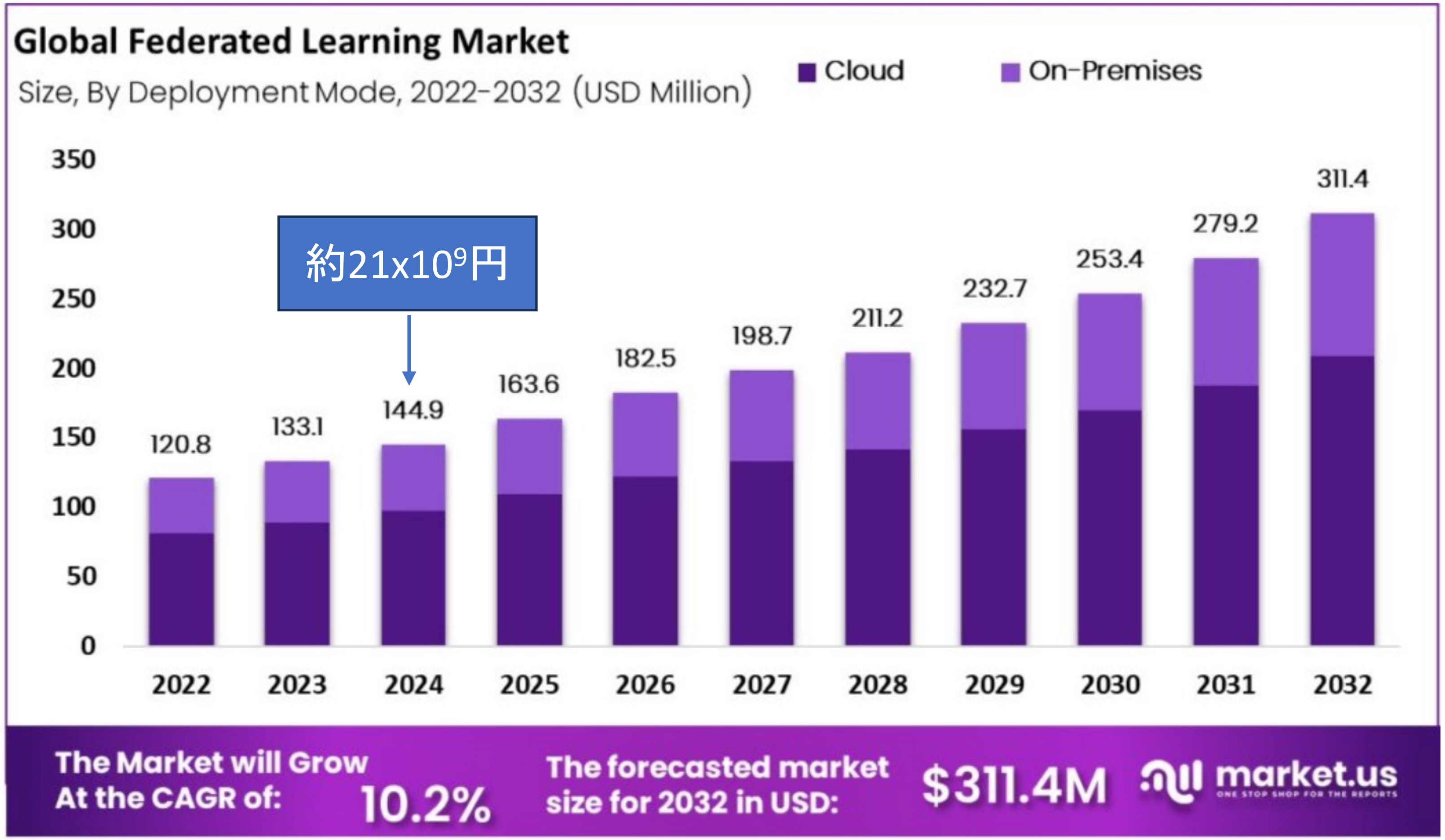
情報セキュリティと分散型AIの融合:
データのプライバシーとユーティリティの保護

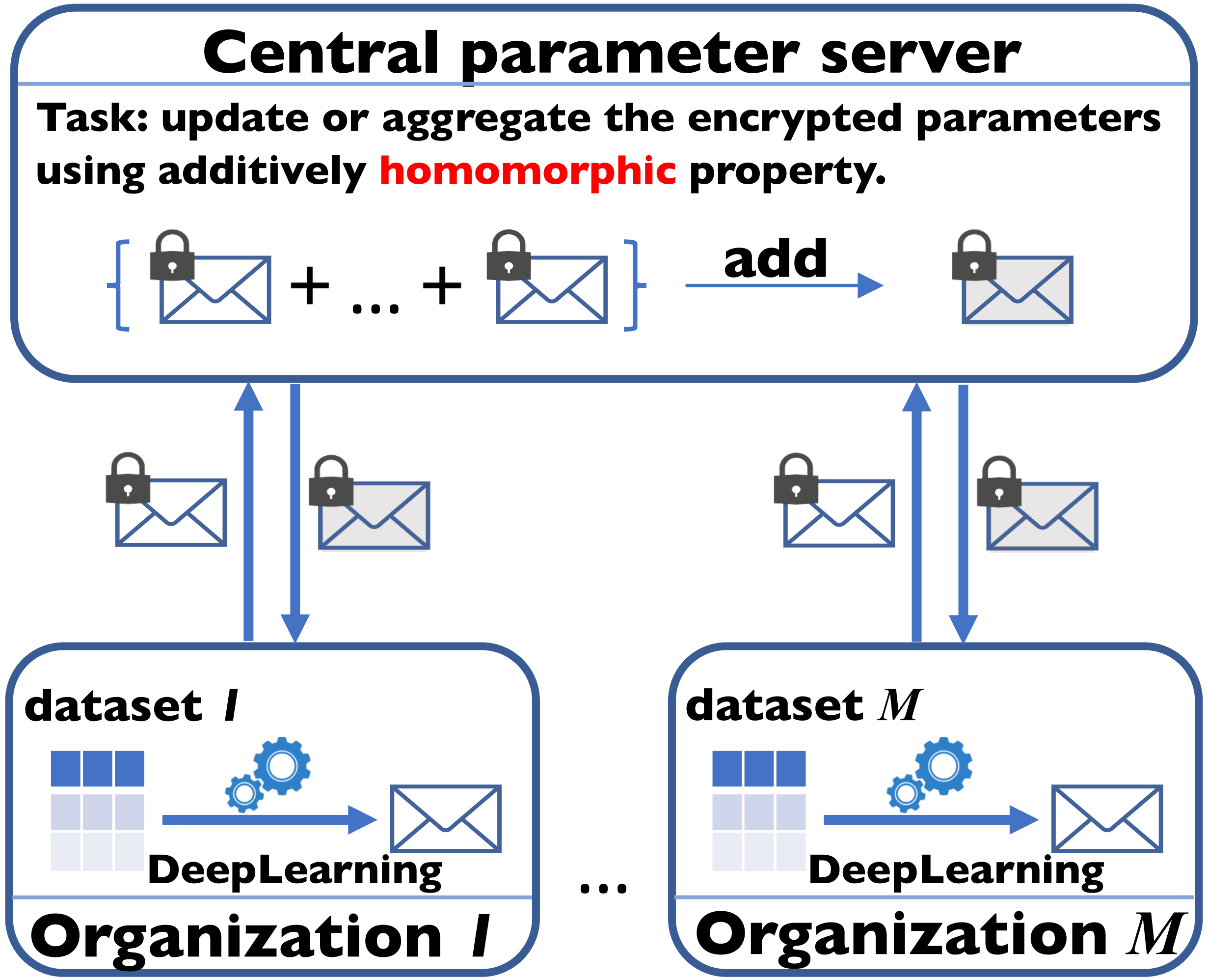
LE TRIEU PHONG
セキュリティ基盤研究室 主任研究員

世界の連合学習の市場（2032まで予測）



https://market.us/report/federated-learning-market/

プライバシー保護連合学習技術 DeepProtect (2018)



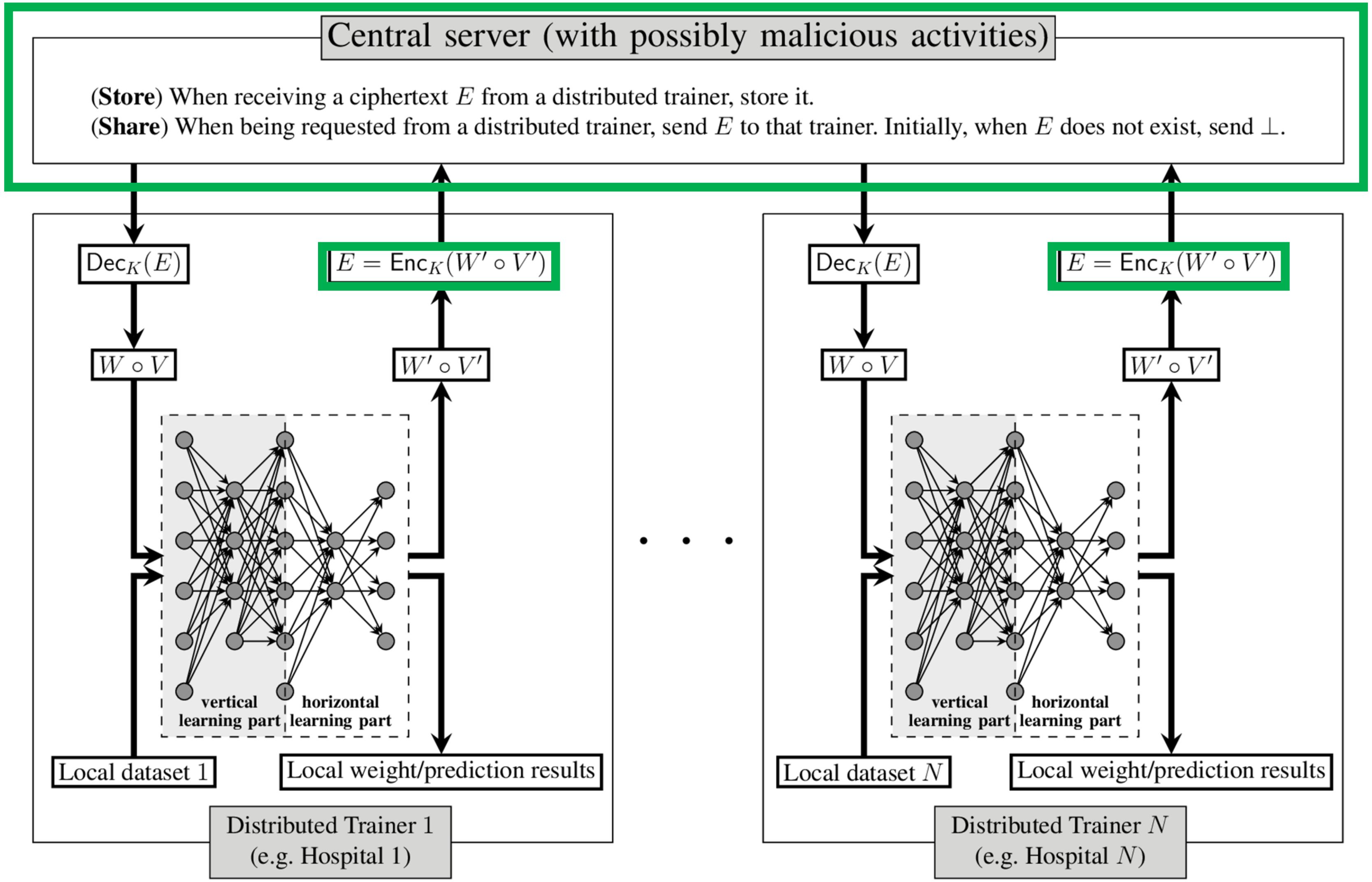
●IEEE Trans. Inf. Forensics Secur. (2018)に掲載

●Google Scholarによる引用数1300以上

●2023 IEEE SPS Best Paper Award受賞

Le Trieu Phong, Yoshinori Aono, Takuya Hayashi, Lihua Wang, Shiho Moriai:
Privacy-Preserving Deep Learning via Additively Homomorphic Encryption. IEEE Trans. Inf. Forensics Secur. (2018)

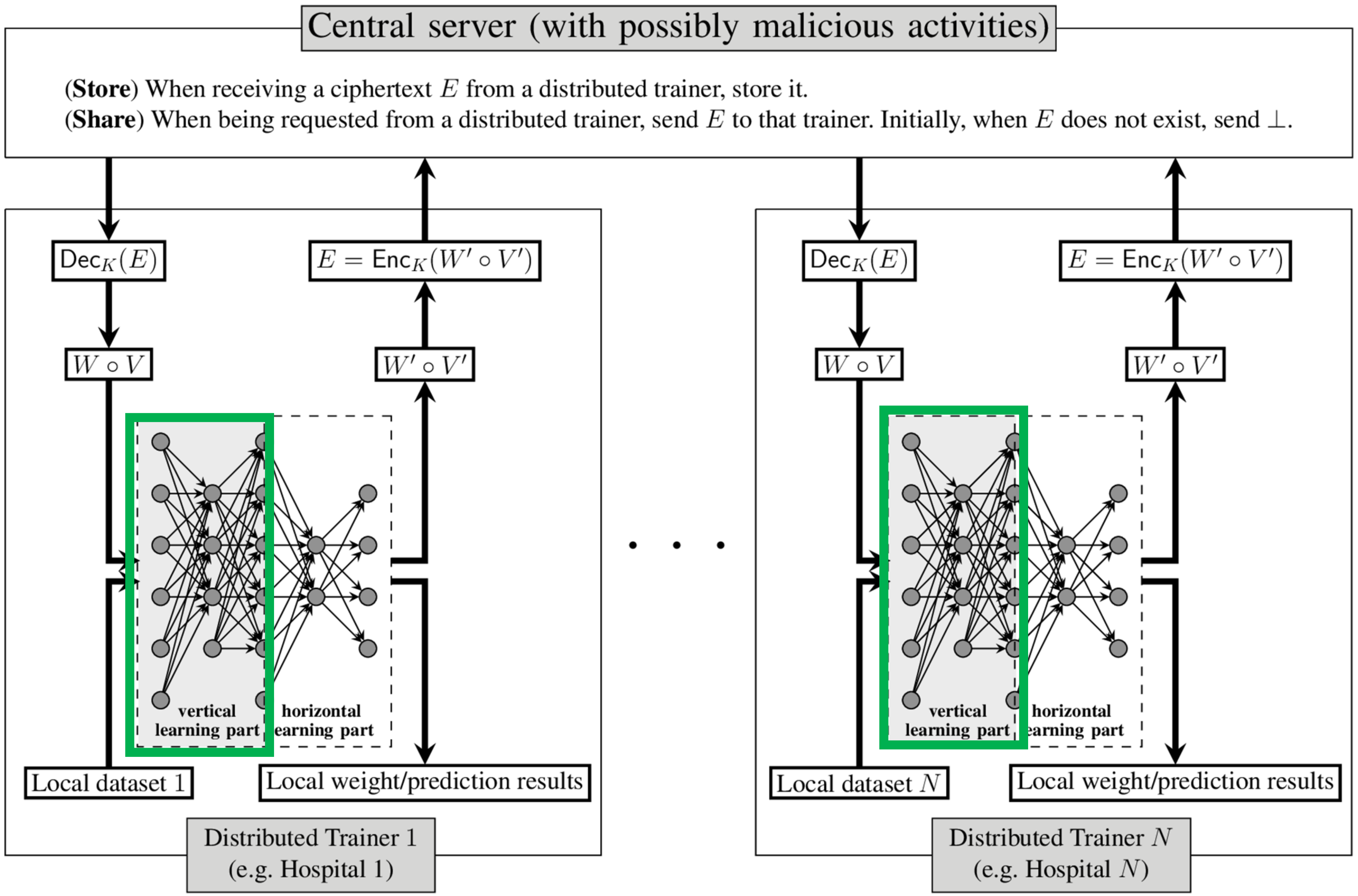
DeepProtect (2018)以降の研究



認証付き暗号により、暗号文を変える攻撃者を対応!

Le Trieu Phong: Secure deep learning for distributed data against malicious central server. PLoS ONE 17(8), 2022

DeepProtect (2018)以降の研究



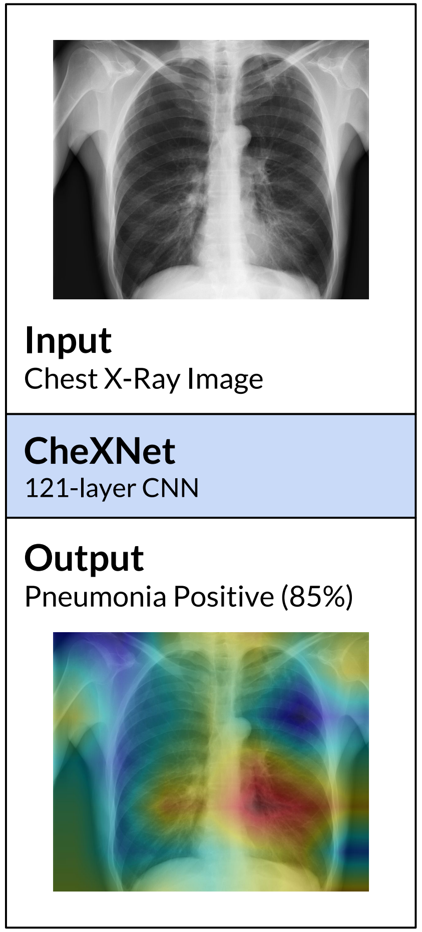
認証付き暗号により、暗号文を変える攻撃者を対応!

一部のパラメータは、別のデータセットの学習結果でもOK!

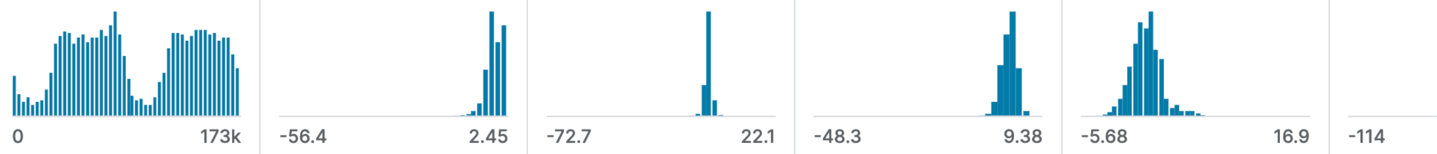
Le Trieu Phong: Secure deep learning for distributed data against malicious central server. PLoS ONE 17(8), 2022

オープンデータを用いた実験

- **<u>医療データ</u>**
  1.MRI
  2.X-Ray画像



- **<u>クレジットカードのデータ</u>**



0	-1.3598071336738	-0.0727811733098497	2.53634673796914	1.37815522427443	-0.33832
0	1.19185711131486	0.26615071205963	0.16648011335321	0.448154078460911	0.060017
1	-1.35835406159823	-1.34016307473609	1.77320934263119	0.379779593034328	-0.50319

MRI: http://www.riteh.uniri.hr/~istajduh/projects/kneeMRI/
X-Ray: https://stanfordmlgroup.github.io/projects/chexnet/
Credit Card: https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud

MRIデータ

MRI1:

- スタンフォード大学医療センターで取得。
- 1,370 件の膝 MRI 検査。
- ラベルは臨床レポートから手動で抽出。

MRI2:

- クロアチアの病院センターで取得。
- 917件の膝 MRI 検査。

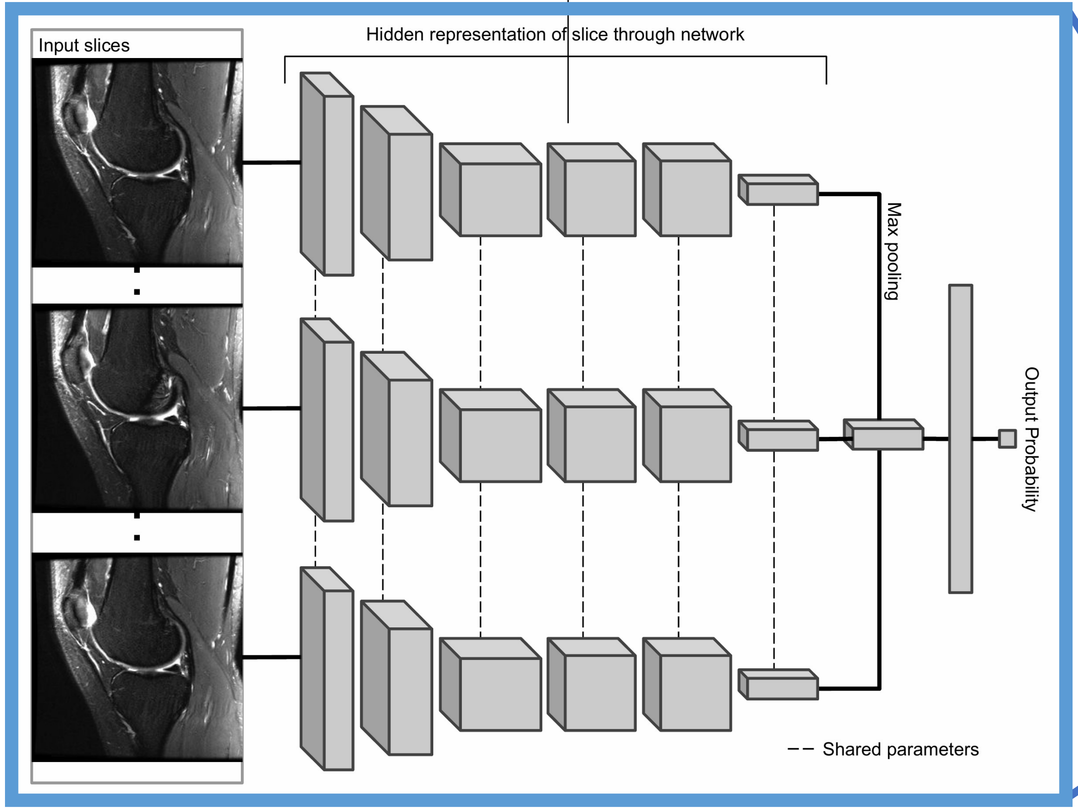
MRI: http://www.riteh.uniri.hr/~istajduh/projects/kneeMRI/

目的：膝の前十字靭帯断裂 (ACL tears)を推測

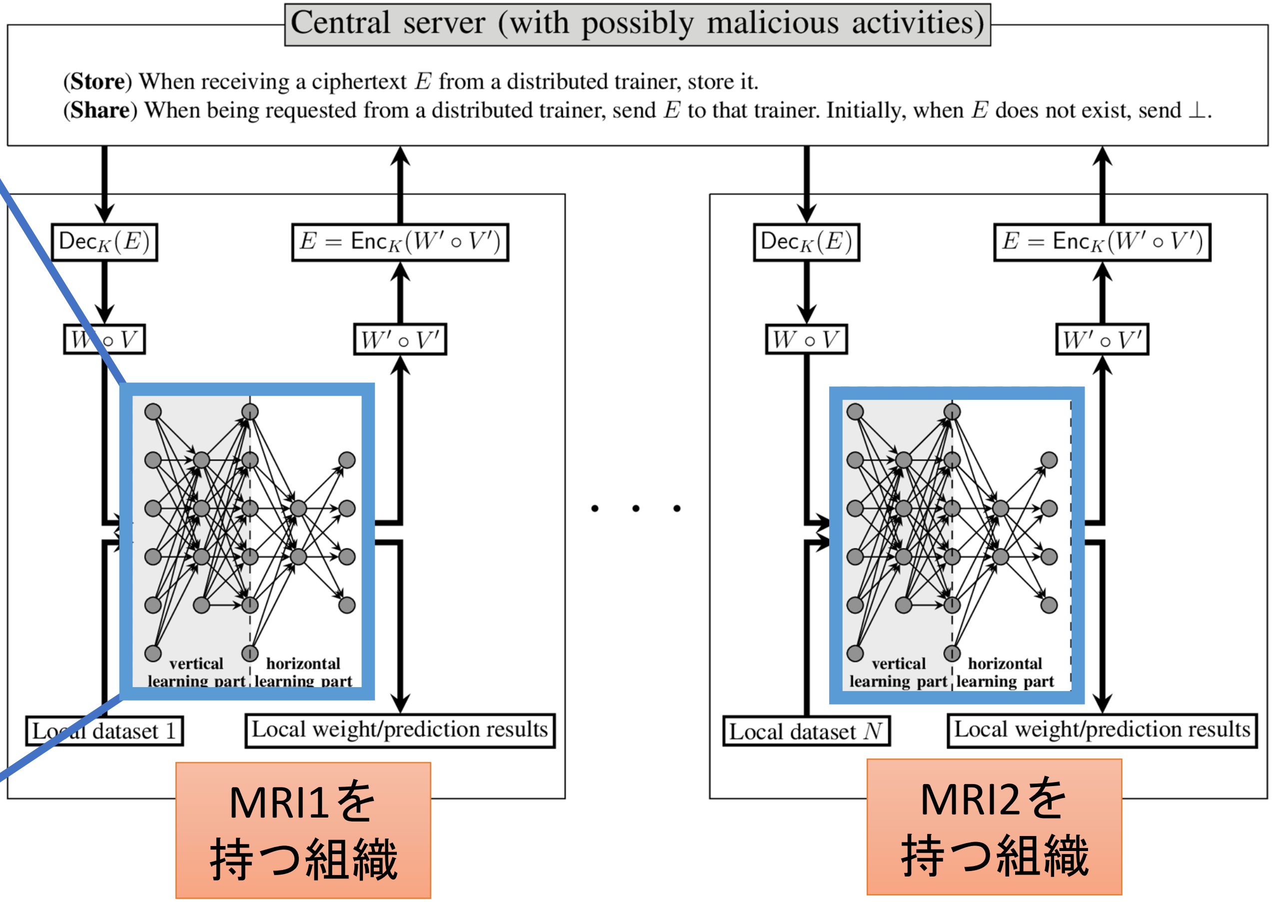
MRNet：我々のシステムの部品

AlexNet

MRNet



https://stanfordmlgroup.github.io/projects/mrnet/



Le Trieu Phong, PLoS ONE 17(8), 2022

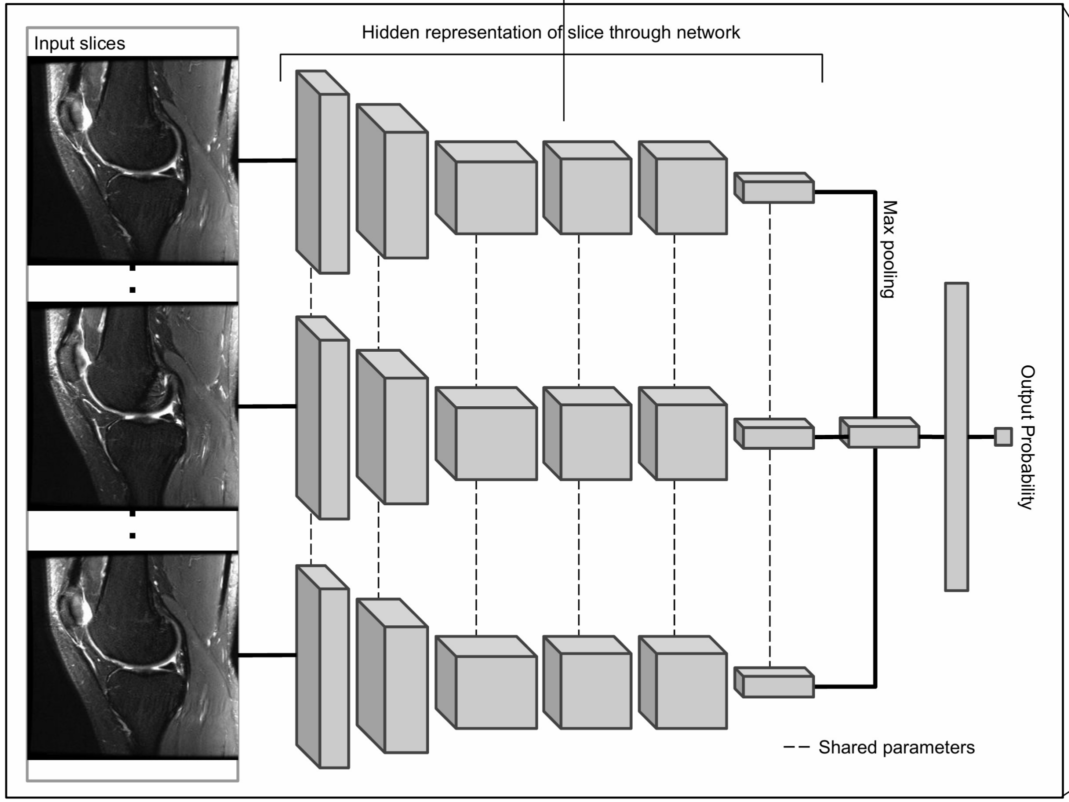
MRNet：我々のシステムの部品

AlexNet

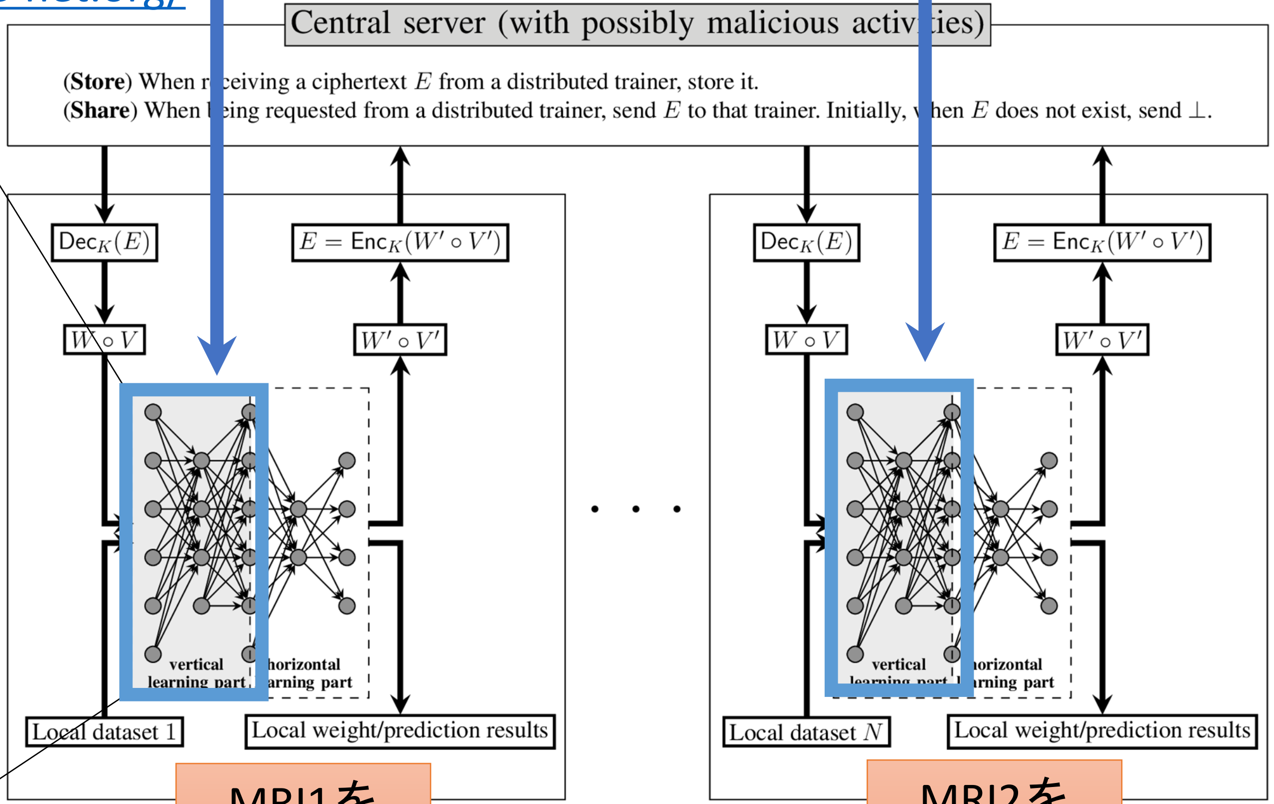
学習結果を良くするために
ImageNetデータセットで学習されたパラメータで初期化

https://www.image-net.org/

MRNet



https://stanfordmlgroup.github.io/projects/mrnet/



MRI1を
持つ組織

MRI2を
持つ組織

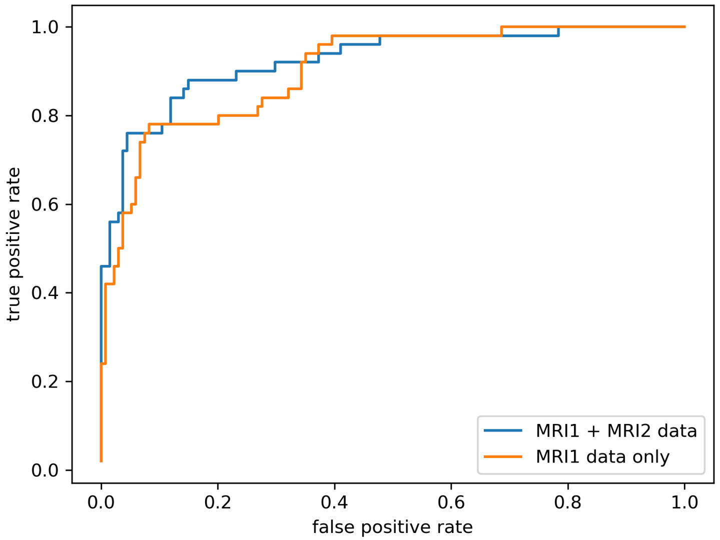
Le Trieu Phong, PLoS ONE 17(8), 2022

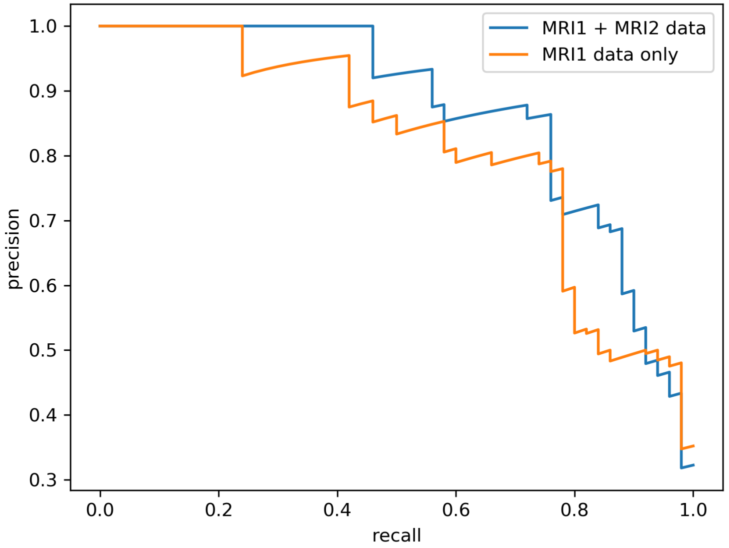
我々のシステムを用い、MRIに関わる実験結果

Area-under-the-curve (AUC) scores of learning methods on MRI datasets.

Paper	Method	AUC score
Stajduhar et al. [15]	Support Vector Machine	0.894
Bien et al. [14]	Neural Network	0.824
Bien et al. [14]	Neural Network	0.911
我々のシステム　（Train on MRI1 + MRI2）	Neural Network	0.924

MRI2 のテストセットで予測





Le Trieu Phong: Secure deep learning for distributed data against malicious central server. PLoS ONE 17(8), 2022
Le Trieu Phong, Tran Thi Phuong, Lihua Wang, Seiichi Ozawa: Frameworks for Privacy-Preserving Federated Learning. IEICE Trans. Inf. Syst. 107(1): 2-12(2024)

学習の詳細 (MRIデータ)

Model (MRNet)	6,100万個のパラメータ (ハードディスクに234MB)
CBC-encrypt-then-mac	3 秒
Training on GPU (one epoch)	13 秒

X-Rayデータ ： NIH ChestX-ray14データセット

NEWS RELEASES

Media Advisory Wednesday, September 27, 2017

NIH Clinical Center provides one of the largest publicly available chest x-ray datasets to scientific community

The dataset of scans is from more than 30,000 patients, including many with advanced lung disease.

https://www.nih.gov/news-events/news-releases/nih-clinical-center-provides-one-largest-publicly-available-chest-x-ray-datasets-scientific-community

ChestX-ray14 データセット

- 14 の異なる胸部疾患に個別にラベル付けされる。
- 112,120 枚の正面胸部 X 線画像がある。

既存結果

AUCスコア（高い方が良い）

(Stanford大学)

病理

Pathology	Wang et al. (2017)	Yao et al. (2017)	CheXNet (ours)
Atelectasis	0.716	0.772	**0.8094**
Cardiomegaly	0.807	0.904	**0.9248**
Effusion	0.784	0.859	**0.8638**
Infiltration	0.609	0.695	**0.7345**
Mass	0.706	0.792	**0.8676**
Nodule	0.671	0.717	**0.7802**
Pneumonia	0.633	0.713	**0.7680**
Pneumothorax	0.806	0.841	**0.8887**
Consolidation	0.708	0.788	**0.7901**
Edema	0.835	0.882	**0.8878**
Emphysema	0.815	0.829	**0.9371**
Fibrosis	0.769	0.767	**0.8047**
Pleural Thickening	0.708	0.765	**0.8062**
Hernia	0.767	0.914	**0.9164**

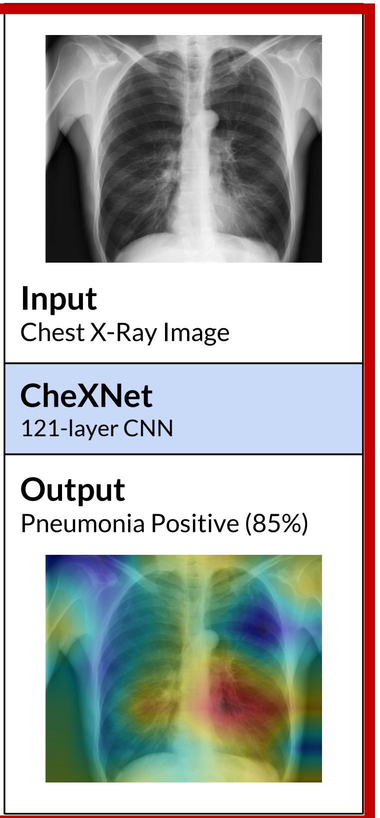


Table 2. CheXNet outperforms the best published results on all 14 pathologies in the ChestX-ray14 dataset. In detecting Mass, Nodule, Pneumonia, and Emphysema, CheXNet has a margin of >0.05 AUROC over previous state of the art results.

https://stanfordmlgroup.github.io/projects/chexnet/

NIH ChestX-ray14 に関わる実験結果

Area-under-the-curve (AUC) scores of learning methods on ChestX-ray14.

病理	Wang et al. [13]	Yao et al. [70]	Zech [71]	我々のシステム	CheXNet [3]
Atelectasis	0.716	0.772	0.8161	0.8176	0.8094
Cardiomegaly	0.807	0.904	0.9105	0.9143	0.9248
Effusion	0.784	0.859	0.8839	0.8842	0.8638
Infiltration	0.609	0.695	0.7077	0.7098	0.7345
Mass	0.706	0.792	0.8308	0.8494	0.8676
Nodule	0.671	0.717	0.7748	0.7829	0.7802
Pneumonia	0.633	0.713	0.7651	0.7675	0.7680
Pneumothorax	0.806	0.841	0.8739	0.8762	0.8887
Consolidation	0.708	0.788	0.8008	0.8077	0.7901
Edema	0.835	0.882	0.8979	0.8931	0.8878
Emphysema	0.815	0.829	0.9227	0.9340	0.9371
Fibrosis	0.769	0.767	0.8293	0.8258	0.8047
Pleural Thickening	0.708	0.765	0.7860	0.7851	0.8062
Hernia	0.767	0.914	0.9010	0.9087	0.9164
Average	**0.7381**	**0.8027**	**0.8358**	**0.8397**	**0.8414**
Securely distributed training?	**no**	**no**	**no**	**yes**	**no**

https://doi.org/10.1371/journal.pone.0272423.t004

Le Trieu Phong, PLoS ONE 17(8), 2022

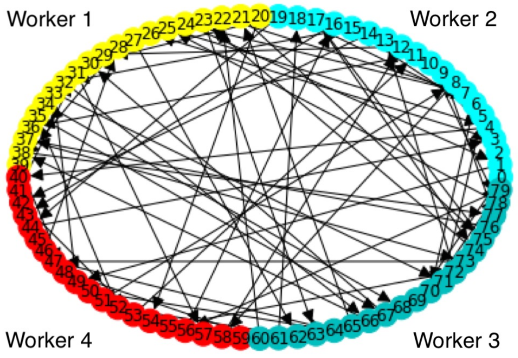
元のデータを分け、4 組織を想定

学習の詳細 (NIH ChestX-ray14 データ)

Model (DenseNet-121)	約 700 万個のパラメータ (ハードディスク 28 MB)
CBC-encrypt-then-mac	0.2 秒
Training on GPU (one epoch)	60 秒

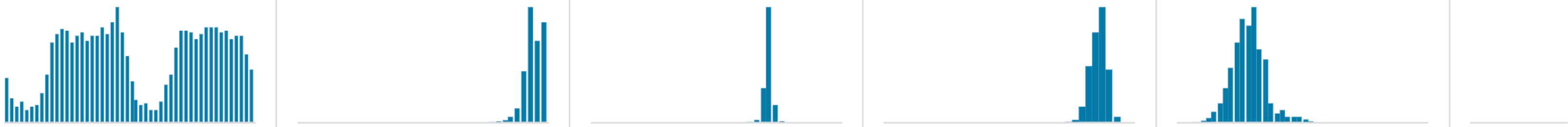
工夫したポイントの一つ：

- 早期に重みパラメータを共有して精度を向上させる。



Le Trieu Phong, Tran Thi Phuong, Lihua Wang, Seiichi Ozawa:
Frameworks for Privacy-Preserving Federated Learning. IEICE Trans. Inf. Syst. 107(1): 2-12(2024)

クレジットカード不正行為の検出のデータセット



# Time	# V1	# V2	# V3	# V4	# V5
Number of seconds elapsed between this transaction and the first transaction in the dataset	may be result of a PCA Dimensionality reduction to protect user identities and sensitive features(v1-v28)				
0 – 173k	-56.4 – 2.45	-72.7 – 22.1	-48.3 – 9.38	-5.68 – 16.9	-114
0	-1.3598071336738	-0.0727811733098497	2.53634673796914	1.37815522427443	-0.33832
0	1.19185711131486	0.26615071205963	0.16648011335321	0.448154078460911	0.060017
1	-1.35835406159823	-1.34016307473609	1.77320934263119	0.379779593034328	-0.50319
1	-0.966271711572087	-0.185226008082898	1.79299333957872	-0.863291275036453	-0.01030

https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud

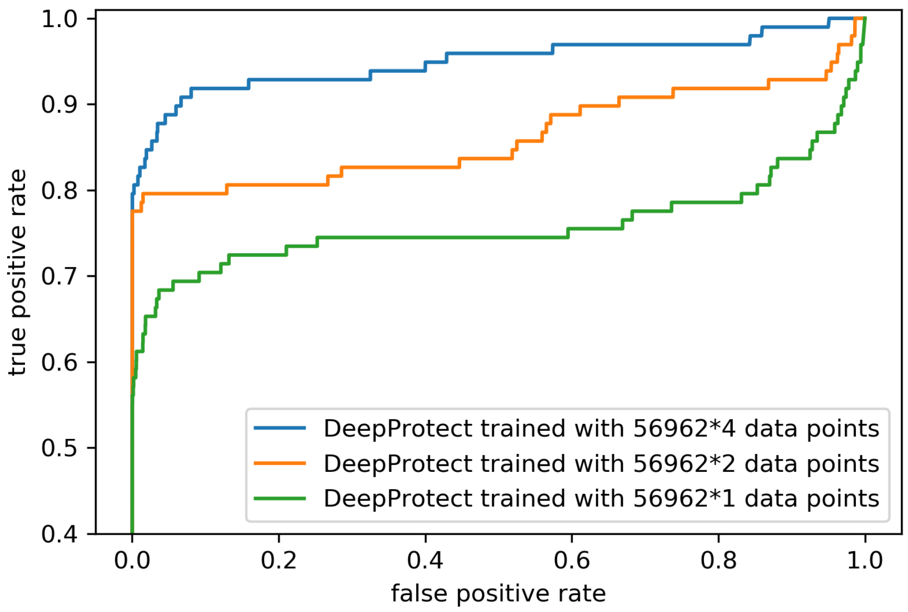
- ヨーロッパのカード所有者によって行われた取引が含まれている。
- 284,807 件の取引のうち 492 件の不正行為が発生した。
- ポジティブ クラス (詐欺) は全取引の 0.172% を占める。

クレジットカードのデータの実験結果

各学習参加者にある
ニューラルネットワークの構成

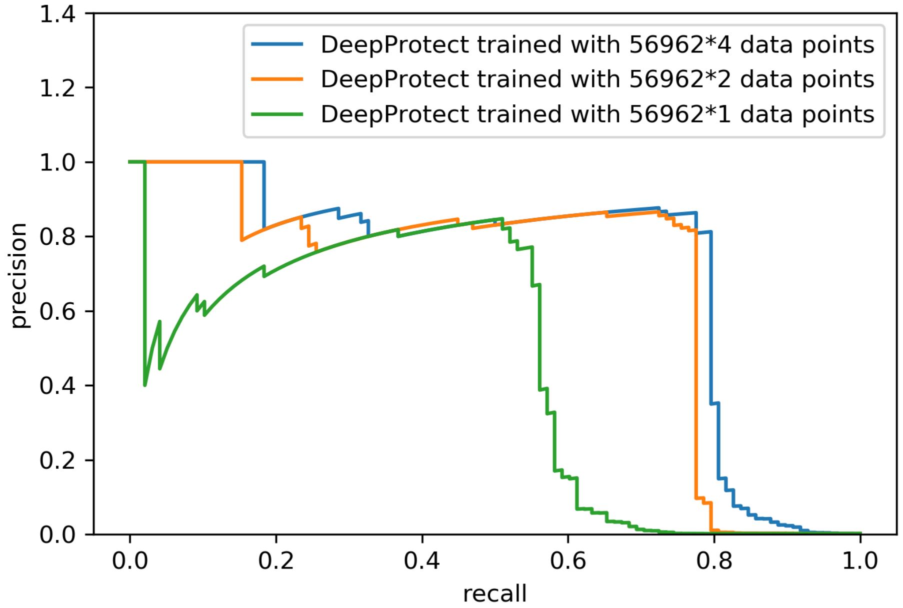
```
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
dense_1 (Dense)              (None, 64)                1984
_________________________________________________________________
dropout_1 (Dropout)          (None, 64)                0
_________________________________________________________________
dense_2 (Dense)              (None, 32)                2080
_________________________________________________________________
dropout_2 (Dropout)          (None, 32)                0
_________________________________________________________________
dense_3 (Dense)              (None, 1)                 33
=================================================================
Total params: 4,097
Trainable params: 4,097
Non-trainable params: 0
_________________________________________________________________
```

Le Trieu Phong, Tran Thi Phuong, Lihua Wang, Seiichi Ozawa: Frameworks for Privacy-Preserving Federated Learning. IEICE Trans. Inf. Syst. 107(1): 2-12(2024)



データが多いほど、
True positive rateが1
False positive rateが0
に近づく。

データが多いほど、学習結果が良くなる！



データが多いほど、
Precisionが1
Recall (True Positive Rate)が1
に近づく。

今後の見通し

- 連合学習には利点があるが、学習に伴うすべての課題（例：データの品質、データの標準化、システムの運用）に対処できるわけではない。
- 連合学習の取り組みと非連合学習の取り組みの両方が必要である。
- すべての技術的な疑問がまだ解決されているわけではない。これからも、活発な研究開発となるだろう。