データの異種性が連合学習モデルの性能に与える影響に関する研究

池奥　裕太† 　高木　駿†† 　曹　洋†† 　吉川　正俊††

† 京都大学工学部情報学科　〒 606–8501 京都府京都市左京区吉田本町
†† 京都大学大学院情報学研究科　〒 606–8501 京都府京都市左京区吉田本町
E-mail: †{ikeoku.yuta.37c,takagi.shun.45a}@st.kyoto-u.ac.jp, ††{yang,yoshikawa}@i.kyoto-u.ac.jp

あらまし　近年，深層学習は様々な領域で用いられている．高い精度の深層学習モデルを得るには多くのデータが必要となるが，一つの機関が単独で得られるデータセットのサイズには限りがある．そこで，複数の機関がデータを持ち寄り深層学習を行う連合学習（federated learning）が用いられる．しかし連合学習では，クライアント間のデータの分布が不均一である場合に，機械学習モデルの精度が低下してしまう問題がある．この論文では，医療画像を用いた病気の診断モデルを学習する事例を考え，データの分布が不均一な場合に連合学習モデルにどのような影響があるのかを実験し，考察する．具体的には，大きなデータセットを持つ専門病院と小さなデータセットを持つ小規模な病院の 2 種類のクライアントを考え，病気データが含まれる割合が異なる場合や，差分プライバシを導入した場合などについて実験を行い，機械学習モデルの精度を調べる．

キーワード　連合学習, 医療画像, Non-IID data

1 はじめに

近年，深層学習は様々な領域で用いられている．高い精度の深層学習モデルを得るには多くのデータが必要となる．一つの機関が単独で得られるデータセットのサイズには限りがあり，複数の機関がデータを持ち寄り深層学習を行う場合がある．しかし，複数の機関でデータを共有することは，プライバシ保護の観点で困難な場合が多い．

連合学習 (federated learning) [1] では，各機関が持つデータを中央サーバに集めることなく機械学習モデル（以降，モデルという）を学習することができる．具体的には，中央サーバが持つグローバルモデルを学習に参加する各機関に送信し，各機関はローカルデータでモデルを更新する．得られたパラメータは中央サーバに送信され，集約され，新しいグローバルモデルが計算される．これを繰り返すことで，データを中央サーバに送信することなく学習を行うことができる．

この論文では，各病院が持つ患者の医療画像により病気の診断を行うモデルを，連合学習の手法を用いて学習する事例を考える．

連合学習の実社会での適用を進める上で考えられる問題の一つに不均一なデータ分布がある．一般的に，多くの場合では各機関が均一なデータを持っているとは考えにくい．例えば病院の規模によって患者数は大きく異なるため，データセットサイズが異なる．また，病院の位置する地域や専門性によって，データセットのラベルの分布が不均一となる．例えば大学病院や専門病院には重い病気を患う患者のデータが多く，一般の病院ではそれほど多くはない場合がある．

この論文では，実社会での医療画像を用いた機械学習に着目し，データが不均一な場合に連合学習モデルにどのような影響を与えるかを実験により観察し，その考察を行う．大きなデータセットを持つ専門病院と小さなデータセットを持つ小規模な病院の 2 種類を考え．各病院が持つデータセットや連合学習の有無，差分プライバシの有無などが，モデルの性能にどのような影響を与えるのかを調べる．

2 関連研究

2.1 連合学習

連合学習 [1] とは，データセットを所有する組織やユーザーが持つデバイスなどの各クライアントに存在するデータを中央サーバに送ることなく機械学習を行う方法である．中央サーバが受け取るのは，グローバルモデルを更新するために必要な最小限の更新値であるため，生の学習データを受け取る場合と比較して，プライバシ保護や通信コストの面で利点がある．

連合学習は，各クライアントが保有するデータの種類によって，水平連合学習 (Horizontal Fedeareted Learning) と垂直連合学習 (Vertical Federated Learning) の 2 種類に大別される．各機関が持つデータの特徴量が等しく，所持しているサンプルが異なるものが水平連合学習，同じサンプルを持つものの，その特徴量を複数の機関が分割して所持しているものが垂直連合学習である．本研究は各病院が異なる患者のデータを保持しているため水平連合学習にあたる．水平連合学習のアルゴリズム，McMahan らによって提案された Federated Averaging (以降，FedAvg とよぶ) アルゴリズム [2] がある．FedAvg アルゴリズムについては，3 章で詳しく述べる．

医療画像で連合学習を行う研究としては，Li らが行なった 4 つの大学病院での fMRI 画像を用いた連合学習に関する研究がある [3]．Li らは各病院での検査機器の種類や検査方法などの違いにより，同様の検査内容であっても病院間でのデータ分布に偏りが生じる問題に，ドメイン適用によって対応している．一方，クライアントによってデータセットサイズが大きく異な

るケースは考えていない．

2.2 非独立同一分布データでの連合学習

Zhao らは，FedAvg を用いて Non-IID データについて連合学習を行なった場合に，モデルの精度が低下することを示している [4]．これは，各クライアントが持つデータの分布が異なるため，ローカルモデルの平均がグローバルモデルと大きく異なることが原因である．文献 [5] は，このような Non-IID データへのアプローチとして主に 3 つの方法を挙げている．

- アルゴリズムベースのアプローチ
- システムベースのアプローチ
- データベースのアプローチ

アルゴリズムベースのアプローチとしては，Per-FedAvg [6] がある．Per-FedAvg は，ニューラルネットワークで抽象的な特徴を表現する浅い層を連合学習のパラメータとして用いる手法である．一方，具体的な分類を行う深い層のパラメータはクライアント内でのみ利用される．システムベースのアプローチは，ローカルデータの分布が類似しているクライアントをクラスタリングし，複数のグローバルモデルを学習する方法である．データベースのアプローチとしては，クライアント間で一部のデータを共有することで，データの分布の偏りを修正する方法がある．具体的には，一様な分布からなる小さなデータセットを各クライアントに配布し，各クライアントでの学習時に，このデータセットのデータとローカルデータの両方を用いる方法が Zhao らによって提案されている [4]．

本研究ではこれらの手法は用いないが，今後検証していきたい．

3 FedAvg の概要

Federated Averaging（以下 FedAvg という）は 2016 年に McMahan らによって提案された手法である [2]．

FedAvg のアルゴリズムについて説明する．FedAvg では以下の流れで学習が行われる．

（1） 中央サーバでグローバルモデルのパラメータを初期化する．

（2） 学習に参加する各クライアントにグローバルモデルを送信する．

（3） 各クライアントは，ローカルデータセットを用いてグローバルモデルの更新を計算し，中央サーバに送る．

（4） 中央サーバでは，各クライアントから送られてきたモデルの更新を集約し，新たなグローバルモデルを計算する．

（5） 上記の（2）から（4）までの流れを 1 ラウンドとし，ラウンドを繰り返すことで学習を進める．

この際，グローバルモデルの 1 度の更新に対し，ローカルデータでのパラメータの更新を複数エポック行うことも可能である．

クライアント数を K，ミニバッチ数を B，ローカルエポック数を E，各ラウンドにおける学習に参加するクライアントの割合を C，学習率を η，クライアント k のデータ数を n_k，全クライアントのデータ数の合計を n とする．t ラウンド目のグローバルモデルを w_t とすると，グローバルモデルの更新は式 (1) によって行われる．ここで S_t は，各ラウンド毎にランダムに選択した $C \times K$ 個のクライアントのインデックスのリストであり， w_{t+1}^k はクライアント k の t ラウンド目の更新終了時のパラメータである．

$$w_{t+1} \leftarrow \sum_{k \in S_t} \frac{n_k}{n} w_{t+1}^k \tag{1}$$

w_{t+1}^k は，直前のラウンドのグローバルモデル w_t を入力として，式 (2) によって更新される．

$$w_{t+1}^k \leftarrow ClientUpdate(w_t, k) \tag{2}$$

サブルーチン $ClientUpdate$ は，w_t を初期値として，ローカルデータを用いた E エポック分の学習を行い，得られたモデルを返却する．

4 差分プライバシの概要

差分プライバシ（differencial privacy）は Dwork によって提案され，以下のように定義される [7]．

定義 1. (ϵ, δ)-差分プライバシ　最大で 1 つの要素が異なる任意のデータセット D_1 と D_2 および全ての $S \subseteq Range(\mathcal{K})$ について，

$$Pr[\mathcal{K}(D_1) \in S] \leqq exp(\epsilon) \times Pr[\mathcal{K}(D_2) \in S] + \delta \tag{3}$$

を満たすとき，メカニズム $\mathcal{K}$ は (ϵ, δ)-差分プライバシを満たす

差分プライバシは，ある特定の要素がデータセットに追加されたり，データセットから削除されたりした場合に，データセットに対して $\mathcal{K}$ を適用しても結果に大きな変化が生じないことを意味する．パラメータ ϵ の大きさを変化させることで保護するプライバシの度合いを決めることができ，ϵ が小さい方がプライバシ保護の度合いが大きい．δ は，直感的には差分プライバシを満たさない確率を示しており，慣習的にデータ数の逆数が用いられる．

差分プライバシを連合学習に適用するアルゴリズムとして Differentially Private SGD Algorithm（以下 DP-SGD という）がある [8]．DP-SGD では，計算された勾配に対してガウス分布を加えることで，差分プライバシを満たしながら連合学習を行うことができる．

DP-SGD ではまず，全てのサンプルから L 個のサンプルをランダムに取り出し，L_t とする．これらのサンプルについて式 (4) のように勾配を計算する．なお，$x_1, ..., x_N$ はサンプル，$\mathcal{L}$ は損失関数である．

$$For\ each\ i \in L_t, compute\ g_t(x_i) \leftarrow \nabla_{\theta_t}(x_i)\mathcal{L}(\theta_t, x_i) \tag{4}$$

次に計算した勾配について，norm clipping を行う．norm

clipping は式 (5) で表される．ここで C は勾配ノルムの最大値で，学習の際に指定するパラメータである．norm clipping によって，勾配のノルムの最大値が C であることが保証される．

$$\bar{g}_t(x_i) \leftarrow g_t(x_i)/\max(1, \frac{||g_t(x_i)||_2}{C}) \tag{5}$$

次に，得られた勾配に対してノイズを加える．$\mathcal{N}$ はガウス分布，σ はノイズの大きさ，L はグループの大きさである．

$$g_t(x_i) \leftarrow \frac{1}{L}(\sum_i \bar{g}_t(x_i) + \mathcal{N}(0, \sigma^2 C^2 \mathbf{I})) \tag{6}$$

以上の手順でノイズを加えることで，差分プライバシを保証する．

DP-SGD に対して，クライアントレベルでのプライバシを保護する連合学習の手法も提案されている [9]．DP-SGD では，あるデータ点が学習データに含まれることがわからないようにプライバシを保護する．一方 [9] で提案されている手法は，学習時に特定のクライアントが学習に参加したことがわからないようにする．

本研究の実験では，患者のデータを保護する必要があるため，DP-SGD を利用する．

5 実　　験

5.1 データセット

実験では, FashionMNIST データセット [10] を用いる．FashionMNIST データセットはファッション関連サービスで利用されている 10 種類のファッションアイテムの画像から成る．60,000 件の学習データと 10,000 件のテストデータがあり，各データは 28×28 のグレースケール画像に処理されている．

この実験では，FashionMNIST データセットを医療画像とみなし，一部のラベルを異常値，それ以外のラベルを正常値とラベリングし直し，二値分類のデータセットとして利用した．

5.2 モ デ ル

連合学習に参加するクライアントを 2 種類設定した．1 つは専門病院や大学病院などの規模の大きい病院を表す大きなデータセットを持つ大規模病院のクライアント，もう 1 つは小さなデータセットを持つ小規模病院のクライアントである．一般的に，大規模病院より小規模病院の方が多く存在するため，クライアント数も大規模病院より小規模病院の方が多くなるようにしている．

ニューラルネットワークモデルは 2 層の畳み込み層と全結合層からなり，活性化関数には ReLU 関数と Softmax 関数を用いた．

5.3 評 価 方 法

学習したモデルの評価については，小規模病院におけるテストデータに対する再現率と AUROC（Area Under the Receiver Operating Characteristics）を用いた．

まず小規模病院のテストデータに対する性能を利用する理由は，この研究で考えている医療画像診断システムが，小規模病院でより有効であると考えられるからである．特に専門病院などの大規模な病院では，専門性の高い医師が在籍し，既に正確な診断が可能である場合が多いと考えられる．一方小規模病院では，特定の病気に対して専門性の高い知識を持つ医師は少なく，このような医療画像の診断システムを利用する利点が大きい．そのため，小規模病院でのテストデータに対する性能を考える．

次に再現率を用いる理由は，偽陰性を減らすことが重要であるためである．小規模病院では，重大な病気を患っている可能性のある患者を大規模病院に送り，より高度な診察や治療を受けさせることがある．このような場合を考えると，病気を患っている可能性の高い人をもれなく見つけ出すことが重要となる．そのため，偽陰性を減らすことが重要であり，評価指標として再現率を利用している．なお，再現率は設定する閾値によって傾向が変わってしまう可能性があるため，閾値に依存しない指標として AUROC も併せて利用する．AUROC は ROC 曲線の下部の面積にあたる値であり，以下の式 (7) で定義される FPR を横軸，再現率を縦軸に取る曲線である [11]．式 (7) で，FP は偽陽性，TN は真陰性を表す．

$$FPR = \frac{FP}{TN + FP} \tag{7}$$

クライアント種別	クライアント数	データ数	異常値の割合
大規模病院	1	30,000	50%
小規模病院	4	3,000	10%

表 1: クライアント種別ごとの実験条件

5.4 連合学習による性能向上を確認する実験

大病院と小規模病院で連合学習を行った場合に，各病院が単体で学習したものと比較して精度が向上するのかを調べる実験を行った．以下の 2 つの場合で実験を行い，大規模病院と小規模病院が連合学習を行った場合との精度の比較を行った．

（1） 小規模病院が単体で学習を行う．

（2） 大規模病院が単体で学習を行う．

実験は表 1 のような条件で行った．小規模病院にあたるクライアントを 4 つ，大規模病院にあたるクライアントを 1 つ設定し，連合学習を行う．大規模病院は 30,000 件，小規模病院は各 3,000 件のデータを所持している．各クライアントは，所持しているデータのうち 80%を学習データ，20%をテストデータとして用いる．大規模病院のデータセットには異常値が 50%含まれ，小規模病院のデータセットには異常値が 10%含まれる．データセットは，Sneaker，Bag，Ankle boot を異常値，これ以外の 7 つのラベルを正常値とラベルして用いた．学習は 10 エポック行い，各クライアントでのモデルの更新 1 回ごとに，グローバルモデルの更新を行う．

5.4.1 小規模病院が単体で学習を行う

通常の連合学習で得られたモデルと，小規模病院単体の機械学習で得られたモデルの精度を比較する．小規模病院単体で

学習方法	recall	precision	auroc
連合学習	0.922	0.889	0.993
小規模病院のみ	0.763	0.666	0.967

表 2: 連合学習と小規模病院のみでの学習結果の比較

学習方法	recall	precision	auroc
連合学習	0.922	0.889	0.993
大規模病院のみ	0.934	0.602	0.970

表 3: 連合学習と大規模病院のみでの学習結果の比較

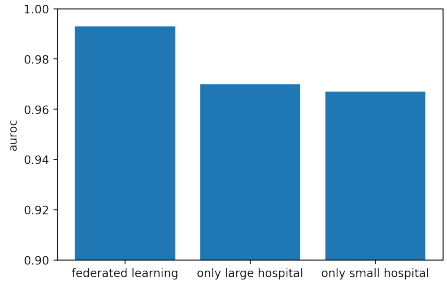


図 1: 連合学習，大規模病院のみの学習，小規模病院のみの学習それぞれの AUROC.

の機械学習は，連合学習時と同じく，データセットサイズは 3,000，異常値の割合は 10%で行った．

得られた結果は表 2 のようになった．小規模病院単体で学習した場合と比較し，連合学習を行なった場合に小規模病院の AUROC が上昇した．これは，連合学習を行うことで学習データ数が大幅に増えたためであると考えられる．

5.4.2 大規模病院が単体で学習を行う

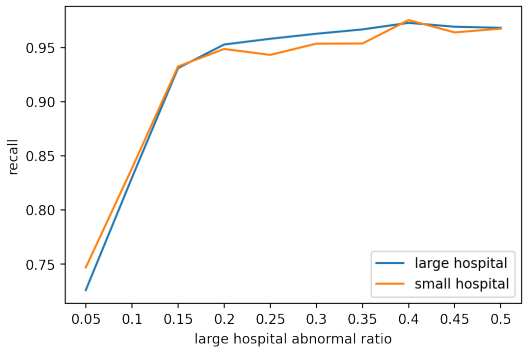
次に，大規模病院のみのデータセットで学習したモデルを小規模病院に適用した場合の精度について実験を行った．大規模病院と小規模病院で連合学習を行う場合と，大規模病院のデータセットのみで学習する場合を比較した．大規模病院のデータセットのみで学習する場合は，連合学習の場合と同じくデータセットサイズは 30,000，異常値の割合は 50%で学習を行う．

実験結果は表 3 のようになった．大規模病院のみで学習を行なったモデルと比較して，連合学習モデルの方が再現率は減少している．これは，大規模病院と小規模病院で異常値データが含まれる割合が異なることに起因していると考えられる．一方，閾値に依存しない AUROC は上昇しているため，連合学習を行う方が診断精度の高いモデルが得られることがわかった．これは連合学習を行うことで学習データ量が増加したためであると考えられる．

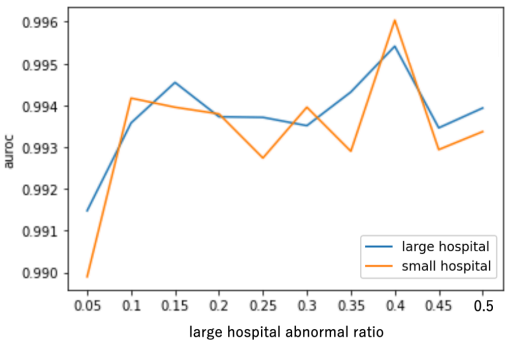
5.4.1，5.4.2 の実験結果をまとめると図 1 のようになった．3 つの条件を比較して，連合学習を行なった場合に小規模病院の AUROC が最も大きくなることがわかった．

5.5 クライアントの異常値の割合を変化させた場合の性能への影響を確認する実験

次に，大規模病院がどのようなデータセットを利用すれば精度が向上するかを確認する．実験では，大規模病院が所持する



(a) 再現率



(b) AUROC

図 2: 大規模病院のデータセットに含まれる異常値の割合を変化させた場合のモデルの (a) 再現率，(b) AUROC の値．小規模病院は各クライアントの平均をプロットした．

データセットのうち異常値のデータが占める割合を変化させ，モデルの精度の変化について調べた．1 つ目の条件では，大規模病院が持つデータセットサイズを固定とし，データセット中の異常値の占める割合を変化させた．2 つ目の条件では，大規模病院が持つ異常値のデータ数を固定とし，データセットサイズを増減させることで異常値の占める割合を変化させた．1 つ目の条件では，全病院が所持する異常値データ数の合計が変化してしまうため，この影響を排除したデータを得るために 2 つ目の条件でも実験を行った．

5.5.1 大規模病院のデータセットサイズを固定した場合

小規模病院にあたるクライアントを 4 つ，大規模病院にあたるクライアントを 1 つ設定し，連合学習を行う．大規模病院は 30,000 件，小規模病院は各 3,000 件のデータを所持している．データセットやエポック数などは，5.4 節の実験と同じ条件で行った．この実験では，大規模病院が持つデータの異常値の割合を 10%から 50%まで 5%刻みで変化させ，精度の変化を調べた．

結果は図 2 のようになった．再現率のグラフを見ると，異常値のデータを増加させることで，大規模病院，小規模病院ともに再現率が上昇していることがわかる．AUROC については振れ幅が大きく判断しづらいが，再現率とほぼ同じ傾向にあると考えられる．

次に，同様の実験を小規模病院の持つデータの異常値の割合が 5%，10%，15%，20%それぞれの場合について行い，大規模

病院と小規模病院の持つデータセットに占める異常値の割合がどのような関係にあるときに精度が高くなるのかを調べる．結果は図 3，図 4 のようになった．再現率は，大規模病院，小規模病院ともに概ね単調増加のグラフになっている．このことから大規模病院は，小規模病院の異常値の割合に関わらず，より多くの割合の異常値をデータセット含めることで再現率が上昇することがわかる．異常値の割合が 10%から離れるほど Non-IID データの度合いが大きくなるため，モデルの精度は低下すると考えられる．しかし，データセットに含まれる異常値データが増加したことによるモデルの精度の上昇の影響が大きく，全体としてモデル精度が向上したと考えられる．

5.5.2 大規模病院の異常値データの数を固定した場合

次に，大規模病院の異常値データの数を固定し，大規模病院の持つデータセットサイズを調整することで，異常値の割合を変化させる．実験は，5.4 節の実験と同じ条件で行った，

結果は図 5 のようになった．グラフより，異常値の割合が小さい間は，割合が増加するに従って再現率も増加していることがわかる．これは 5.5.1 の実験結果の再現率のグラフ（図 2 (c)）と同じ傾向である．また小規模病院の異常値が 15%より大きい範囲では，5.5.1 と比較して全体的に再現率が低くなっている．これは今回の条件下では異常値の割合が増加すると全体の学習データ数が少なくなるためであると考えられる．AUROC については試行回数が少なく，グラフから傾向を判断できなかった．

5.6 差分プライバシのモデルへの精度の影響を調べる実験

次に，プライバシー保持のために差分プライバシを満たす雑音を加えた場合に，モデルの性能にどのような影響があるのかを調べる．大規模病院のクライアントを 1 つ，小規模病院のクライアントを 4 つ設定して，DP-SGD アルゴリズムによる連合学習を行った．実験の条件は 5.4 節と同様である．差分プライバシのライブラリとして Opacus*を用いてプログラムを実装した．本実験では，δ の値は全て 10^{-5} とした．

5.6.1 差分プライバシを導入して連合学習を行う

まず，差分プライバシを導入した状態で連合学習を行うことでモデルの精度が向上するか否かを調べる．差分プライバシを適用した大規模病院単体で学習したモデルと，全ての病院に差分プライバシを適用して連合学習を行った場合のモデルで精度を調べる．

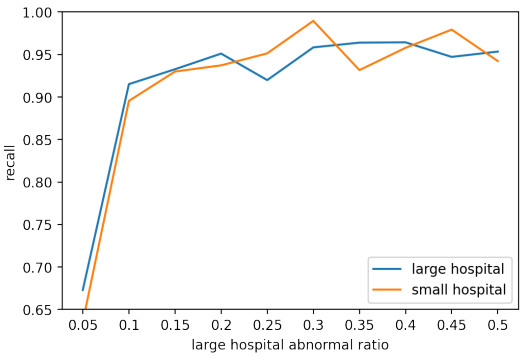
結果は図 6 のようになった．今回の条件下では，差分プライバシを導入した状態で連合学習を行った場合は，大病院単体で学習を行った場合よりも再現率，AUROC 共に低下することがわかった．

5.6.2 差分プライバシを適用するクライアントを変更する

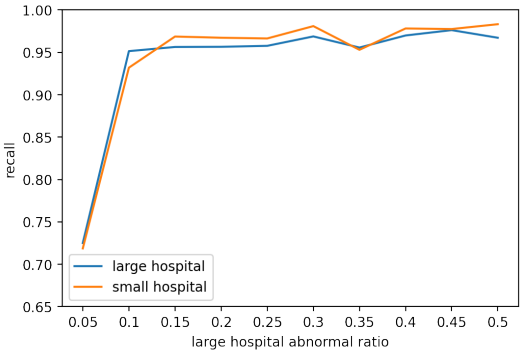
以下の三つの条件で差分プライバシを導入し，プライバシーの度合いを表す ϵ の値ごとに精度の変化を調べた．

（1） 大規模病院，小規模病院ともに差分プライバシを導入
（2） 小規模病院のみ差分プライバシを導入
（3） 差分プライバシを導入しない

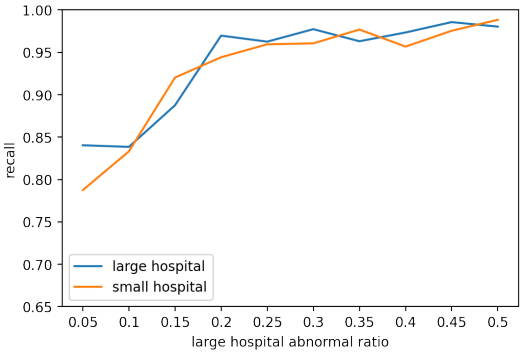
*：https://opacus.ai/api/index.html



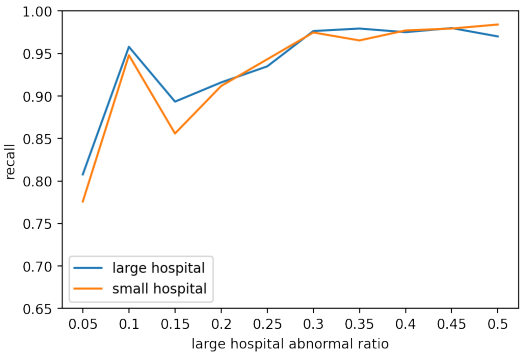
(a) 5%



(b) 10%

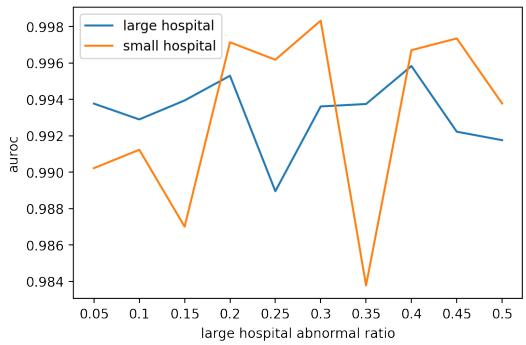


(c) 15%

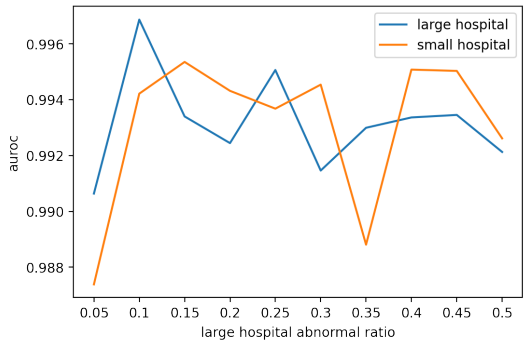


(d) 20%

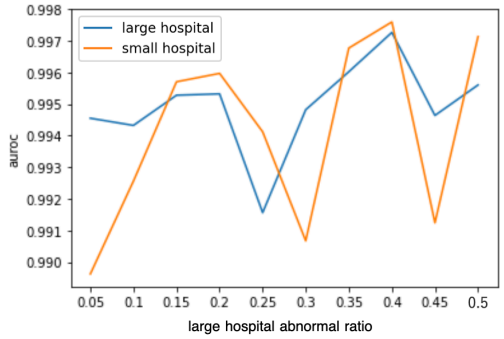
図 3: 大規模病院のデータセットに含まれる異常値の割合を変化させた場合の，大規模病院および小規模病院の再現率．小規模病院は，4 つのクライアントの平均をプロットしている．小規模病院のデータセットに含まれる異常値の割合が (a) 5%，(b) 10%，(c) 15%，(d) 20%の場合について実験を行った．



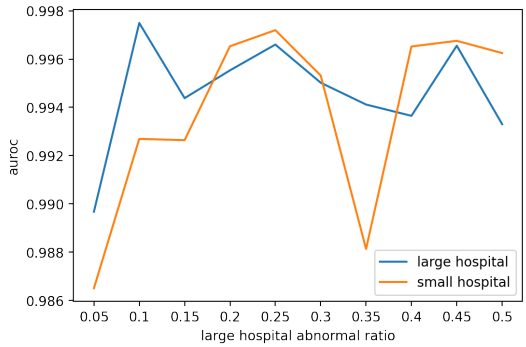
(a) 5%



(b) 10%

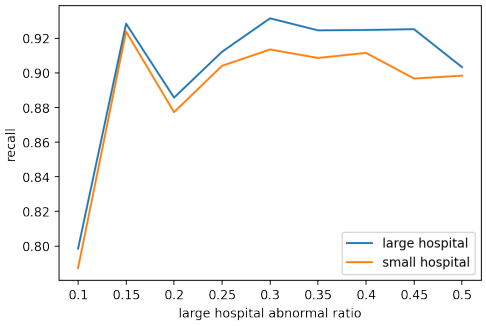


(c) 15%

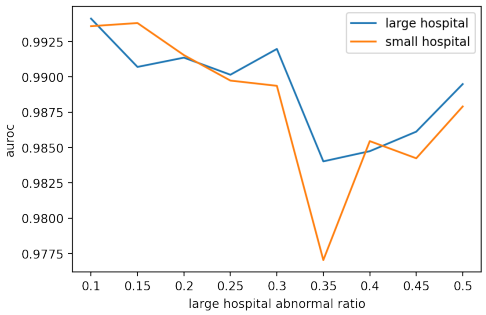


(d) 20%

図 4: 大規模病院のデータセットに含まれる異常値の割合を変化させた場合の，大規模病院および小規模病院の AUROC の値．小規模病院は，4 つのクライアントの平均をプロットしている．小規模病院のデータセットに含まれる異常値の割合が (a) 5%，(b) 10%，(c) 15%，(d) 20%の場合について実験を行った．

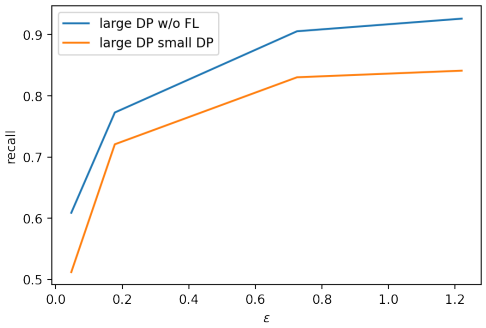


(a) 再現率

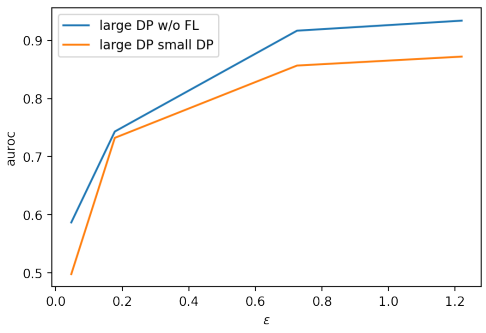


(b) AUROC

図 5: 異常値の割合を変化させた場合の (a) 再現率, (b) AUROC の値．小規模病院の値は各クライアントの平均をプロットした．



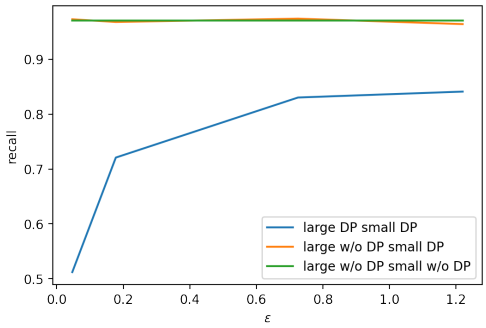
(a) 再現率



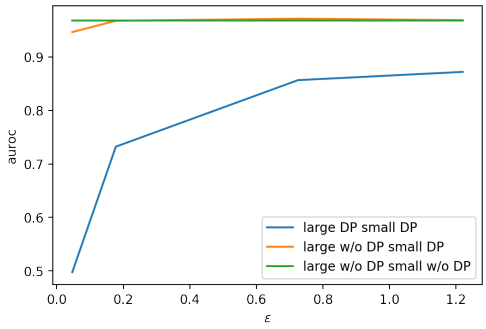
(b) AUROC

図 6: 差分プライバシを導入した状態で，大規模病院単体で学習を行った場合と連合学習を行った場合の (a) 再現率，(b) AUROC の値．

結果は図 7 のようになった．ϵ が 0.2 より大きい範囲では，小規模病院への差分プライバシ導入の有無は，モデルの再現率や AUROC への影響は少ないことがわかった．一方，大規模病院



(a) 再現率



(b) AUROC

図 7: 全ての病院に差分プライバシを導入，小規模病院のみに差分プライバシを導入，差分プライバシを導入しない場合の (a) 再現率，(b) AUROC の値.

への差分プライバシの導入の影響は大きい．このことから，モデルの精度を低下させず，小規模病院のプライバシーを保護することが可能であることがわかった.

6 結　　論

本研究では，実社会での医療画像を用いた連合学習において，クライアント間のデータ分布の不均一性がモデルの性能に与える影響を考察した．具体的には，データセットサイズが大きく，異常値のデータが占める割合も大きいという特徴をもつ専門病院と，データセットサイズが小さく，異常値のデータが占める割合も小さいという特徴を持つ小規模な病院を設定し連合学習を行い，異常値の割合や差分プライバシの有無がモデルの性能に与える影響を考察した.

7 今後の課題

本研究では FashionMNIST データセットを用いて実験を行った．これを実際の医療画像のデータセットで行った場合にも，同様の結果となるのかを確認したい.

また，文献 [5] で挙げられている，データベース，アルゴリズムベース，システムベースの手法についても，今回の事例で有効であるか検証したい．特に，データベースのアプローチとして [4] で挙げらている手法については医療画像のケースでは有効であると考えている．[4] ではデータセットのサブセットを共有データセットとして複数クライアント間で共有することにより精度を上昇させている．今回の病院の事例では，事前に患者の同意を得られたデータのみを共有データセットとして用いることで精度向上に寄与できるのではないかと考えている.

謝　　辞

本研究は，JST CREST JPMJCR21M2, JST SICORP JPMJSC2107, 科学研究費 21K19767, 19K20269 の支援を受けたものである.

文　　献

[1] Qiang Yang, Yang Liu, Yong Cheng, Yan Kang, Tianjian Chen, and Han Yu. Federated learning. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, Vol. 13, No. 3, pp. 1–207, 2019.

[2] Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas. Communication-efficient learning of deep networks from decentralized data. In *Artificial intelligence and statistics*, pp. 1273–1282. PMLR, 2017.

[3] Xiaoxiao Li, Yufeng Gu, Nicha Dvornek, Lawrence H Staib, Pamela Ventola, and James S Duncan. Multi-site fmri analysis using privacy-preserving federated learning and domain adaptation: Abide results. *Medical Image Analysis*, Vol. 65, p. 101765, 2020.

[4] Yue Zhao, Meng Li, Liangzhen Lai, Naveen Suda, Damon Civin, and Vikas Chandra. Federated learning with non-iid data. *arXiv preprint arXiv:1806.00582*, 2018.

[5] Hangyu Zhu, Jinjin Xu, Shiqing Liu, and Yaochu Jin. Federated learning on non-iid data: A survey. *arXiv preprint arXiv:2106.06843*, 2021.

[6] Alireza Fallah, Aryan Mokhtari, and Asuman Ozdaglar. Personalized federated learning with theoretical guarantees: A model-agnostic meta-learning approach. *Advances in Neural Information Processing Systems*, Vol. 33, pp. 3557–3568, 2020.

[7] Cynthia Dwork. Differential privacy. In *International Colloquium on Automata, Languages, and Programming*, pp. 1–12. Springer, 2006.

[8] Martin Abadi, Andy Chu, Ian Goodfellow, H Brendan McMahan, Ilya Mironov, Kunal Talwar, and Li Zhang. Deep learning with differential privacy. In *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security*, pp. 308–318, 2016.

[9] Robin C Geyer, Tassilo Klein, and Moin Nabi. Differentially private federated learning: A client level perspective. *arXiv preprint arXiv:1712.07557*, 2017.

[10] Han Xiao, Kashif Rasul, and Roland Vollgraf. Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms. *arXiv preprint arXiv:1708.07747*, 2017.

[11] Sarang Narkhede. Understanding auc-roc curve. *Towards Data Science*, Vol. 26, pp. 220–227, 2018.