Federated Semi-Supervised Medical Image Classification via Inter-Client Relation Matching

Muhammed Said ÖZDEMİR
mohammedozdamir@gmail.com

Mais SABBAGH
maissabbagh24@gmail.com

Assoc. Prof. Ömer KORÇAK

Prof. Dr. Çiğdem EROĞLU ERDEM

PROBLEM DESCRIPTION

Problems:

* ✳ **Label scarcity** in healthcare data.
* ✳ **Data privacy** concerns for patient information.

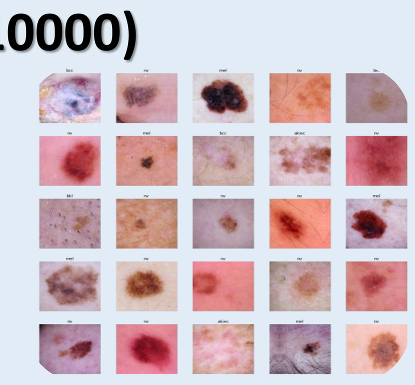
Solutions:

* ✳ **Semi-supervised learning** for leveraging unlabeled data.
* ✳ **Federated learning** for decentralized, privacy-preserving model training.

DATASETS

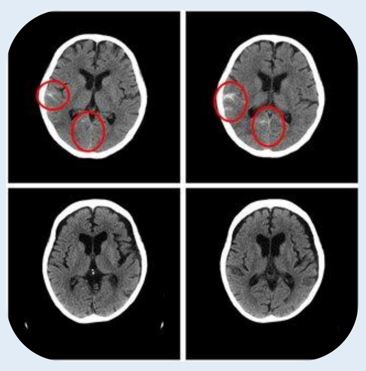
Human Against Machine (HAM10000)

* ✳ Large, multi source images of pigmented skin lesions
* ✳ 10015 Dermoscopic images.
* ✳ Images categorized into 7 classes.



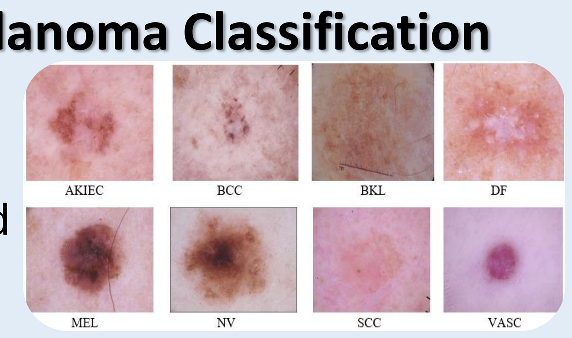
RSNA Intracranial Hemorrhage

* ✳ Brain Hemorrhage dataset.
* ✳ 753,000 training images and 121,000 testing images.
* ✳ CT scans categorized into 5 classes of intracranial brain hemorrhaging



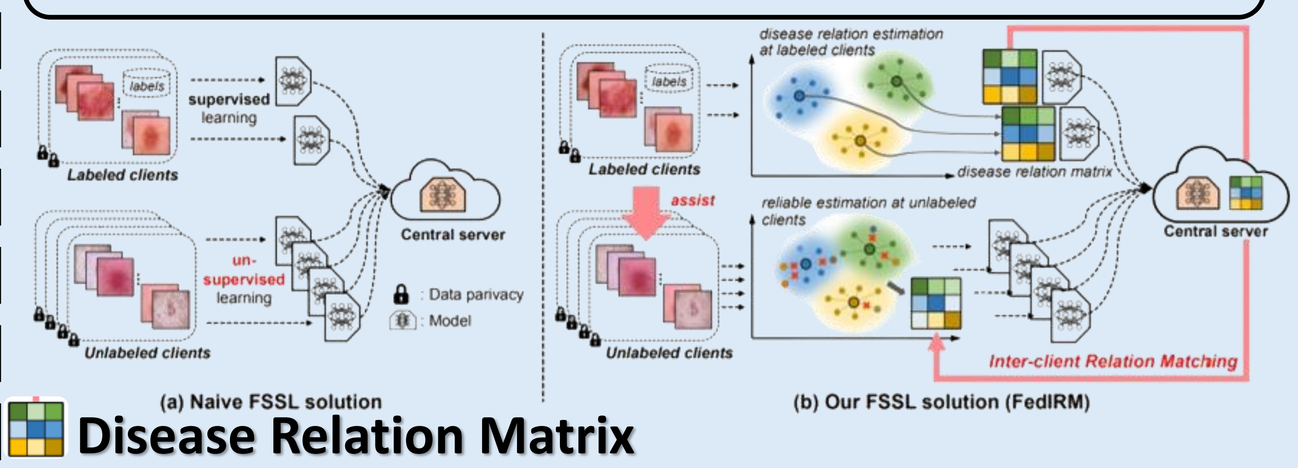
Skin Lesion Images for Melanoma Classification

* ✳ Oncological dataset.
* ✳ 25,331 images.
* ✳ Dermoscopic images categorized into 8 disease classes.



FedIRM

An algorithm combining Semi-Supervised and Federated Learning to leverage unlabeled data with local training at each client. [1]



Disease Relation Matrix

* ✳ **Mean feature vector** computed for each class.
* ✳ Mean vectors scaled to **soft labels** using softened softmax function. **(Equation 1)**
* ✳ Combination of all soft labels form a **disease relation matrix.** **(Equation 2)**
* ✳ Labeled clients' DRMs averaged to form **general disease relation matrix.**
* ✳ Disease relation matrix utilized to aid unlabeled clients in learning.

1
$$V_c^l = \frac{1}{N_c^l}\sum_{i=1}^{N^l} \mathbf{1}_{[y_i^l=c]}\hat{f}_{\theta^l}(x_i^l)$$

2
$$\mathcal{M}^l = [\, S_1^l, \ldots, S_C^l]$$

Disease Relation Estimation at Unlabeled Clients

* ✳ Predicted probability vectors generate **pseudo labels**.
* ✳ **Predictive entropy** filters out unreliable pseudo labels.
* ✳ Remaining pseudo labels estimate **disease relation matrix**. **(Equation 3)**
* ✳ Unlabeled clients align disease relation matrix with labeled clients'.

3
$$V_c^u = \frac{\sum_{i=1}^{B} \mathbb{1}_{[(y_i=c).(w_i^u<h)]} \cdot p_i^u}{\sum_{i=1}^{B} \mathbb{1}_{[(y_i=c).(w_i^u<h)]}} \quad M^u = [S_1^u, \ldots \ldots, S_C^u]$$

Learning Objectives and IRM Loss Function

* ✳ Labeled clients employ **Cross Entropy loss**.**(Equation 4)**
* ✳ Unlabeled clients utilize **Consistency regularization**.**(Equation5)**
* ✳ **KL-Divergence** between disease relation matrices of labeled and unlabeled clients minimized for IRM loss.**(Equation 6)**

4
$$\mathcal{L}^l = \mathcal{L}_{ce}(D^l, \theta^l)$$

5
$$\mathcal{L}^u = \lambda(\omega)(\mathcal{L}_c + \mathcal{L}_{IRM})$$

6
$$\mathcal{L}_{IRM} = \frac{1}{C}\sum_{c=1}^{C}(\mathcal{L}_{KL}(\mathcal{M}_c||\mathcal{M}_c^u) + \mathcal{L}_{KL}(\mathcal{M}_c^u||\mathcal{M}_c)) \textbf{ with}$$
$$\mathcal{L}_{KL}(\mathcal{M}_c||\mathcal{M}_c^u) = \sum_j \mathcal{M}_{c(j)} \log\frac{\mathcal{M}_{c(j)}}{\mathcal{M}_{c(j)}^u}$$

TECHNOLOGIES USED

Imbalanced learn

kaggle

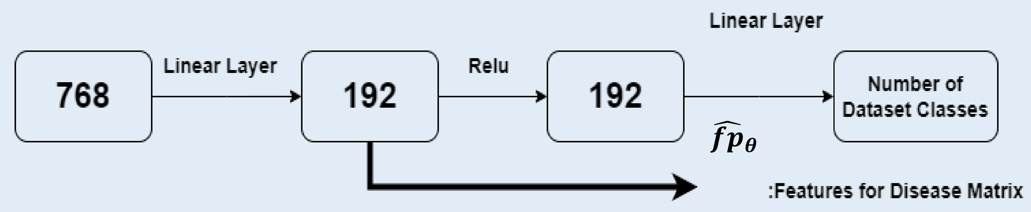
IMPROVED FedIRM

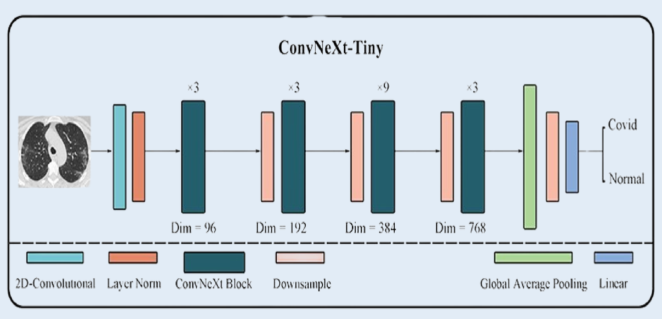
Changing the Backbone Architecture

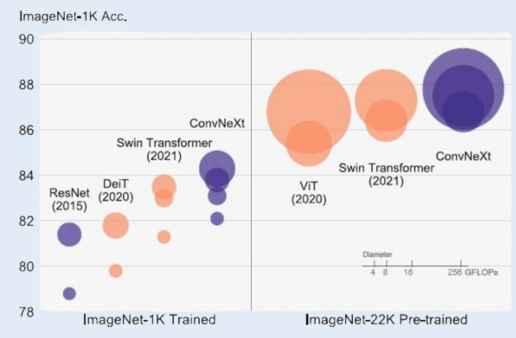
This improvement involves changing the backbone architecture from **DenseNet121** to **ConvNeXt Tiny**.

Why ConvNeXt ?

* ✳ Features **"Bottleneck Residual Blocks"** , reducing number of parameters for learning.
* ✳ More parameters for **complex feature learning** and enhanced accuracy.
* ✳ Employs **Global Response Normalization** "GRN" , to enhance feature diversity and mitigate feature collapse.
* ✳ Employs **"GELU activation"**, for improving network efficiency.
* ✳ We added a **linear layer** for a richer representation of the disease relation matrix using more features.







Covariance Matrix for Class Distributions

* ✳ Represents **class distribution** with pre-category mean feature vectors.**(Equation 7)**
* ✳ Alternative to original **Disease Relation Matrix**.
* ✳ Vectorized, normalized matrices for each class combined into a single matrix. **(Equation 9)**
* ✳ Similar computation for unlabeled clients.
* ✳ Goal: Minimize KL divergence for **better accuracy** in unlabeled clients. **(Equation 8)**

7
$$M_c^l = \frac{1}{N_c^l}\sum_{i=1}^{N_c^l}\mathbb{1}_{[y_i^l=c]}[\hat{f}_{\theta^l}(x_i^l) - V_c^l][\hat{f}_{\theta^l}(x_i^l) - V_c^l]^T$$

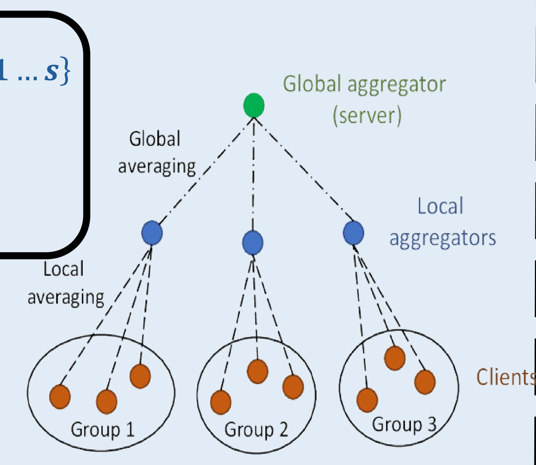
8
$$\mathcal{L}_{IRM,COV} = \frac{1}{C}\sum_{c=1}^{C}(\mathcal{L}_{KL}(M_{cov}^l||M_{cov}^u) + \mathcal{L}_{KL}(M_{cov}^u||M_{cov}^l))$$

9
$$M_{cov} = [Z_1^l, \ldots \ldots, Z_C^l]$$

Aggregation using Group-Based Averaging

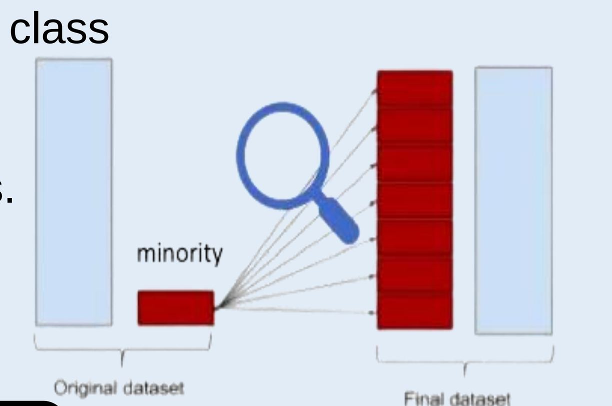
* ✳ Alternative to **Federated Averaging**.
* ✳ **Reduces dissimilarity** in local gradient updates.
* ✳ Server collects model weights and forms equal sized **random groups**.
* ✳ Averaging is done within groups.
* ✳ Clients receive **group's averaged weights**.**(Equation 10)**

10
$$\begin{cases} w_{avg,i}^{t+1} = (w_s^t + \sum_{k\in G_i^t} w_k^t)/(|G_i^t| + 1), \forall i \in \{1 \ldots s\} \\ w_{avg}^{t+1} = (\sum_{i=1}^{S} w_{avg,i}^{t+1}/s) \end{cases}$$



Synthetic Minority Over-Sampling Technique

* ✳ Synthesizes new minority class samples by interpolation.
* ✳ Achieves better results.
* ✳ **Avoids** majority class bias.
* ✳ **Enhances unsupervised training** accuracy with unlabeled data.



REFERENCES

[1] Liu, Q., Yang, H., Dou, Q., & Heng, P. A. (2021, September). Federated semi-supervised medical image classification via inter-client relation matching. In International Conference on Medical Image Computing and Computer-Assisted Intervention (pp. 325-335). Springer, Cham.

[2] RSNA Intracranial Hemorrhage Detection. (2019). Kaggle. https://www.kaggle.com/competitions/rsna-intracranial-hemorrhage-detectio n/data (Date of access: 2024, January 26).

[3] Skin Lesion Images for Melanoma Classification. (2019). Kaggle. https://www.kaggle.com/datasets/andrewmvd/isic-2019. (Date of access: 2024, January 26).

[4] Skin Cancer MNIST: HAM10000. (2018). Kaggle. https://www.kaggle.com/datasets/kmader/skin-cancer-mnist-ham10000 (Date of access: 2024, January 26).

[5] Zhuang Liu, Hanzi Mao, Chao-Yuan Wu, Christoph Feichtenhofer, Trevor Darrell, Saining Xie "A ConvNet for the 2020s" arXiv:2201.03545v2 [cs.CV] 2 Mar 2022.

[6] G. Tian, Z. Wang, C. Wang and J. Chen, "A deep ensemble learning-based automated detection of COVID-19 using lung CT images and Vision Transformer and ConvNeXt," 2022.

[7] H. Wei, "Medium," [Online]. Available: https://medium.com/@haataa/fighting-imbalance-data-set-with-code-examples-f2a3880700a6. [Accessed 21 05 2024]

EXPERIMENTAL RESULTS

HAM10000 Dataset Results

* ✳ Overall accuracy improved by **1.22%** when combining all improvements.
* ✳ Among the different improvements, **SMOTE** yielded the most significant results for all datasets.

Method	HAM10000				
	Accuracy	F1-Score	AUROC	Sensitivity	Specificity
FedAvg (Liu,Yang, Dou, & Heng) Architecture: DensNet121 (10L / 0U)	97.09	96.99	97.42	79.44	97.14
FedAvg (Liu,Yang, Dou, & Heng) Architecture: DensNet121 (2L / 0U)	94.57	94.31	92.11	64.04	95.33
FedIRM (Liu,Yang, Dou, & Heng) Architecture: DensNet121 (Confidence Threshold: 0.3) , (2L / 8U)	94.92	94.64	92.88	60.03	95.82
FedIRM (Liu,Yang, Dou, & Heng) Architecture: DensNet121 (Confidence Threshold: 0.87) , (2L / 8U)	95.20	94.81	93.89	58.49	95.42
FedIRM (Covariance Matrix) Architecture:DensNet121 (Confidence Threshold: 0.87) , (2L / 8U)	95.34	94.99	94.03	62.61	95.18
FedIRM Architecture:Convnext-Tiny (Confidence Threshold: 0.87) , (2L / 8U)	95.34	95.10	95.49	62.40	95.90
FedIRM (Group based averaging) Architecture:Convnext-Tiny (Confidence Threshold: 0.87) , (2L / 8U)	95.55	95.16	95.70	57.66	96.12
FedIRM (Org. Group Based Avg) Architecture:Convnext-Tiny (Confidence Threshold: 0.87) , (2L / 8U)	95.51	95.19	95.26	59.08	96.38
FedIRM (SMOTE) Architecture:Convnext-Tiny (Confidence Threshold: 0.87) , (2L / 8U)	95.88	95.71	96.70	69.19	95.98
FedIRM (Combined) Architecture:Convnext-Tiny (Confidence Threshold: 0.87) , (2L / 8U)	96.14	95.99	95.70	71.34	96.56

RSNA Dataset Results

* ✳ Overall accuracy improved by **1.86%** when combining all improvements.

Method	RSNA				
	Accuracy	F1-Score	AUROC	Sensitivity	Specificity
FedAvg (Liu,Yang, Dou, & Heng) Architecture: DensNet121 (10L / 0U)	91.87	91.56	88.70	62.09	95.00
FedAvg (Liu,Yang, Dou, & Heng) Architecture: DensNet121 (2L / 0U)	89.10	88.89	83.48	57.69	92.99
FedIRM (Liu,Yang, Dou, & Heng) Architecture: DensNet121 (Confidence Threshold: 0.3) , (2L / 8U)	89.63	89.20	87.02	55.92	93.89
FedIRM (Liu,Yang, Dou, & Heng) Architecture: DensNet121 (Confidence Threshold: 0.87) , (2L / 8U)	90.00	89.41	86.46	54.32	93.83
FedIRM (Covariance Matrix) Architecture:DensNet121 (Confidence Threshold: 0.87) , (2L / 8U)	89.81	89.21	86.85	53.84	93.90
FedIRM Architecture:Convnext-Tiny (Confidence Threshold: 0.87) , (2L / 8U)	91.30	90.83	89.91	57.70	94.91
FedIRM (Group based averaging) Architecture:Convnext-Tiny (Confidence Threshold: 0.87) , (2L / 8U)	91.19	90.75	87.33	58.93	94.07
FedIRM (Org. Group Based Avg) Architecture:Convnext-Tiny (Confidence Threshold: 0.87) , (2L / 8U)	91.25	90.87	87.38	59.83	94.23
FedIRM (SMOTE) Architecture:Convnext-Tiny (Confidence Threshold: 0.87) , (2L / 8U)	91.40	90.95	89.81	58.53	94.60
FedIRM (Combined) Architecture:Convnext-Tiny (Confidence Threshold: 0.87) , (2L / 8U)	91.49	91.06	89.60	59.21	94.90

ISIC2019 Dataset Results

* ✳ Overall accuracy improved by **1.43%** when combining all improvements.

Method	ISIC2019				
	Accuracy	F1-Score	AUROC	Sensitivity	Specificity
FedAvg (Liu,Yang, Dou, & Heng) Architecture: DensNet121 (10L / 0U)	96.70	96.61	96.25	77.31	97.60
FedAvg (Liu,Yang, Dou, & Heng) Architecture: DensNet121 (2L / 0U)	93.59	93.38	89.39	52.96	95.74
FedIRM (Liu,Yang, Dou, & Heng) Architecture: DensNet121 (Confidence Threshold: 0.3) , (2L / 8U)	93.52	93.05	87.57	37.82	95.71
FedIRM (Liu,Yang, Dou, & Heng) Architecture: DensNet121 (Confidence Threshold: 0.87) , (2L / 8U)	93.89	93.26	91.44	43.93	96.32
FedIRM (Covariance Matrix) Architecture:DensNet121 (Confidence Threshold: 0.87) , (2L / 8U)	94.09	93.45	91.60	45.26	96.60
FedIRM Architecture:Convnext-Tiny (Confidence Threshold: 0.87) , (2L / 8U)	93.96	93.48	91.95	44.24	97.02
FedIRM (Group based averaging) Architecture:Convnext-Tiny (Confidence Threshold: 0.87) , (2L / 8U)	94.06	93.55	91.46	44.41	96.56
FedIRM (Org. Group Based Avg) Architecture:Convnext-Tiny (Confidence Threshold: 0.87) , (2L / 8U)	94.31	93.94	92.75	50.86	96.76
FedIRM (SMOTE) Architecture:Convnext-Tiny (Confidence Threshold: 0.87) , (2L / 8U)	94.85	94.34	94.10	46.97	97.48
FedIRM (Combined) Architecture:Convnext-Tiny (Confidence Threshold: 0.87) , (2L / 8U)	94.95	94.64	94.10	55.51	96.89

CONCLUSION

* ✳ This project aims to enhance the original algorithm through an **ablation study**.
* ✳ Changes included altering **backbone architecture**, integrating a **covariance matrix**, offering an **alternative aggregation method**, and implementing **oversampling** for balanced class distributions.
* ✳ All these modifications outperformed the original research metrics.
* ✳ In the future, FedIRM can be aligned with **non-independent and non-identically** distributed settings, improving its robustness and performance with real-world datasets.