flow across federated dental institutions with privacy-preserving prompt exchange.

Introduction

Federated, privacy-preserving multimodal AI is urgently needed in healthcare, particularly in areas like dental imaging where large, diverse, and sensitive patient datasets are indispensable for advancing diagnosis and treatment. Yet the risks of traditional data sharing are profound: between 2009 and 2025 there have been 6,759 U.S. healthcare data breaches of 500 or more records, exposing or impermissibly disclosing the protected health data of 846,962,011 individuals.In 2023 alone, 725 such large breaches were reported, exposing over 133 million records. These breaches set new records in both number of incidents and number of records breached.At the same time, medical imaging analysis stands at a critical juncture where the transformative potential of vision–language models (VLMs) collides with the immutable constraints of healthcare data governance. Recent advances in VLMs have shown unprecedented multimodal reasoning capabilities, yet their deployment in real-world healthcare environments exposes a fundamental paradox: the models that show the greatest promise require precisely the type of large-scale, cross-institutional data sharing that regulatory frameworks explicitly prohibit. This tension is more than a technical bottleneck—it is a systemic barrier that prevents the medical community from fully leveraging state-of-the-art AI while upholding the privacy guarantees essential to patient trust and compliance.

Research Questions

RQ1. *How can federated multi-modal prompt learning frameworks be designed to reconcile the privacy–utility paradox, enabling vision–language models to achieve clinically useful diagnostic performance in dental imaging without direct data sharing?*

RQ2. *To what extent can hierarchical optimization strategies in federated settings address the statistical heterogeneity of dental imaging data across institutions, ensuring stable convergence and generalizable model performance?*

RQ3. *Can privacy-preserving, Byzantine-resilient aggregation mechanisms safeguard against both adversarial clients and regulatory non-compliance while maintaining efficiency and robustness in large-scale multimodal healthcare training?*

Proposed Algorithm

Algorithm 1 Multi-Modal Prompt Space Construction (MMPS)

Require: Raw prompts P_v, P_t, P_{3D}; contrastive parameters (τ, batch_size)
Ensure: Unified embeddings $\psi(p_v, p_t, p_{3D})$, learned fusion function F

Initialize embedding functions ϕ_v, ϕ_t, ϕ_{3D} with random weights
Initialize fusion network F with Xavier initialization
▷ Phase 1: Individual modality embedding learning
for each modality $m \in \{v, t, 3D\}$ **do**
  for epoch = 1 to E_1 **do**
    Sample batch of prompts $\{p_m^{(i)}\}$
    Compute embeddings $z_m^{(i)} = \phi_m(p_m^{(i)})$
    Update ϕ_m via contrastive loss minimization
  end for
end for
▷ Phase 2: Cross-modal fusion learning
for epoch = 1 to E_2 **do**
  Sample multi-modal triplets $(p_v^{(i)}, p_t^{(i)}, p_{3D}^{(i)})$
  Compute modality embeddings $z_v^{(i)} = \phi_v(p_v^{(i)})$, $z_t^{(i)} = \phi_t(p_t^{(i)})$, $z_{3D}^{(i)} = \phi_{3D}(p_{3D}^{(i)})$
  Compute fused embedding $\psi^{(i)} = F(z_v^{(i)}, z_t^{(i)}, z_{3D}^{(i)})$
  Update F via contrastive loss on $\psi^{(i)}$
end for
return $\psi(p_v, p_t, p_{3D})$, F

Algorithm 2 Hierarchical Multi-Modal Optimization (HMMO)

Require: Global prompt candidates P_{global}, participation probability p_m
Ensure: Optimal prompt distribution θ, client adaptations $\{\theta_k\}$

Initialize global prompt parameters $\theta^{(0)}$
for round $t = 1$ to T **do**
  Select subset of clients $S_t \subseteq [K]$ with probability p_m
  for each client $k \in S_t$ **do**
    Evaluate prompt candidates on local data: $q_k(p) = \text{quality}(M(x_k, p), y_k)$ for $p \in P_{\text{global}}$
    Generate local prompt ranking: $r_k = \text{argsort}(q_k, \text{descending} = \text{True})$
    Compute alignment statistics: $a_k = \text{cross\_modal\_alignment}(r_k)$
    Send encrypted (r_k, a_k) to server
  end for
  Aggregate rankings via Byzantine-resilient mechanism: $\theta^{(t)} = \text{BRCMA}(\{r_k, a_k\}_{k \in S_t})$
end for
return $\theta^{(T)}, \{\theta_k^{(T)}\}$

Algorithm 3 Byzantine-Resilient Cross-Modal Aggregation with Privacy-Preserving Selection (BRCMA-PMMS)

Require: Client rankings $\{r_k\}$, privacy parameters (ε, δ), prompt set P
Ensure: Aggregated prompt parameters θ, privacy-preserving selection

Initialize global prompt parameters $\theta^{(0)}$
for round $t = 1$ to T **do**
  Collect client rankings $\{r_k^{(t)}\}$ and quality scores $\{q_k^{(t)}\}$
  Apply exponential mechanism to select top prompts:
  $p_{\text{select}}(p) \propto \exp\left(\frac{\varepsilon \cdot q(D,p)}{2\Delta q}\right)$
  Filter out Byzantine clients using cross-modal consistency check:
  $S_t^{\text{valid}} = \{k : \text{consistency}(r_k, \{r_j\}_{j \neq k}) > \tau_{\text{byz}}\}$
  Aggregate valid client rankings using median-based robust estimator
  Add calibrated Gaussian noise for (ε, δ)-differential privacy
  Update global prompt distribution: $\theta^{(t)} = \text{weighted\_aggregate}(S_t^{\text{valid}})$
end for
return $\theta^{(T)}$

We gratefully acknowledge the support and guidance received during our COM S 4590X: Security and Privacy in Cloud Computing final project at Iowa State University, which made this work possible. We would also like to express our appreciation to the Translational AI Center as a token of gratitude for their encouragement and support throughout this research.