ИССЛЕДОВАНИЕ МЕТОДОВ АГРЕГАЦИИ МОДЕЛЕЙ МАШИННОГО ОБУЧЕНИЯ С СОХРАНЕНИЕМ ПРИВАТНОСТИ МЕДИЦИНСКИХ ИЗОБРАЖЕНИЙ

Подготовил:
студент 4 курса
Зеленковский Виктор Петрович

Научный руководитель:
Ковалёв В.А.

Постановка задачи

Цель: исследование и разработка методов агрегации моделей машинного обучения с сохранением приватности медицинских изображений, в частности методов федеративного обучения

Задачи:

1. Исследование существующих методов агрегации моделей машинного обучения
2. Оценка возможности использовать существующие методы для медицинских данных
3. Разработка методов, позволяющих сохранить приватность данных
4. Оценка разработанных методов

Преимущества и недостатки агрегации

Преимущества:

- повышение точности модели;
- снижение риска переобучения;
- улучшение устойчивости модели.

Недостатки:

- увеличение времени обучения;
- сложность настройки параметров;
- риск утечки данных.

Федеративное обучение: определение

Клиенты: $\{C_1, C_2, ..., C_N\}$

Сервер: S

Этапы обучения:

1. Инициализация

   $$\{w_k^t, k \in [1, 2, ..., N]\}.$$

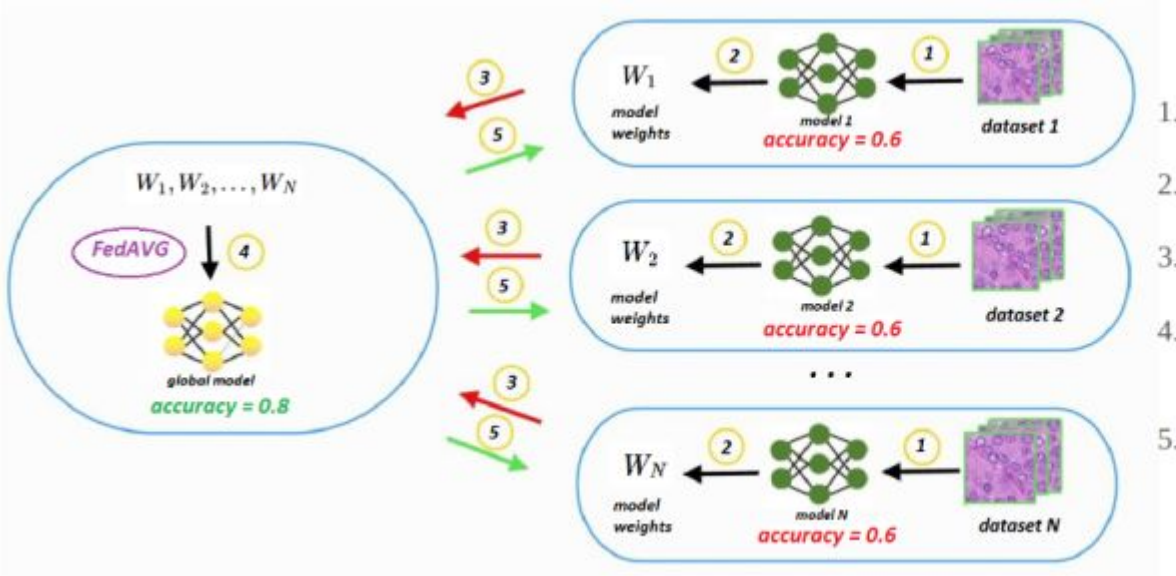
2. Локальное обучение

   $$w_k{}^t \leftarrow w_k{}^{t-1}(\eta, D_k)$$

3. Агрегация моделей

   $$w_{global}^t \leftarrow Aggregation(w_k{}^t; k \in [1, 2, ..., N])$$

Федеративное обучение: этапы обучения



1. Обучение локальных моделей на клиентах
2. Определение передаваемых весов модели
3. Передача весов модели
4. Агрегация весов (построение глобальной модели)
5. Отправка агрегированных весов обратно (обновление локальных моделей)

Федеративное обучение: основные методы

1. **FedAvg**

$$w_{global}^{t} \leftarrow \sum_{C_k \in S_t} \frac{n_k}{n} {w_k}^t$$

2. **FedProx**

$$\frac{\mu}{2} ||{w_k}^t - w_{global}^{t}||^2$$

3. **FedNova**

нормализация и масштабирование локальных обновлений от каждого клиента

Федеративное обучение: преимущества и недостатки

Преимущества:

- масштабируемость;
- упрощение модели;
- быстрая сходимость;

Недостатки:

- потеря приватности данных;
- трудность агрегации моделей различной архитектуры;

Агрегация с помощью части весов: определение

Алгоритм:

1. Отправка части весов серверу
2. Преобразование весов

$$F : R_k \to R^{h \times h} \qquad w^{t'}_k \leftarrow F(w^t_k)$$

3. Агрегация весов сервером

$$G : R^{h \times h} \to R^{h \times h} \qquad w^t_{global} \leftarrow G(w^{t'}_k)$$

4. Обратное преобразование и отправка клиентам

$$F^{-1} : R^{h \times h} \to R_k$$

Агрегация с помощью части весов: преимущества и недостатки

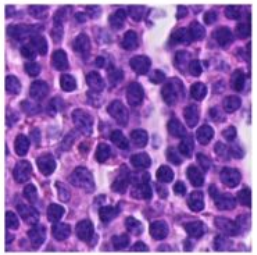
Преимущества:

- сохранение приватности изображений.

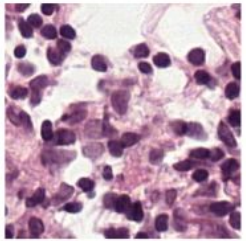
Недостатки:

- необходимость дополнительно обучать модель сервера;
- отсутствие улучшения качества моделей (мало весов);
- потенциальное ухудшение качеств (слишком разные архитектуры).

Эксперименты: данные



Класс 1 (normal)



Класс 2 (tumor)

Размер набора данных для обучения для каждого клиента и сервера: 1176.

Размер тестового набора данных: 2520.

Количество клиентов: 4.

Веса моделей для обмена: матрица весов линейного слоя 1024x1024

Эксперименты: модели

1. **Одинаковой архитектуры**

```
SimpleModel(
  (conv1): Conv2d(3, 16, kernel_size=(3, 3), stride=(1, 1))
  (pool): MaxPool2d(kernel_size=3, stride=3, padding=0, dilation=1, ceil_mode=False)
  (conv2): Conv2d(16, 16, kernel_size=(3, 3), stride=(1, 1))
  (conv3): Conv2d(16, 32, kernel_size=(3, 3), stride=(1, 1))
  (flatten): Flatten(start_dim=1, end_dim=-1)
  (fc): Linear(in_features=16928, out_features=1024, bias=True)
  (shared): Linear(in_features=1024, out_features=1024, bias=True)
  (classifier): Linear(in_features=1024, out_features=2, bias=True)
)
```

2. **Различной архитектуры**

   Клиент 1: SimpleModel

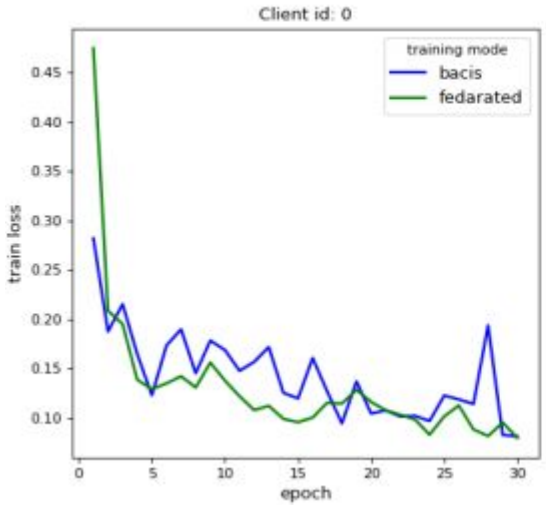
   Клиент 2: mobilenetv3 large

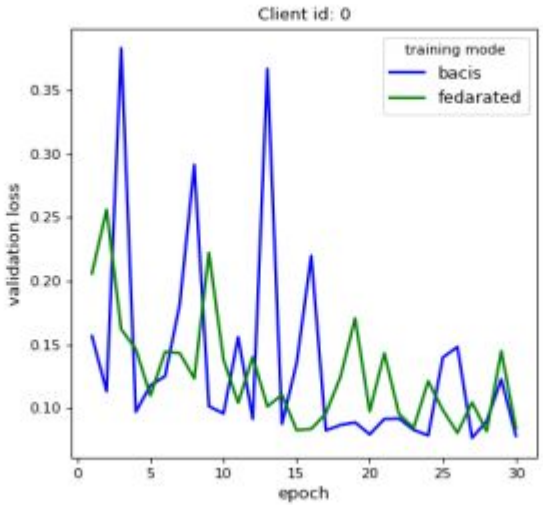
   Клиент 3: mobilenetv3 small

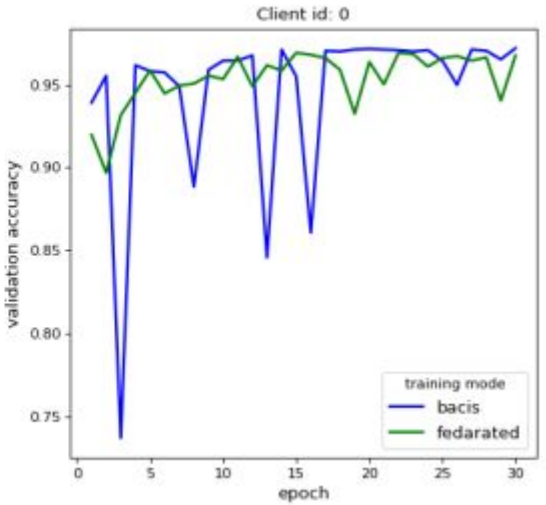
   Клиент 4: densenet121

Эксперименты: результаты

Одинаковая архитектура

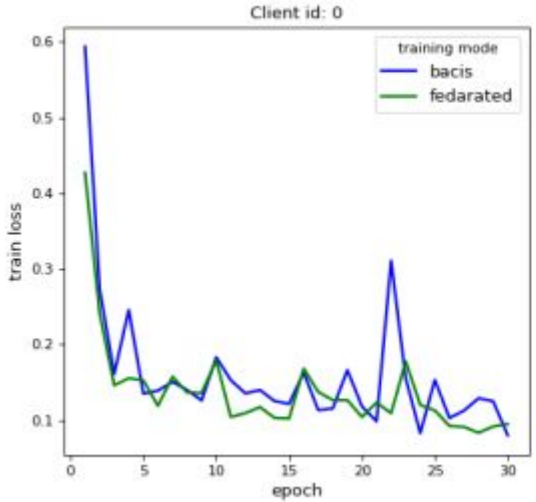


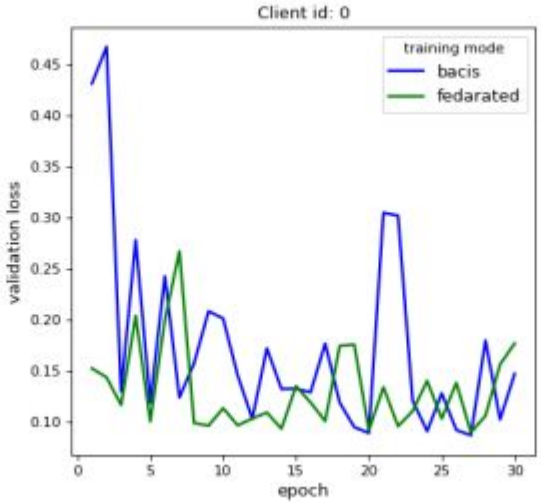


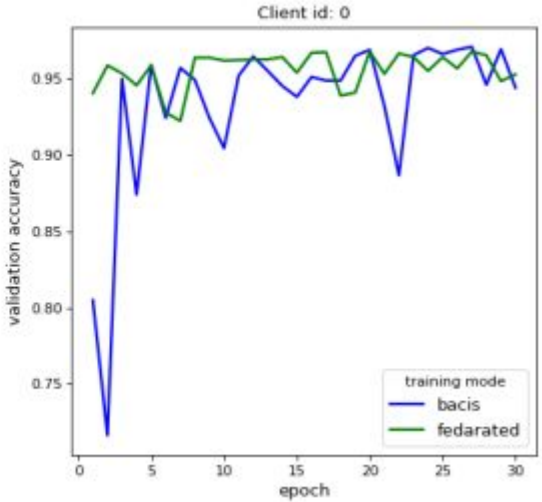


Эксперименты: результаты

Различная архитектура : SimpleModel

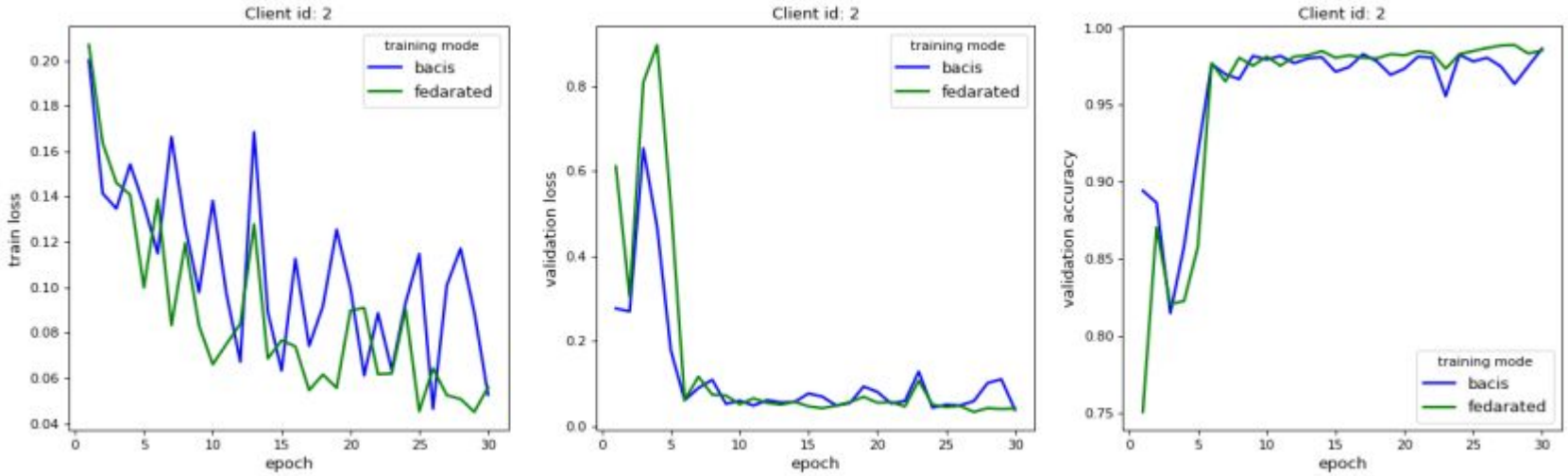






Эксперименты: результаты

Различная архитектура: mobilenetv3 large



Выводы

1. Невозможность использования базовых методов в условиях приватности медицинских изображений

2. Разработанный метод позволяет улучшить стабильность обучения моделей машинного обучения

Спасибо за внимание!!!