Federated Learning for Personalized Fitness Tracking on Mobile Devices

---

1. Research Objectives and Scope

The objective of this project is to develop and evaluate a **Federated Learning (FL)** framework for **personalized fitness tracking** on mobile devices. The focus is on building an end-to-end solution that addresses model efficiency, privacy preservation, and real-world deployment feasibility. The system will:

- Utilize **mobile sensor data** (accelerometer and gyroscope) to recognize fitness activities.
- Implement **FL with privacy-preserving techniques**, ensuring sensitive user data remains on the device.
- Apply **model optimization strategies** to enhance accuracy, reduce communication overhead, and increase scalability.
- Provide **comprehensive evaluations** with detailed metrics, ablation studies, and comparative baselines.
- Contribute novel insights and optimizations that strengthen its chances for acceptance in top-tier conferences.

---

2. Technical Architecture and Design

Dataset Selection and Preprocessing

- **Dataset:** Use the **UCI HAR Dataset** with smartphone accelerometer and gyroscope data, collected from 30 individuals performing six activities (walking, sitting, standing, etc.).
- **Preprocessing Steps:**
  - Normalize the data (0 to 1 range) to prevent large variances in sensor values.
  - Split the dataset by subjects, simulating distributed clients (one subject = one device).
  - Augment data with slight noise and rotations to simulate real-world device variations.
  - Convert time series data into **feature vectors** representing activity patterns.

Model Architecture

- Use a **Feed-Forward Neural Network (FNN)** with the following structure:
  - Input Layer: 561 features (preprocessed data vector).
  - Hidden Layer 1: 128 neurons, ReLU activation.
  - Hidden Layer 2: 64 neurons, ReLU activation.

    - Output Layer: 6 neurons (one per activity class) with Softmax activation.
- **Compilation Parameters:**
    - Optimizer: Adam
    - Loss: Categorical Cross-Entropy
    - Metrics: Accuracy

Federated Learning Setup

- **Aggregation Algorithm:** Use **Federated Averaging (FedAvg)** for model parameter aggregation.
- **Client-Server Interaction:**
    - Each client (mobile device) trains the model locally on its private dataset.
    - The server receives only the **model updates (gradients)**, not the raw data.
    - The server aggregates the updates and refines the global model.
- **Communication Rounds:**
    - Define **10–30 communication rounds** with **random client selection** in each round.
    - Track performance metrics at each round to monitor convergence.

---

3. Privacy-Preserving Techniques

To strengthen the project's contribution, implement the following privacy and security techniques:

- **Differential Privacy (DP):**
    - Add controlled noise to local model updates before sending them to the server.
    - Use a privacy budget ε to balance privacy and model accuracy.
- **Secure Aggregation:**
    - Encrypt model updates using homomorphic encryption or secure multi-party computation (SMPC).
    - Ensure that individual model updates remain hidden even from the server.
- **Personalized FL:**
    - Implement local fine-tuning after global aggregation.
    - Use client-specific model adjustments to enhance individual accuracy.

---

4. Optimization Strategies

To enhance model efficiency and reduce communication overhead, apply:

- **Model Compression:**
    - Use **quantization** (reducing model precision to 16-bit or 8-bit) to reduce model size.
    - Apply **pruning** by removing low-importance weights, reducing the communication payload.
- **Adaptive Client Selection:**
    - Instead of using all clients in each round, select a subset based on **accuracy performance** or **data heterogeneity**.
    - This reduces overhead and speeds up convergence.
- **Regularization Techniques:**
    - Add **dropout layers** and **L2 regularization** to prevent overfitting.
    - Improve model generalization across diverse clients.

---

5. Model Evaluation and Benchmarking

- **Evaluation Metrics:**
    - Accuracy, F1-score, and Recall for activity classification.
    - Privacy overhead and communication costs.
    - Model convergence rates over multiple rounds.
- **Benchmarking:**
    - Compare FL performance against **centralized training** and **local-only training**.
    - Include an ablation study by removing privacy techniques and comparing accuracy.
- **Statistical Validation:**
    - Use statistical significance tests (e.g., **Wilcoxon signed-rank test**) to validate improvements.
    - Report **confidence intervals** for accuracy and privacy scores.

---

6. Results and Insights

- **Model Performance:**
    - Showcase detailed performance charts, including accuracy over rounds, loss reduction, and privacy-accuracy trade-offs.
- **Privacy Analysis:**

   - Demonstrate how **DP and secure aggregation** reduce privacy risks.
   - Quantify privacy leakage using formal metrics.
- **Efficiency Gains:**
   - Show the reduction in communication overhead using model compression and adaptive client selection.
- **Ablation Study:**
   - Include experiments with and without privacy techniques to highlight their impact.

7. Conference-Ready Paper Structure

Title:
Federated Learning on Mobile Devices for Personalized Fitness Tracking: Privacy-Preserving Model Optimization and Performance Analysis

Abstract:
A concise summary covering the motivation, methodology, and key findings. Mention the novelty in applying FL with privacy preservation to mobile fitness tracking.

1. Introduction:

- Motivation behind FL for privacy-preserving fitness tracking.
- Problem statement and objectives.

2. Related Work:

- Overview of FL and privacy-preserving AI.
- Review existing FL applications in fitness and mobile health.

3. Methodology:

- Dataset description and preprocessing.
- FL architecture, model details, and aggregation algorithm.
- Privacy-preserving techniques.

4. Experiments and Results:

- Detailed evaluation results with accuracy, privacy scores, and communication overhead.
- Ablation studies and statistical validation.

5. Discussion:

- Interpretation of the results.
- Privacy vs. accuracy trade-offs.
- Limitations and future work.

6. Conclusion:

- Summarize findings and highlight the contributions.
- Suggest potential directions for further research.

7. References:

- Cite relevant FL, privacy, and optimization papers.

---

8. Submission and Finalization Plan

- **Preparation Timeline:**
  - **Week 1-2:** Refine the FL model and optimize communication strategies.
  - **Week 3-4:** Integrate privacy-preserving techniques (DP and secure aggregation).
  - **Week 5-6:** Conduct extensive experiments and benchmarking.
  - **Week 7-8:** Write and review the paper.
  - **Week 9:** Submit to a top-tier conference.
- **Review and Iteration:**
  - Incorporate reviewer feedback and iterate before final submission.

Final Deliverables

- **Source Code:** FL implementation with complete preprocessing, training, and evaluation scripts.
- **Research Paper:** Well-structured manuscript ready for submission.
- **Supplementary Materials:** GitHub repository with detailed documentation and instructions.

✅ This project plan covers everything from technical implementation, privacy features, and optimization strategies to result analysis and conference submission, ensuring the project meets the standards for top-tier AI conferences.