Intelligent Handling of Noise in Federated Learning with Co-Training for Enhanced Diagnostic Precision

Farah Farid Babar, Faisal Jamil and Faiza Fareed Babar

September 9, 2024

Intelligent Handling of Noise in Federated Learning with Co-training for Enhanced Diagnostic Precision

Farah Farid Babar[1][0009−0009−0073−2045], Faisal Jamil[1][0000−0003−1994−6907], and Faiza Fareed Babar[2][0009−0007−2899−2397]

[1] Department of Computer Science, University of Huddersfield, Queensgate, HD1 3DH Huddersfield, United Kingdom farah.babar@hud.ac.uk;f.jamil@hud.ac.uk;
[2] School of Electronics and Information Engineering, Hebei University of Technology, Tianjin, China faizababar2016@yahoo.com

Abstract. Federated learning (FL) allows multiple distributed clients to train a model while protecting their data. Medical data, especially brain MRIs, might be misdiagnosed due to capture noise and scanner abnormalities. Existing noise-handling technologies use data transmission, raising communication burdens and privacy risks. To address these challenges, we propose a novel Adaptive Sample Weighting Federated Learning (ASW-FL) approach incorporating co-training into the FL framework. The local and global models in FL have different learning abilities, which we use to our advantage. The two models "teach each other" to ignore noisy labels by exchanging samples with their confident predictions. Our method improved accuracy from 83.05% to 85.20% using various aggregation algorithms on a benchmark dataset of 1300 brain MRIs and our own Biobank UK data. Our methodology for accurate, privacy-preserving medical image analysis is adequate. The proposed model is precise but requires more processing resources, making it more appropriate for powerful servers than personal devices.

Keywords: Federated Learning · Co-training · Noise Handling · Diagnostic Precision· Machine Learning Approach.

1 Introduction

The latest advancements in artificial intelligence (AI) and machine learning have ushered in a revolutionary age in cancer imaging and research, which is expected to influence medical practice significantly. These technologies can automate manual tasks associated with clinical image interpretation. Numerous aspects of cancer care, such as identity, prognosis, subtype class, and optimisation, heavily depend on AI [1]. Deep neural networks have undergone extensive training to study digital pathology slides and radiological pictures across many cancers. For instance, AI models now exhibit expert-level overall performance in detecting mammographic lesions [2] [3]. However, the privacy issue has long been afflicted by deep learning, particularly in healthcare [4]. Medical images typically involve patient-sensitive data that must remain private [5]. Federated studying (FL) has emerged as a promising solution to those challenges, permitting model training to co-occur on multiple distributed clients. This method addresses privacy, security, and bandwidth constraints issues, making it attractive for programs in healthcare, finance, and other domains [6–9].

The manual medical annotation process during dataset preparation is prone to errors due to the high level of expertise required for clinical diagnosis. Consequently, label noise often arises, challenging the assurance of label accuracy [10]. In medical datasets, label noise commonly stems from intra- and inter-observer variability, which can confuse supervised training. Label noise has been extensively studied in centralized settings within medical [11], but its impact on federated learning in medical contexts is still unexplored.

To address label noise in medical, federated learning frameworks require powerful denoising techniques. Existing methods often utilize strategies that involve the transmission of overhead information. For instance, [12] suggests a process that exchanges class-wise centroids of local data on every device. Similarly, [13] suggests sending data quality and the amount of training data from each client to the server in each round.

This study addresses the need for robust FL frameworks to manage label noise in medical datasets. This study uses FL with co-training in semi-supervised learning [14]. These are the main contributions:

1. We propose an effective method that easily integrates into existing FL frameworks to handle mislabeled local data. Our method ensures privacy and preserves both privacy and communication efficiency.

2. Our federated co-training scheme comes with a performance guarantee.

3. Extensive evaluations on two datasets have been done. Which validate the superiority of our method over existing approaches, including FedAvg [15], co-teaching [16], and noise-tolerant FL schemes [12].

2 Related work

Federated Learning is a decentralized model [15] which involve four steps in each round. First, the central server sends the global model to all the nodes. Second, the local node updates the model with its local data. Then, the updated models are transmitted back to the central server. Finally, the global model is updated from their aggregates [17]. In medical imaging, FL has demonstrated promise in various applications. [18] integrated differential-privacy techniques into FedAvg for brain MRI segmentation, ensuring patient data confidentiality. Explored FL to detect COVID-19 abnormalities in lung CT scans across multiple hospitals. Extended FL to semi-supervised CT segmentation by leveraging unlabeled data, highlighting its versatility in medical imaging tasks. Additionally, [19] tackled data heterogeneity by aligning server and client prototypes, addressing class imbalance issues. Our work proposes further advancing noise removal FL by integrating Co-Training with federated learning. We aim to enhance the model's robustness in labelling noise.

Noise management Federated Learning Medical imaging relies on deep learning techniques for diagnosis and analysis and suffers from label noise [20]. Classical deep learning models use clean data, which is impossible in medical imaging. Firstly, constructing a noise transition matrix is crucial. Various techniques have been used, such as constrained linear layers or non-linear networks. Forward and backward correction methods have been proposed to rectify outputs and loss [21]. Secondly, the design of noise-tolerant loss functions plays an essential role in robust model training in medicine [22]. Symmetric cross-entropy loss has shown promise in enhancing model robustness [23]. Finally, MentorNet utilizes pre-trained networks to filter out noisy samples [16]. Co-teaching randomly initializes networks with different parameters and exchanges clean samples for mutual learning and updating [24]. In summary, noise management learning methodologies used in medical imaging applications hold great potential for improving diagnostic accuracy. Addressing these challenges gives a promising startup for more robust and interpretable Federated learning models in medical diagnosis. Table 1. Shows an overview of all the previous literature techniques used to handle noise in federated learning.

Table 1: Overview of the existing Noise Tolerance Federated Learning.

Paper	**Techniques**	**Benefit**	**Limitation**
[25]	Noise adaptation layer	Improves model robustness to noise	May not be suitable for all noise types, complex design
[16]	Co-training	Improves model robustness to noise, reduces overfitting	Requires additional training data, sensitive to hyperparameters
[26]	MentorNet	Improves model robustness to noise, reduces overfitting	Requires additional training data, sensitive to hyperparameters
[23]	Symmetric cross-entropy loss	More robust to label noise than traditional cross-entropy	May not be universally applicable
[21]	Loss correction approach	Improves model convergence, reduces noise impact	May not be suitable for all noise types, complex design
[27]	Noise transition matrix	Reduces label noise, improves model accuracy	Requires additional data, computationally expensive
[28]	Enhanced co-training	Improves effectiveness of co-training in noisy settings	Builds on limitations of co-training

3 Methodology

Federated learning (FL) is a decentralized approach to machine learning that facilitates collaborative model training over distributed datasets while upholding data privacy. Meanwhile, co-training is a semi-supervised learning method wherein multiple models are trained on distinct subsets of data. These models then exchange information iteratively to enhance their performance. Our proposed framework integrates federated learning with co-training and adaptive weighting sample methodologies to improve brain tumor diagnosis from MRI scans. This framework addresses noise in federated learning, enhances the identification of brain cancers in MRI images, mitigates the impact of data uncertainties, and ensures privacy. Through this approach, we aim to develop a robust and dependable AI model for identifying brain tumors.

3.1 Proposed Architecture

This section presents an overview of the proposed architecture designed to handle noise in federated learning using co-training to detect brain tumors in MRI images. The proposed system conceptual architecture is presented in Fig.1. The dataset we use in this study comprises MRI images sourced from two repositories; one is the commonly used benchmark Brain Tumor Detection V1.0 dataset, and the other is obtained from the Biobank. The initial phase of the model involves data preprocessing, which includes resizing all the images and using normalization (Z-score) and augmentation. After that, the dataset is divided into three subsets for training, validation, and testing purposes. The convolutional Neural Network (CNN) model trains the dataset. The dataset contains brain MRI images to detect brain tumors. CNNs can effectively learn discriminative features indicative of tumor presence, making them well-suited for our dataset. The proposed architecture combines Federated Learning with co-training and adaptive weighting sample techniques to enhance brain tumor detection from MRI images. Co-training involves training two models on different subsets of the dataset, which exchange information to improve performance. The adaptive weighting sample technique helps in handling noise and unlabeled data. This integrated strategy aims to minimize the influence of data uncertainties and ensure privacy while creating a robust and reliable AI model for brain tumor identification.

3.2 Data Description

The datasets used in this study consist of brain MRI images sourced from two different repositories: the Brain Tumor Detection V1.0 benchmark dataset and data obtained from the UK Biobank. The UK Biobank dataset includes information from 2819+ brain tumor patients, focusing on MRI images used for surgical planning in brain tumor treatment. These images vary in in-plane resolutions of 256 x 256. For a visual representation, a sample of the prepared data is in Fig.2a.

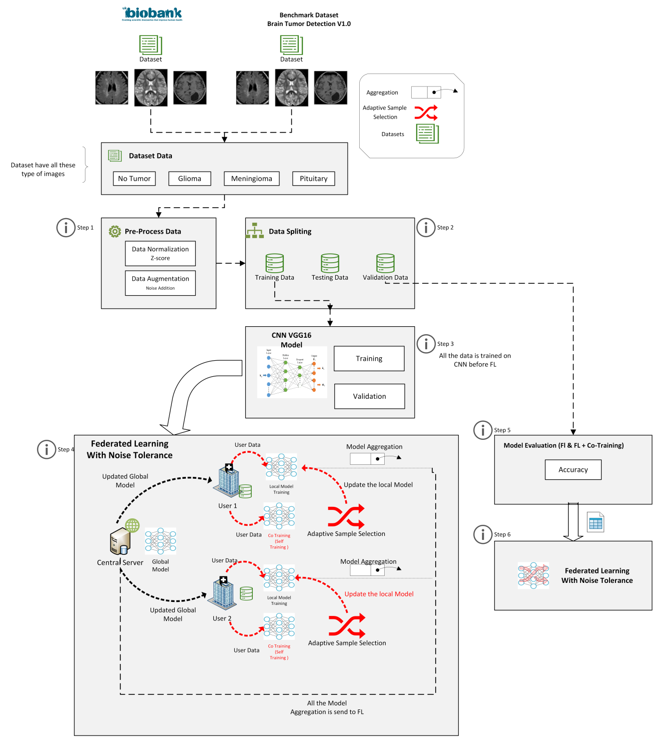


Fig. 1: Architecture of the proposed system

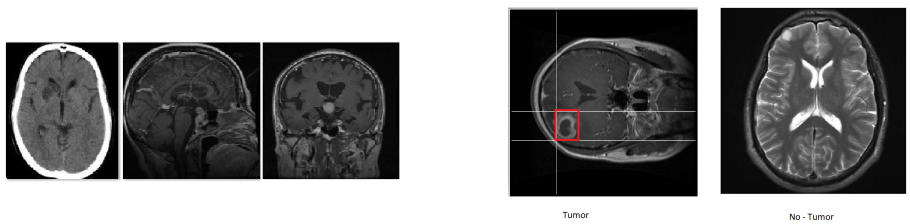
Noise Types The MRI images are susceptible to various types of noise, including acquisition noise, motion artefacts, and image distortions. Acquisition noise arises from imperfections in the imaging process, while motion artefacts occur due to movement during imaging, and image distortions encompass irregularities in the MRI images. These noise types can impact training by distorting image features and introducing inaccuracies.

3.3 Data Pre-processing

Pre-processing data is a machine learning technique that turns unprocessed data into a desirable format. A few variables were applied when processing the MRI data. These elements are listed as follows:

Image Extraction The data sets were organised using 3D Nifty (Neuroimaging Informatics Technology Initiative). The first pre-processing entailed taking 2D-shaped slices out of the 3D images. The "nibabel" library was used to change it into PNG format.

Data Labelling and Segmentation After that, all images were classified as either "Tumor" or "No - Tumor". Sample images illustrating both scenarios in Fig.2b. The dataset comprised a total of 1300 images. Of these, 80% were allocated for training and 20% for validation.



(a) Sample Images of Brain Tumor Dataset

(b) Sample images Tumor and No-tumor

Fig. 2: Brain Tumor Dataset

After that, the normalization process is implemented using the z-score method. This technique standardized the pixel values of the images, resulting in a mean of 0 and a standard deviation of 1. The z-score normalization formula applied can be expressed as:

$$z = \frac{x - \mu}{\sigma} \tag{1}$$

3.4 CNN Architectures

Our proposed system uses the VGG16 CNN model using TensorFlow and Keras libraries. This model is first pre-trained on the sample images dataset, providing a solid foundation for our training. We adjust the CNN architecture by removing fully connected layers and adding max pooling layers to reduce feature map sizes. After flattening the features, we introduced a dense layer with 1024 neurons, followed by a dropout layer to prevent overfitting. ReLU activation was used for the dense layers, softmax for the output, and the Adam optimizer with a learning rate 0.001. We optimized the model through grid search by exploring various combinations of activation functions, loss functions, optimizers, and neuron counts.

3.5 CNN Model Training

The CNN model training process involved several key parameters and techniques to ensure effective learning and convergence. The dataset comprised 1040 MRI slices, with 720 images in the "Tumor" class and 320 in the "No-Tumor" class. To rectify the imbalance in class representation, a greater weight was assigned to each image belonging to the "Tumor" class throughout the training process.

The outcome was a weight of 2.25 assigned to the photos classified as "Tumor" and 1 assigned to the images classified as "No-Tumor. The training includes ten epochs, utilizing a batch size of sixteen and sixty-five step sizes for the total loaded data. Additionally, the model employed optimization techniques such as stochastic gradient descent (SGD) to adjust the model parameters and minimize the loss function during training.

3.6 Data Augmentation

In data augmentation, we use Gaussian and salt-and-pepper noise techniques. These methods add variations to the images(Fig. 3), enhancing the training dataset and the model's ability.

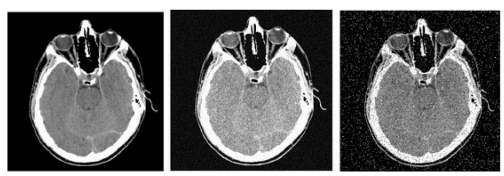


Fig. 3: Data Augmentation by Adding Noise to Images.

3.7 Federated Learning with Noise Tolerance

In our federated co-training framework Fig. 1, each local training round involves the client downloading the global model from the central server and using it as the starting point for training the local model. Unlike disregarding the trained local model from the previous iteration, we integrate knowledge from other clients and the local client's experience using a co-training paradigm. The workflow comprises two primary steps: adaptive sample selection and model co-training. During adaptive sample selection, confident samples, denoted as $S'_{i,l}$ and $S'_{i,g}$, are chosen based on predictions from the two classifiers. In addition, an adaptive weighting mechanism is employed to assign weights to these samples, prioritizing those that are more informative while mitigating the impact of noisy data. These weighted samples are then exchanged between the global and local models, and both models are updated using stochastic gradient descent (SGD). Importantly, the global and local models are independently initialized at the start of each round to ensure their divergence. Algorithm 1 outlines the adaptive sample weight noise management federated learning process, where T represents the number of communication rounds, M represents the number of edge nodes, η denotes the learning rate, E denotes the number of local epochs, and τ denotes the fixed threshold. Each edge node's final global and private models are returned as outputs.

The returned final global model $\Theta^{(T+1)}$ represents the model obtained after T communication rounds. It is important to note that this model does not include any updates or modifications from the $(T+1)$-th round of communication.

Algorithm 1 Adaptive Sample Weighting with Co-Training in Federated Learning

Require: Number of communication rounds T, Number of edge nodes M, Learning rate η, Number of local epochs E, Fixed threshold τ
Ensure: Final global model
for $t = 0$ to T **do**
  Initialize the global model $\Theta^{(t)} = 0$
  for $i = 1$ to M in parallel **do**
    Send the global model to edge i
    $\Theta_i^g, \Theta_i^l, S'_{i,l}, S'_{i,g} \leftarrow$ **LocalUpdate**$(\Theta^{(t)})$
  end for
  Aggregate the local models $\Theta^{(t+1)} = \frac{1}{M}\sum_{i=1}^{M} \Theta_i^g$
  Update sample weights based on the loss from the global model and local models
  Adjust the threshold τ based on the communication round t
end forreturn Final global model $\Theta^{(T+1)}$

Algorithm 2 LocalUpdate

Function LocalUpdate(Θ)
Perform local training using Θ and local data
Identify confident samples ($S'_{i,l}$ and $S'_{i,g}$) based on predictions from Θ and local model
Assign weights to samples in $S'_{i,l}$ and $S'_{i,g}$ based on informativeness and noise mitigation
Perform other local update steps
return Updated local model (Θ_i^g), updated global model (Θ_i^l), confident samples ($S'_{i,l}, S'_{i,g}$)

3.8 Convergence analysis

One drawback of the noisy labels to FL is that they create incorrect decision boundaries among local clients, making it challenging for separate local models to agree upon during the global model's aggregation phase. Because of that, the number of communication rounds between the local clients and the central server has increased. In this section, We analyze the convergence of the Federated Learning (FedAvg), the noise-tolerant FL (Sample selection), and our suggested method, Adaptive sample weight with co-training(ASW-FL).

As shown in Fig.4, our suggested Adaptive sample weight with co-training (ASW-FL) method outperforms both conventional FL and noise-tolerant Federated Learning approaches in terms of convergence speed while maintaining classification accuracy. A far lower number of iterations are needed to get the same precision. For example, with a noise rate of 0.2, the test accuracy of 60% requires about 50 rounds for regular FedAvg, whereas noise-tolerant Federated Learning reaches the same accuracy in about 15 rounds. On the other hand, this accuracy level is reached using our Adaptive sample weight with co-training (ASW-FL) approach in just 8 rounds. This highlights how our method works as a communication-friendly FL framework, lowering communication costs by eliminating the requirement to send overhead messages from local to central components.

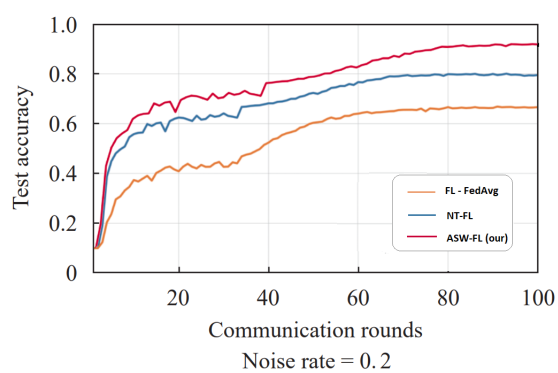


Fig. 4: Comparison of Federated Learning (FedAvg), Noise Tolerance FL (with two classifiers) and our proposed ASW-FL

4 Experiments

This section presents detailed experiment results to verify the robustness and accuracy of our proposed algorithm AWS-FL.

4.1 Experiments setup

We evaluated our suggested method using the Brain Tumor Detection V1.0 benchmark dataset and a real-world dataset from Biobank that comprised Brain MRIs of actual UK cancer patients. In our simulation, each client receives the training samples randomly, and the test set is the only thing the central server keeps around to assess overall performance.

4.2 Computational Resources

The experimentation setup employed a Windows 10 workstation with 32 gigabytes of RAM and an NVIDIA GeForce GTX 1070 GPU (11 GB VRAM). While this configuration allowed for initial exploration, training was performed on an RTX 2080Ti GPU for significantly faster training times. The RTX 2080Ti's parallel processing capabilities and ample VRAM (24 GB) accelerated the training process, especially for handling large datasets and complex models. Compatibility with CUDA and TensorFlow 2.0 ensured seamless integration and optimization.

4.3 Implementation

The PyTorch framework was employed in our implementation. We trained the network in 100 global communication rounds, with 5 local training. We conducted experiments using two versions of our suggested technique with different loss functions. We used the conventional cross-entropy loss (CE) and robust loss function Generalized Cross Entropy (GCE) [22]. In earlier studies, it enhanced model performance when label noise was present.

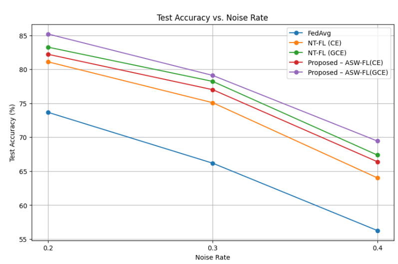
4.4 Evaluation

We evaluated the performance of our proposed Adaptive Sample Weight with Co-training (ASW-FL) approach on two datasets, Brain Tumor Detection V1.0 and Biobank Dataset. We compared ASW-FL against several baseline methods, including FedAvg and Noise Tolerance Federated Learning (NT-FL) with both Cross-Entropy (CE) and Generalized Cross-Entropy (GCE) loss functions. The experiments utilized different noise levels (0.2, 0.3, and 0.4) on the Brain Tumor Detection V1.0 dataset to evaluate the method's noise-handling capabilities.

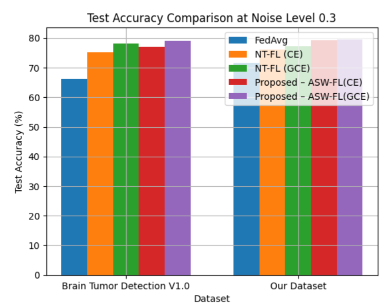
As shown in table2, Our evaluations demonstrated that ASW-FL consistently outperformed other approaches in test accuracy on the Brain Tumor Detection V1.0 dataset. Notably, with a noise level of 0.3, ASW-FL achieved an accuracy of 79.11% compared to 75.09% for NT-FL (CE), the second-best performing method. This improvement highlights ASW-FL's effectiveness in handling noise. The Generalized Cross-Entropy (GCE) loss function enhanced performance, with ASW-FL achieving the highest accuracy of 85.21% at a noise level of 0.2. On our Biobank dataset, which contains real-world noisy data, ASW-FL maintained its superiority. Here, ASW-FL with the GCE loss function achieved the highest accuracy of 79.56%, demonstrating its robustness in practical settings. These results strongly support the efficacy of ASW-FL for improving the accuracy of federated learning in medical image analysis, particularly when dealing with noisy data.

Table 2: Overview of the Test Accuracies (%).

Dataset	**Noise**	**FedAvg**	**NT-FL (CE)**	**NT-FL (GCE)**	**Proposed ASW-FL(CE)**	**Proposed ASW-FL(GCE)**
Brain Tumor Detection V1.0	0.2	73.67	81.11	83.27	82.22	85.21
Brain Tumor Detection V1.0	0.3	66.18	75.09	78.23	77.01	79.11
Brain Tumor Detection V1.0	0.4	56.26	64.02	67.38	66.40	69.45
Our Dataset	0.3	71.63	76.13	77.25	79.32	79.56



(a) Test accuracy (%) on Noise level 0.2, 0.3, and 0.4.



(b) Test accuracy (%) on Brain Tumor Detection V1.0 vs our dataset.

Fig. 5: Test accuracy on both datasets.

5 Discussion and Conclusion

The novel ASW-FL method effectively manages noise in federated learning for medical images, preserving privacy. Integrating CNNs with FL and adaptive weighting mitigates noise without additional data sharing. Our analysis demonstrates ASW-FL's superior noise tolerance compared to existing approaches. However, it requires additional co-training, increasing local user workload. ASW-FL is well-suited for cross-silo federated learning but may not be ideal for cross-device scenarios due to privacy concerns and varying device capabilities. A benefit of ASW-FL is that users retain private local models after training. Future research will explore this feature with label-noise learning to address non-IID data issues.

References

1. Faisal Jamil and Ibrahim A Hameed. Toward intelligent open-ended questions evaluation based on predictive optimization. *Expert Systems with Applications*, page 120640, 2023.
2. William Lotter, Abdul Rahman Diab, Bryan Haslam, Jiye G Kim, Giorgia Grisot, Eric Wu, Kevin Wu, Jorge Onieva Onieva, Yun Boyer, Jerrold L Boxerman, et al. Robust breast cancer detection in mammography and digital breast tomosynthesis using an annotation-efficient deep learning approach. *Nature medicine*, 27(2):244–249, 2021.
3. Faiza Fareed Babar, Shi Lukui, Farah Farid Babar, and Fayaz Muhammad. Enhanced weather forecasting using the meteronet model: A comprehensive ensemble approach. *International Journal of Advanced Multidisciplinary Research*, 10(8):20–38, 2023.
4. Yanbu Wang, Linqing Liu, and Chao Wang. Trends in using deep learning algorithms in biomedical prediction systems. *Frontiers in Neuroscience*, 17:1256351, 2023.
5. Mohammed Adnan, Shivam Kalra, Jesse C Cresswell, Graham W Taylor, and Hamid R Tizhoosh. Federated learning and differential privacy for medical image analysis. *Scientific reports*, 12(1):1953, 2022.
6. Dinh C Nguyen, Quoc-Viet Pham, Pubudu N Pathirana, Ming Ding, Aruna Seneviratne, Zihuai Lin, Octavia Dobre, and Won-Joo Hwang. Federated learning for smart healthcare: A survey. *ACM Computing Surveys (Csur)*, 55(3):1–37, 2022.
7. Faisal Jamil, Shabir Ahmad, Taeg Keun Whangbo, Ammar Muthanna, and Do-Hyeun Kim. Improving blockchain performance in clinical trials using intelligent optimal transaction traffic control mechanism in smart healthcare applications. *Computers & Industrial Engineering*, 170:108327, 2022.
8. Faisal Jamil, Faiza Qayyum, Soha Alhelaly, Farjeel Javed, and Ammar Muthanna. Intelligent microservice based on blockchain for healthcare applications. *Computers, Materials & Continua*, 69(2), 2021.
9. Faisal Jamil and DoHyeun Kim. Enhanced kalman filter algorithm using fuzzy inference for improving position estimation in indoor navigation. *Journal of Intelligent & Fuzzy Systems*, 40(5):8991–9005, 2021.
10. Chuang Zhu, Wenkai Chen, Ting Peng, Ying Wang, and Mulan Jin. Hard sample aware noise robust learning for histopathology image classification. *IEEE transactions on medical imaging*, 41(4):881–894, 2021.
11. Jiarun Liu, Ruirui Li, and Chuan Sun. Co-correcting: noise-tolerant medical image classification via mutual label correction. *IEEE Transactions on Medical Imaging*, 40(12):3580–3592, 2021.
12. Seunghan Yang, Hyoungseob Park, Junyoung Byun, and Changick Kim. Robust federated learning with noisy labels. *IEEE Intelligent Systems*, 37(2):35–43, 2022.
13. Kahou Tam, Li Li, Bo Han, Chengzhong Xu, and Huazhu Fu. Federated noisy client learning. *IEEE Transactions on Neural Networks and Learning Systems*, 2023.
14. Avrim Blum and Tom Mitchell. Combining labeled and unlabeled data with co-training. In *Proceedings of the eleventh annual conference on Computational learning theory*, pages 92–100, 1998.
15. 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*, pages 1273–1282. PMLR, 2017.

16. Bo Han, Quanming Yao, Xingrui Yu, Gang Niu, Miao Xu, Weihua Hu, Ivor Tsang, and Masashi Sugiyama. Co-teaching: Robust training of deep neural networks with extremely noisy labels. *Advances in neural information processing systems*, 31, 2018.
17. Peter Kairouz, H Brendan McMahan, Brendan Avent, Aurélien Bellet, Mehdi Bennis, Arjun Nitin Bhagoji, Kallista Bonawitz, Zachary Charles, Graham Cormode, Rachel Cummings, et al. Advances and open problems in federated learning. *Foundations and trends® in machine learning*, 14(1–2):1–210, 2021.
18. Wenqi Li, Fausto Milletarì, Daguang Xu, Nicola Rieke, Jonny Hancox, Wentao Zhu, Maximilian Baust, Yan Cheng, Sébastien Ourselin, M Jorge Cardoso, et al. Privacy-preserving federated brain tumour segmentation. In *Machine Learning in Medical Imaging: 10th International Workshop, MLMI 2019, Held in Conjunction with MICCAI 2019, Shenzhen, China, October 13, 2019, Proceedings 10*, pages 133–141. Springer, 2019.
19. Zhen Chen, Chen Yang, Meilu Zhu, Zhe Peng, and Yixuan Yuan. Personalized retrogress-resilient federated learning toward imbalanced medical data. *IEEE Transactions on Medical Imaging*, 41(12):3663–3674, 2022.
20. Davood Karimi, Haoran Dou, Simon K Warfield, and Ali Gholipour. Deep learning with noisy labels: Exploring techniques and remedies in medical image analysis. *Medical image analysis*, 65:101759, 2020.
21. Giorgio Patrini, Alessandro Rozza, Aditya Krishna Menon, Richard Nock, and Lizhen Qu. Making deep neural networks robust to label noise: A loss correction approach. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1944–1952, 2017.
22. Zhilu Zhang and Mert Sabuncu. Generalized cross entropy loss for training deep neural networks with noisy labels. *Advances in neural information processing systems*, 31, 2018.
23. Aditya Krishna Menon, Ankit Singh Rawat, Sashank J Reddi, and Sanjiv Kumar. Can gradient clipping mitigate label noise? In *International Conference on Learning Representations*, 2019.
24. Lie Ju, Xin Wang, Lin Wang, Dwarikanath Mahapatra, Xin Zhao, Quan Zhou, Tongliang Liu, and Zongyuan Ge. Improving medical images classification with label noise using dual-uncertainty estimation. *IEEE transactions on medical imaging*, 41(6):1533–1546, 2022.
25. Jacob Goldberger and Ehud Ben-Reuven. Training deep neural-networks using a noise adaptation layer. In *International conference on learning representations*, 2022.
26. Lu Jiang, Zhengyuan Zhou, Thomas Leung, Li-Jia Li, and Li Fei-Fei. Mentornet: Learning data-driven curriculum for very deep neural networks on corrupted labels. In *International conference on machine learning*, pages 2304–2313. PMLR, 2018.
27. Sainbayar Sukhbaatar, Joan Bruna, Manohar Paluri, Lubomir Bourdev, and Rob Fergus. Training convolutional networks with noisy labels. *arXiv preprint arXiv:1406.2080*, 2014.
28. Cheng Xue, Lequan Yu, Pengfei Chen, Qi Dou, and Pheng-Ann Heng. Robust medical image classification from noisy labeled data with global and local representation guided co-training. *IEEE transactions on medical imaging*, 41(6):1371–1382, 2022.