MiFL: Multi-Input Neural Networks in Federated Learning

Bruno Casella [1], Walter Riviera [1], Marco Aldinucci [1], and Gloria Menegaz [1]

[1]Affiliation not available

October 30, 2023

Abstract

Driven by the Deep Learning (DL) revolution, Artificial intelligence (AI) has become a fundamental tool for many Bio-Medical tasks, including AI-assisted diagnosis. These include analysing and classifying images (2D and 3D), where, for some tasks, DL exhibits superhuman performance. Diagnostic imaging, however, is not the only diagnostic tool. Tabular data, such as personal data, vital signs, and genomic/blood tests, are commonly collected for every patient entering a clinical institution. However, it is rarely considered in DL pipelines, although it carries diagnostic information. The training of DL models requires large datasets, so large that every institution might need more data that should be pooled from different sites. Data pooling generates newfound concerns about data access and movement across other institutions spawning multiple dimensions, such as performance, energy efficiency, privacy, criticality, and security. Federated Learning (FL) is a cooperative learning paradigm aiming at addressing these concerns by moving models instead of data across different institutions. This paper proposes a Federated multi-input model that leverages images and tabular data, providing a proof of concept of the feasibility of multi-input FL architectures. The proposed model was evaluated on two showcases: the prognosis of CoViD-19 disease and the patients' stratification for Alzheimer's disease. Results show that enabling multi-input architectures in the FL framework allows for improving the performance regarding both accuracy and generalizability with respect to non-federated models while ensuring security and data protection peculiar to FL.

Posted on 7 Apr 2020 — CC-BY 4.0 — https://doi.org/10.36227/techrxiv.22492732.v1 — e-Prints posted on TechRxiv are preliminary reports that are not peer reviewed. They should not b...

MiFL: Multi-Input Neural Networks in Federated Learning

Bruno Casella[1§], Walter Riviera[*,2,3,§], Marco Aldinucci[1†], and Gloria Menegaz[2†]

[*]Lead contact
[1]University of Turin, Turin, Italy
[2]University of Verona, Verona, Italy
[3]Intel Corp., Italy
[§]Equal contribution
[†]Equal supervision
[†]emails: [bruno.casella, marco.aldinucci]@unito.it,[walter.riviera, gloria.menegaz]@univr.it

Summary

Driven by the Deep Learning (DL) revolution, Artificial Intelligence (AI) has become a fundamental tool for many Bio-Medical tasks, including AI-assisted diagnosis. These include analysing and classifying images (2D and 3D), where, for some tasks, DL exhibits superhuman performance. Diagnostic imaging, however, is not the only diagnostic tool. Tabular data, such as personal data, vital signs, and genomic/blood tests, are commonly collected for every patient entering a clinical institution. However, it is rarely considered in DL pipelines, although it carries diagnostic information.

The training of DL models requires large datasets, so large that every institution might need more data that should be pooled from different sites. Data pooling generates newfound concerns about data access and movement across other institutions spawning multiple dimensions, such as performance, energy efficiency, privacy, criticality, and security. Federated Learning (FL) is a cooperative learning paradigm aiming at addressing these concerns by moving models instead of data across different institutions.

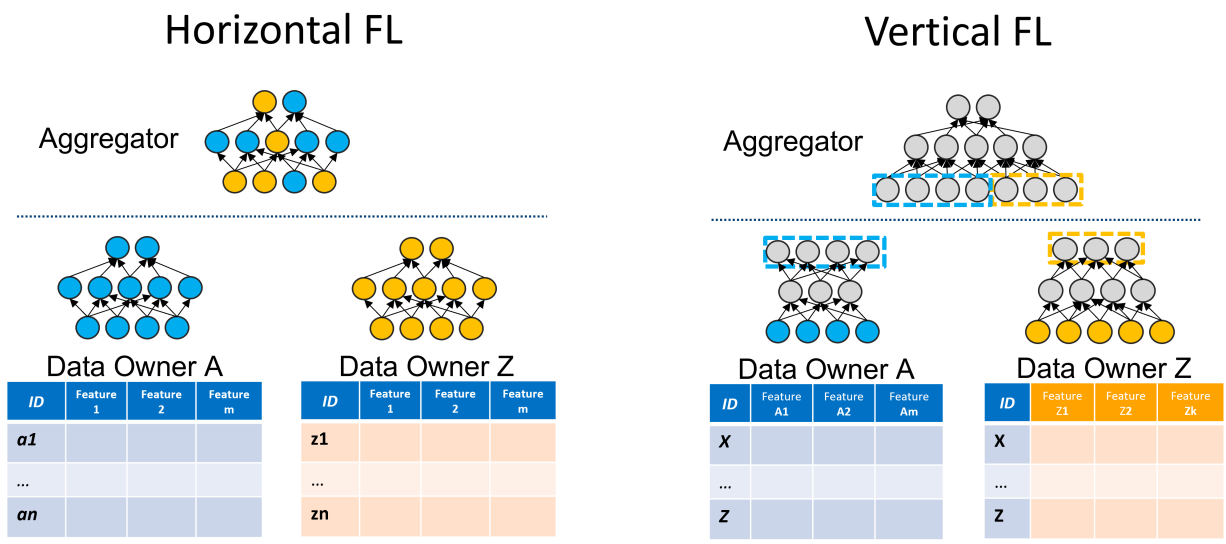
This paper proposes a *Federated multi-input* model that leverages images and tabular data, providing a proof of concept of the feasibility of multi-input FL architectures. The proposed model was evaluated on two showcases: the prognosis of CoViD-19 disease and the patients' stratification for Alzheimer's disease. Results show that enabling multi-input architectures in the FL framework allows for improving the performance regarding both accuracy and generalizability with respect to non-federated models while ensuring security and data protection peculiar to FL.

Keywords

Multi-input classification, Federated Learning, Mixed Data Deep Learning, Federated Classification, *Bio-Medical* Imaging

Introduction

Artificial intelligence techniques, such as Machine Learning (ML) and Deep Learning (DL), are increasingly being studied as tools for tackling challenges in various research fields, starting from the *Bio-Medical* one. One of the strengths of ML models is the capability to capture hidden relationships among complex multi-dimensional data. They have been explored for several tasks, inter-alia, disease classification[1–3], human body segmentation[4–7], the definition of diagnostic



(a) HFL (b) VFL

Figure 1: FL settings: **(a)** HFL, with collaborating institutions sharing the same features space but different samples. In this setting, only one model is being trained locally by each institution that is also shared and aggregated. **(b)** VFL, with collaborating institutions having different features for the same samples. In this case, each C_i has its model that flows into a unified 2nd-level model shared across the federation.

scores[8], drug-discovery[9,10] and data augmentation through the generation of synthetic samples[11–14]. Despite most of these examples leveraging medical images as the preferred data format, a variety of data types are collected by hospitals, clinical laboratories, and other healthcare institutions. Beyond magnetic resonance imaging (MRI), x-rays, or other types of body scans, there are also time series like the ones coming from the electrocardiograms or devices for monitoring vital signs, video from cameras recording patients' movements and positions overnight or during a rehabilitation therapy, and last but not least, text or tabular data coming from surveys, administrative and vital records. Very often, these data are used singularly to solve specific problems.

Although these data are different from each other, they have one common feature: they are critical data for privacy and security and must be treated appropriately by their owners. According to the Data Protection ACT (DPA) published by the Government of the United Kingdom (UK) in 2018 (Part 7, section 205), *health data* are defined as *personal data relating to the physical or mental health of an individual, including the provision of health care services, which reveals information about their health status*. The DPA is just an example of data regulation, but it is enough to demonstrate the efforts that a country might be willing to take to regulate and control how personal data can be used by organizations, businesses, and or the government. The UK is not an isolated example; in fact, DPA was created following the General Data Privacy Regulations (GDPR)[15] in place in the European Union. Other data regulations specifically created for regulating the access and use of Health Data are: the Health Insurance Portability and Accountability Act (HIPAA)[16] in the USA, which provides guidelines and constraints to be respected in order to access data, and the Protection of Personal Information Act (POPIA)[17]. While protecting sensitive information is a critical mission from a governance perspective, from an AI perspective,

introducing all these regulations limits access to data.

Computing paradigms like Federated Learning (FL)[18] can help addressing this challenge. Initially deployed by Google for predicting text input on mobile devices, FL has been adopted by many other industries[19]. In an FL scenario, multiple institutions C_i holding proprietary and critical data collaborate to train a global AI model M. The crucial aspect of the FL is that data belonging to a specific C_i never exits from the IT facility of its owner. Instead of sharing or exchanging data, different institutions iteratively aggregate local models in a single global model M that approximates the one that can be achieved by gathering all data in a single data lake. Typically, the model training algorithm underneath FL is a variant of the Gradient Descent, even if, recently, the FL approach has been generalized to a broader class of ML algorithms, including AdaBoost[20], which can be very effective on tabular data[21].

FL is generally categorised along two main approaches: Horizontal (HFL) and Vertical (VFL) Federated Learning[22].

In both cases, multiple institutions C_i holding proprietary data are willing to train a shared model while never sharing their sensitive information. The training routine can either follow a synchronous or asynchronous schema. The main difference between HFL and VFL resides in the assumption of how data are split. In the HFL, the basic assumption is that each C_i taking part in the federation has the same feature space (i.e., data format, like MRIs) but different sample instances (i.e., different patients). Conversely, in the VFL environment, each C_i contributes to the federation by leveraging different data with different feature spaces, while the sample instances are the same. An example of HFL and VFL settings can be found in Figure **1**.

In this paper, we propose a multi-input model that can handle different types of data, namely images and tabular text, being trained with the orchestration of the HFL environment. We demonstrate the goodness of our approach by solving classification tasks for 2D and 3D images and by showing how the performance of the federated multi-input model can outperform the results obtained by taking the inputs from different data owners in insulation. As datasets, we leveraged the CoViD-19 CXR for the 2D classification and the ADNI dataset for the 3D variant.

This work aims at providing the first proof of concept of a federated multi-input architecture, besides additional evidence that exploiting multi-modal information improves classification performance. Indeed, outperforming the state of the art of the same tasks in a centralised setting is not the scope of this contribution because FL can only produce a model that approximates the one that can be achieved by pooling all the distributed data in a global dataset. In addition, is typically slower in converging and slightly less performing than centralized models[23]. Moreover, it is well known[24,25] that when the datasets are non-i.i.d. distributed, the performance of FL methods degrades. Our multi-input approach reduces the loss in performance when we shift from a centralized paradigm to a federated one, also in cases of non-i.i.d. data.

Inline with open science initiatives, to help reproduce our results and facilitate the customization of the model, we provide the code and instructions on how to use it in Section EXPERIMENTAL PROCEDURES.

Federated Learning in the Bio-Medical field

In recent years, the evolution of machine learning techniques has been dominated by the need to improve the performance of models up to levels that can be used for practical purposes. The introduction of DL has been a real revolution for many tasks, especially for image and text analysis. Unfortunately, DL is data-hungry: larger and larger DL models require more and more data to be trained, and data pooling data is often needed to build an appropriate dataset. Pooling increases the model generalizability[26,27], i.e. the ability of a model to deal with unseen data reliably[28]. Various efforts have been made in the research community in this respect[29–33]. The

studies show how higher exposure to data can increase generalizability, hence the need for real-life federations. FL allows the virtual pooling of different datasets while maintaining the data private and allowing the model to learn features appearing in examples from different data owners.

An example of a HFL setting could be represented by a set of institutions, like hospitals, aiming at training a shared model (i.e., for brain tumor segmentation) by leveraging MRI scans. In this scenario, the assumption is that each C_i would be able to provide MRI scans of their patients. Patients differ from one hospital to another if we consider a worldwide federation with institutions from other countries. In a typical HFL setting, there is only one model, and all the C_i are responsible for ensuring that data gets normalized to feed it. Despite being relatively new, FL has already demonstrated its value in addressing the generalizability challenge[34,35]. Due to the high sensitivity of the data, many other works have been done to explore FL for the *Bio-Medical* field[23,27,36]. While the effectiveness of FL in increasing the model generalizability has already been proven, finding the best way to aggregate the contribution from all the data owners remains an open challenge.

Multi-input classification

The idea of using multiple input sources to perform a classification task is well known in the literature[37]. It can be found with different names like multi-modal[38], multi-view[39], multi-channel[40], or mixed data ML[41]. For the sake of clarity, we are going to refer to this topic using the multi-input label. One of the driving factors beyond the adoption of multi-input AI approaches is data availability. As one can easily imagine, video data lead the list as they consist in tuples of images and audio signals[42,43]. However, the list goes beyond video data, and several multi-modal datasets can be found publicly[44–47]. When it comes to the *Bio-Medical* field, it is widespread to see images being coupled with other types of images like, for example, in the breast tumor classification[48], Alzheimer's disease by using the fluorodeoxyglucose positron emission tomography (FDG-PET) with MRIs[49], unregistered craniocaudal (CC) and mediolateral oblique (MLO) mammograms[50], or the fusion of the multiresolution representation of the same image[51]. In other cases, the fusion involves different data types, like time series[52], or features extracted from different sources[53,54].

Regardless of the type of data being considered, according to Sleeman et al.[38], there are two main possible settings of a multi-input-based approach: co-training and co-regularization. The co-training refers to those problems where the information from one source can help to estimate missing details (like labels) on the other, as it happens in semi-supervised learning[55–57]. Concerning FL, the co-training setting can be easily mapped into the VFL scenario. In the co-regularization case, the various input sources are considered as contributors to build a common descriptor for representing the instances of the problem being addressed. In this last case, the fusion occurs within the classification model itself after each input has been encoded independently[58]. For this work, we relied on the co-training setting to train models in an HFL architecture.

Federated Multi-input

Despite a few works that have addressed the challenge of having a multi-input model employed in an FL setting, the field remains less well-explored in the literature. Most of the articles propose multi-input-based approaches that are directly associated with the VFL setting. Huang et al.[59] describe a scenario where *M views are distributed across M devices*. Same considerations apply for other recent works[60–62].

When it comes to the *Bio-Medical* environment, Che et al.[63] propose a detailed approach that can work in both HFL and VFL settings. However, the main focus is pipeline orchestration and

data leakage prevention. Furthermore, the examples provided refer to preserving the privacy of sequential data, like real-world keyboard data collected from the BiAffect study[64]. Qayyum et al.[62] propose a clustered FL approach where the clinical institutions are organized in two clusters depending on what type of data they own: X-ray data cluster and Ultrasound data cluster, which can be easily related to a standard VFL setting. Mahbub Ul and Rahim[65] present a FL multi-input approach working on Internet-of-Medical-things (IoMT). However, as claimed by the authors, the main focus is to study the heterogeneity of the hardware equipment used to simulate the different clients (medical institutions), and the evaluations are performed on time-series models (like GRU, LSTM and BERT) and Generative Adversarial Networks (GANs). Regarding the tasks, a recent work[66] focuses on the classification of signals coming from Internet-of-Things (IoT) devices, which use autoencoders to extract common representations of the different data sources. Unfortunately, this contribution is not directly linked to the *Bio-Medical* field. Bernecker et al.[67] tackle the liver segmentation problem by proposing a multi-input normalization technique, which tries to encode CT and MRIs into a common representation.

Study relevance

In this work we propose a multi-input model that works in a HFL setting to solve classification tasks. The model is just meant to be a proof of concept and further customization by the community is encouraged. We demonstrate the feasibility of our approach by evaluating the model on two classification tasks: prognosis of CoViD-19 (2D data) disease and Alzheimer's disease relying on the "ADNI" initiative (3D data). While our final goal is not to improve the state of the art performance for the two specific problems, it is important to highlight how the two challenges are open and might benefit from the proposed alternative approach[68]. Alternative routes to solve the classification task have been explored by considering different feature extractions techniques[69,70]. Similar considerations apply for the classification of Alzheimer disease (AD). Many works demonstrate multiple efforts in addressing the challenge by using transfer-learning[71], multiple-kernel[72] and multi-input techniques[73–75].

In summary, the few works available in the literature that mention the implementation of multi-input models in federated settings are directly referring to the VFL setting which was designed for this purpose, non-relatable to the *Bio-Medical* environment, or do not address the classification task. For these reasons, to the best of our knowledge, we believe that this paper is the first in addressing the problem of having a multi-input classification model in an HFL setting for the *Bio-Medical* field.

RESULTS

In this paper, we propose a new approach for solving classification tasks for *Bio-Medical* environment in a privacy-compliant way. The method introduces an HFL setting with the advantage of leveraging multiple input sources. The basic assumption for the proposed approach is that each C_i taking part in the federation has both data types, that is, images and tabular data, locally available and accessible. We demonstrated the goodness of our approach by running several tests based on 2D and 3D images combined with tabular data and by comparing the results obtained by the multi-input model with the only-images and only-tabular models.

Results show that enabling multi-input architectures in the FL framework allows improving the performance regarding both accuracy and generalizability with respect to non-federated models while complying with data protection practices. In particular, we tested our method on two classification tasks: prognosis of CoViD-19 disease from chest X-rays data (COVID-CXR dataset)

and detection of Alzheimer's disease from neuroimaging data (ADNI dataset).

Experimental evaluations

For both datasets, we ran six experiments. More precisely, we evaluated the performance by comparing models using different types of input:

- Image-only, using a ResNet-18[76] model;
- Tabular-only, leveraging an MLP model;
- Multi-input, using a concatenation of the two models as shown in Figure **3**.

In the image-only and multi-input cases, we used the 2D version of ResNet-18 for the experiments based on the CoViD-19 CXR dataset and the 3D version for experiments on the ADNI dataset. The list of experiments was performed in three settings:

- **Isolated**: AI models were trained assuming no collaboration among institutions and tested on the data belonging to the same organization. In this scenario, we have three models (one for each type of input) for each of the six hospitals of the CoViD-19 dataset and each ADNI (ADNI1, ADNI2, and ADNI3) study. Results are shown in Table 1. Moreover, *isolated* models have also been tested on the data of other organizations in order to discuss generalizability properties. Results are shown in Figure 2.
- **Centralized**: assuming collaboration among institutions, AI models were also trained on all the data hosted in a single machine. Results are shown in Table 2.
- **Federated**: here, the collaboration was achieved by FL, which guarantees a high level of privacy. Results are shown in Table 2.

Table 1: **Accuracy in the isolated setting.** Results (mean $\pm$ standard deviation) obtained with 5-fold cross-validation.

	CoViD-19 Hospital					
Input	**A**	**B**	**C**	**D**	**E**	**F**
Only images	0.833 ± 0.06	0.605 ± 0.14	0.640 ± 0.14	0.700 ± 0.07	0.674 ± 0.08	0.538 ± 0.03
Only tabular	0.883 ± 0.05	0.780 ± 0.06	0.723 ± 0.09	$\mathbf{0.775 \pm 0.01}$	$\mathbf{0.933 \pm 0.05}$	0.542 ± 0.12
Multi-input	$\mathbf{0.909 \pm 0.09}$	$\mathbf{0.857 \pm 0.09}$	$\mathbf{0.840 \pm 0.07}$	0.763 ± 0.04	0.890 ± 0.10	$\mathbf{0.777 \pm 0.02}$

Input	**ADNI1**	**ADNI2**	**ADNI3**
Only images	0.510 ± 0.04	$\mathbf{0.644 \pm 0.05}$	0.813 ± 0.01
Only tabular	0.725 ± 0.00	0.615 ± 0.03	$\mathbf{0.881 \pm 0.03}$
Multi-input	$\mathbf{0.750 \pm 0.07}$	0.641 ± 0.02	0.779 ± 0.02

Table 2: **Accuracy in centralized and federated setting.** Results (mean $\pm$ standard deviation) obtained with 5-fold cross-validation.

	CoViD-19 CXR		ADNI	
Input	**Centralized**	**Federated**	**Centralized**	**Federated**
Only images	0.596 ± 0.06	0.565 ± 0.01	0.664 ± 0.02	0.653 ± 0.06
Only tabular	$\mathbf{0.775 \pm 0.01}$	0.719 ± 0.02	0.640 ± 0.02	0.707 ± 0.01
Multi-input	0.733 ± 0.01	$\mathbf{0.731 \pm 0.02}$	$\mathbf{0.789 \pm 0.01}$	0.728 ± 0.02

The datasets were split into three subsets for the isolated and centralized experiments: train, validation, and test using 80%, 10%, and 10% quotas, respectively. We kept the test set fixed and used the train and validation sets to train the models on five stratified folds in each centralized experiment. The reported accuracy values were obtained by evaluating the best model with the best validation accuracy on the test set.

Datasets were partitioned only in train and test sets for the federated experiments. This is mainly due to two reasons: 1) since each C_i would perform only one iteration before sending the model back to the aggregator, running a cross-validation step would not bring any benefits as there would not be any iterative process to optimize. 2) the tool used for running the federated experiments would not allow a five-fold cross-validation phase without massive intervention to the low-level code. To have comparable results, we preserved the percentage of train data by splitting the dataset into train and test sets, respectively, using 80% and 20%. We ran the federated experiments five times to ensure that the performance would not depend on a specific data split.

Discussion

We first evaluated the performance (in terms of classification accuracy) of the multi-input models in the isolated setting and compared it to the baselines taking as input only images and only tabular, respectively. The results, reported in Table 1, show that combining images and tabular as input data appear to be the most promising approach. Multi-input models outperform baseline models in four out of six cases for the CoViD-19 dataset and in one out of three ADNI studies. In cases where the baseline is better than the multi-input model, the loss of accuracy is minimal.

We compared the isolated settings results against the centralized and federated results, reported in Table 2. Isolated strategies are preferable when tested on their local data, but they suffer from low generalizability. To better investigate this aspect, we performed a cross-test evaluation by testing the best model of each isolated setting against a small testing dataset belonging to the other institutions.

More in detail, we evaluated each model trained on the data of a specific CoViD-19 hospital (or ADNI partition) on a testing set coming from each of the other CoViD-19 hospitals (or ADNI partitions). The results, reported in Figure 2, show that *isolated* models do not generalize well on new data (from other institutions) compared to the accuracy obtained when tested on their proprietary data. Different considerations apply to centralized strategies. This setting generally performs slightly worse than isolated training because of the non-i.i.d. nature of the data. However, it can generalize better because training uses a richer and more diverse dataset. Moreover, in some scenarios, such as the hospitals D and F results of the CoViD-19 dataset and the ADNI1 and ADNI2 studies, the centralized approach outperforms or is on par with the isolated training.

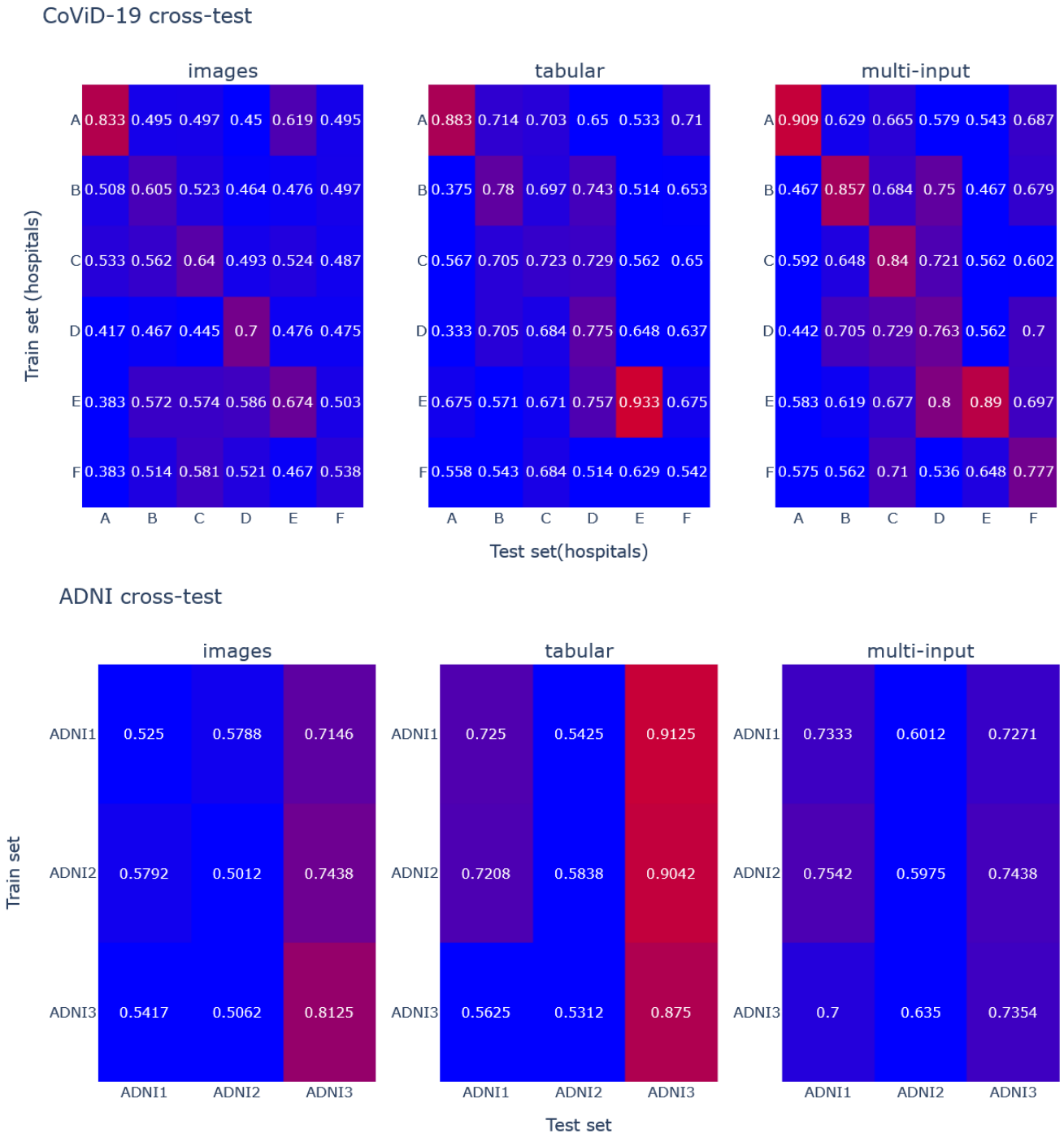


Figure 2: Accuracies obtained by testing the *isolated* models on data belonging to other institutions. Highlighted in red are the values above the federated performances reported in table 2

In real-life scenarios, however, a centralized setting is unrealistic due to data regulations preventing medical institutions from sharing data. For this reason, it is fundamental that FL performance remains aligned with centralized performance. Sheller et al.[23] demonstrated that FL approaches are a great option when we want to increase the ability of a model to generalize by leveraging multiple data sets, but when compared to a centralized setting leveraging the same data sets is just slower in converging and slightly less performing. This behaviour is confirmed for most of the results reported in Table 2. The proposed FL approach using multi-input NNs outperforms the federated baselines and remains aligned with the centralized counterpart. Furthermore, we argue that a minimal loss of accuracy is more than compensated for by the privacy preservation

inherent in FL, especially in industries and institutions in the *Bio-Medical* field.

EXPERIMENTAL PROCEDURES

Lead contact

Request for information and resources used in this article should be addressed to Walter Riviera (walter.riviera@univr.it).

Resource availability

The code used for experimental evaluation is publicly available at https://github.com/alpha-unito/Multi-Input-Neural-Networks-in-Federated-Learning.

All the experiments for CoViD-19 prognosis were performed on the HPC4AI[77] facility located in Turin (node: 8 cores per CPU, AMD EPYC-IPBP, 1 NVIDIA A40 GPU). For FL experiments, we adopted Open Federated Learning (OpenFL)[78], the new framework for FL developed by the Intel Internet of Things Group (IOTG) and Intel Labs. FL experiments were executed on a distributed environment encompassing six Collaborators (clients in the federation that train a global model on a local dataset) and one Aggregator (that aggregates the model updates received from Collaborators), each of which ran on the previously described node.

For the ADNI case study, all the experiments were performed on a 4 nodes cluster of dual socket machines equipped with Intel(R) Xeon(R) Platinum 8380 CPU @ 2.30GHz, with 40 physical cores per socket.

Method

To demonstrate our point, we evaluated our approach on two different datasets: the COVID-CXR dataset (2D) and ADNI dataset (3D). Both Datasets are publicly available.

This section illustrates the adopted architecture and describes the data pre-processing phases for both datasets. Finally, we share the experimental results.

Architecture

Figure 3 displays the architecture of a multi-input NN. The general idea is to aggregate two different NNs trained on the same dataset but using different data types. In particular, the ultimate goal is to aggregate a Convolutional Neural Network (CNN)[79] and a Multi Layer Perceptron (MLP)[80], respectively trained using as input features a set of images and a tabular data frame. CNN and MLP are used as feature extractors. We used a ResNet-18[76] with ten output features for the CNN and a MLP consisting of an input layer, three hidden layers (with 50, 50 and 10 neurons, respectively) and one output layer. We concatenated the outputs of the CNN and the MLP, consisting of 10 features each which resulted in a layer with a dimension of 20. Finally, a ReLU activation function was followed by a linear layer for classification.

Evaluation metric: for each experiment, we returned the test accuracy value defined as the ratio between correct guesses among all guesses, more precisely:

$$Accuracy = (TP + TN)/(TP + TN + FP + FN) \quad (1)$$

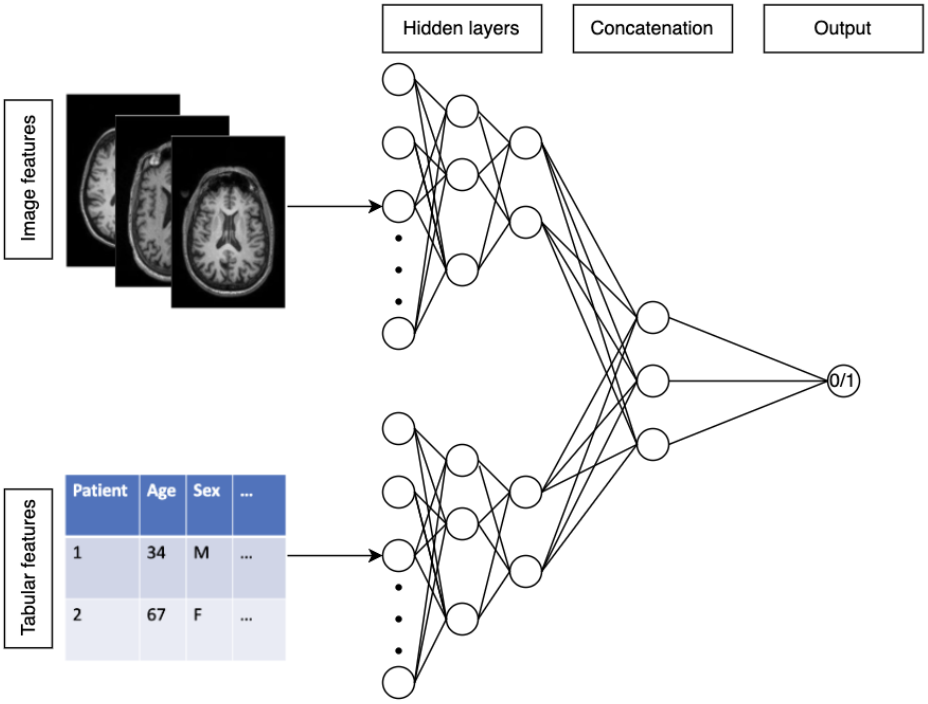


Figure 3: A general Multi-Input Neural Network Architecture

Model: models were trained by minimizing the binary cross-entropy loss with mini-batch gradient descent using the Adam optimizer with learning rate $1e{-}4$ and OneCycleLR as scheduler. The local batch size was 8. The number of training epochs and FL rounds on the CoViD-19 classification task was set to 500, while on the Alzheimer's detection task was set to 200.

Datasets

We tested the multi-input NN on two tasks:

1. Prognosis of CoViD-19 disease from chest X-rays data (CXR), using the COVID-CXR dataset
2. Detection of Alzheimer's disease from neuroimaging data, using the ADNI dataset.

The federated setting emulates a realistic medical non-i.i.d. scenario, where each C_i is hosted on an independent computing node using its dataset, contrasting with standard procedures where non-i.i.d distributions are often simulated by splitting a single source dataset hosted in a single machine.

CoViD-19 dataset

This task relied on real-world data consisting of CXR and clinical parameters, divided into train and test sets. Data were collected from six hospitals in emergency conditions during the first outbreak in Northern Italy in collaboration with Centro Diagnostico Italiano and Bracco Imaging. Due to the different data collection procedures, the distribution of image features varies between each hospital, leading to the well-known problem of non-iidness[24,25]. The CoViD-19

CXR dataset consists of 1589 patients. Each of them is provided with a CXR and some clinical parameters (namely age, sex, positivity at admission, temperature, days of fever, cough, difficulty in breathing, WBC, RBC, CRP, glucose, LDH, INR, PaO2, PaCO2, pH, high blood pressure, diabetes, dementia, BPCO, cancer, CKD and respiratory failure). The dataset details are summarized in Table 3. Additional information about this dataset can be found at `https://aiforcovid.radiomica.it/`.

Table 3: Statistics of the CoViD-19 CXR dataset

Hospital	Samples	Positives	Negatives
A	120	85	35
B	104	59	45
C	151	81	70
D	139	76	63
E	101	55	46
F	974	546	428

This dataset also exhibits a clear quantity skew distribution because, as shown in Table 3, more than 60% of the data is contained in hospital F. However, recent works[24,25] in FL literature show that quantity skew does not degrade the model's performance because most FL algorithms, such as FedAvg[18], adopt a weighted averaging of the parameters. As a result, the distribution of samples (except for the quantity) is uniform among parties, which is the easiest setting. All the images, provided in JPEG format, were rescaled to 256x256. As for data augmentation, we performed random horizontal flips and random crops with a probability of 50%.

ADNI dataset

The Alzheimer's Disease Neuroimaging Initiative (ADNI) represents an ongoing, longitudinal, and multicenter study, the main landmark repository currently available for AD. Beginning in October 2004, ADNI has the primary goal of defining outcome measures to be used in clinical trials for assessing the treatment effectiveness in AD patients. However, its scope has been further widened over the years, pointing to identifying early-diagnosis biomarkers in the pre-dementia stage. A comprehensive set of clinical, neuropsychological, neuroimaging (MRI and positron emission tomography), genetic, and biochemical data are currently collected in large cohorts of healthy elderly subjects, mild cognitive impairment, and AD patients. In particular, this study has been organized into different subsequent phases, the main ones being ADNI-1 (2004-2011), ADNI-2 (2011-2016), and ADNI-3 (2016-ongoing), each of which has witnessed the enrollment of a significant number of new subjects over time and the progressive expansion of the adopted technologies and collected data[81,82]. Up-to-date information is available at `https://adni.loni.usc.edu`.

In our study, the 3D T1-weighted MRI scans acquired at baseline were considered as imaging data and downloaded for all the healthy controls (CN) and AD patients available in ADNI-1, ADNI-2, and ADNI-3. Details about the acquisition protocols in terms of scanners, sequences and corresponding parameters can be found at `https://adni.loni.usc.edu/methods/documents/mri-protocols/`. Moreover, age, gender, and APOE4 (ε4 allele of Apolipoprotein E) were retained as additional tabular features to feed the models, according to the coded information in the updated "ADNIMERGE.csv" file. The latter, in particular, represents the most decisive known genetic risk factor for AD and assumes either 0, 1, or 2 according to the number of ε4 alleles

of the APOE gene. Subjects with incomplete data were removed, leading to the final samples reported in Table 4.

ADNI	Samples	AD	CN	Gender		Age		APOE4		
				F	M	avg	std.	type 0	type 1	type 2
1	411	184	227	198	213	75.58	6.21	229	143	39
2	288	143	145	130	158	73.69	7.35	149	107	32
3	262	51	211	136	126	72.01	6.44	169	75	18

Table 4: Main demographic and clinical data for the three ADNI study cohorts. Age is reported as mean $\pm$ standard deviation values, gender as number males/females, while APOE4 refers to the number of $\varepsilon 4$ alleles (0, 1, or 2, respectively).

ADNI Pre-processing: The individual 3D T1-weighted volumes were minimally pre-processed, including reorientation, bias-field correction and non-linear registration to the MNI152-2mm standard space with dimensions of 91x109x91 (*fsl_anat* tool[83]).

Limitations and future work

As mentioned above, this paper's main goal was to demonstrate the feasibility of a horizontal federated multi-input architecture suitable for the biomedical field. Consequently, optimizing the performance in the non-federated conditions was not targeted, and improvements concerning state-of-the-art in this respect could not be demonstrated. However, making a federated architecture available enables the exploitation of multiple sources of unshared data that allows building on top of current cutting-edge single-institution solutions overcoming the low data numerosity issue while improving the generalization capability of the overall system.

The proposed approach does not consider the problem of missing views, which also affects clinical data processing. However, we are confident that the openness and flexibility of the proposed approach will foster research in the field, marking a step in data sharing and distributed processing.

Among the main future directions are

- Investigating the integration of different input sources;
- Exploring different ways of concatenating the different input sources;
- Understanding how different aggregation functions might impact the federated results;
- Proposing a VFL setting where each client has a different type of dataset (images, tabular, text) and DL model.

Acknowledgments

Research conducted on ADNI dataset was made possible thanks to the pre-processing and the consultancy of Ilaria Boscolo Galazzo, with the University of Verona. We also thank Diego Sona, with FBK, who helped in retrieving the CoViD-19 CXR dataset that he also contributed to assemble during the early CoVID-19 pandemia years to stimulate the grow of AI-based diagnostic methods.

Author contributions

Bruno Casella proposed the method, designed and conducted the experiments.
Walter Riviera conducted the experiments and wrote the paper.
Marco Aldinucci and Gloria Menegaz supervised the experiments and wrote the paper.
Equal contribution of Bruno Casella and Walter Riviera.
Equal supervision of Marco Aldinucci and Gloria Menegaz.

References

1. Al-Issa, Y., and Alqudah, A. M. (2022). A lightweight hybrid deep learning system for cardiac valvular disease classification. Scientific Reports *12*, 1–20.

2. Mansour, R. F., Alfar, N. M., Abdel-Khalek, S., Abdelhaq, M. S., Saeed, R. A., and Alsaqour, R. A. (2022). Optimal deep learning based fusion model for biomedical image classification. Expert Syst. J. Knowl. Eng. *39*. doi:10.1111/exsy.12764.

3. Lai, X., Zhou, J., Wessely, A., Heppt, M., Maier, A., Berking, C., Vera, J., and Zhang, L. (2022). A disease network-based deep learning approach for characterizing melanoma. International Journal of Cancer *150*, 1029–1044.

4. Song, Y., Ren, S., Lu, Y., Fu, X., and Wong, K. K. L. (2022). Deep learning-based automatic segmentation of images in cardiac radiography: A promising challenge. Comput. Methods Programs Biomed. *220*, 106821. doi:10.1016/j.cmpb.2022.106821.

5. Weston, A. D., Korfiatis, P., Kline, T. L., Philbrick, K. A., Kostandy, P., Sakinis, T., Sugimoto, M., Takahashi, N., and Erickson, B. J. (2019). Automated abdominal segmentation of ct scans for body composition analysis using deep learning. Radiology *290*, 669–679.

6. Bhatt, P., Sahoo, A. K., Chattopadhyay, S., and Bhatt, C. A study on deep learning models for medical image segmentation. In: *Artificial Intelligence in Industrial Applications* (161–174). Springer (2022):(161–174).

7. Liu, X., Song, L., Liu, S., and Zhang, Y. (2021). A review of deep-learning-based medical image segmentation methods. Sustainability *13*, 1224.

8. D'Ascenzo, F., De Filippo, O., Gallone, G., Mittone, G. et al. (2021). Machine learning-based prediction of adverse events following an acute coronary syndrome (PRAISE): a modelling study of pooled datasets. The Lancet *397*, 199–207. doi:10.1016/S0140-6736(20)32519-8.

9. Gawehn, E., Hiss, J. A., and Schneider, G. (2016). Deep learning in drug discovery. Molecular informatics *35*, 3–14.

10. Zhang, L., Tan, J., Han, D., and Zhu, H. (2017). From machine learning to deep learning: progress in machine intelligence for rational drug discovery. Drug discovery today *22*, 1680–1685.

11. Frid-Adar, M., Diamant, I., Klang, E., Amitai, M., Goldberger, J., and Greenspan, H. (2018). Gan-based synthetic medical image augmentation for increased CNN performance in liver lesion classification. Neurocomputing *321*, 321–331. doi:10.1016/j.neucom.2018.09.013.

12. Zhao, A., Balakrishnan, G., Durand, F., Guttag, J. V., and Dalca, A. V. Data augmentation using learned transformations for one-shot medical image segmentation. In: *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, USA, June 16-20, 2019*. Computer Vision Foundation / IEEE (2019):(8543–8553). doi:10.1109/CVPR.2019.00874.

13. Sedik, A., Iliyasu, A. M., Abd El-Rahiem, B., Abdel Samea, M. E., Abdel-Raheem, A., Hammad, M., Peng, J., Abd El-Samie, F. E., and Abd El-Latif, A. A. (2020). Deploying machine and deep learning models for efficient data-augmented detection of covid-19 infections. Viruses *12*, 769.

14. Cruciani, F., Brusini, L., Zucchelli, M., Pinheiro, G. R., Setti, F., Galazzo, I. B., Deriche, R., Rittner, L., Calabrese, M., and Menegaz, G. (2021). Interpretable deep learning as a means for decrypting disease signature in multiple sclerosis. Journal of Neural Engineering *18*, 0460a6. doi:10.1088/1741-2552/ac0f4b.

15. Hjerppe, K., Ruohonen, J., and Leppänen, V. The general data protection regulation: Requirements, architectures, and constraints. In: Damian, D. E., Perini, A., and Lee, S., eds. *27th IEEE International Requirements Engineering Conference, RE 2019, Jeju Island, Korea (South), September 23-27, 2019*. IEEE (2019):(265–275). doi:10.1109/RE.2019.00036.

16. Sello, P., Bagula, A. B., and Ajayi, O. Laws and regulations on big data management: The case of south africa. In: Zitouni, R., Agueh, M., Houngue, P., and Soude, H., eds. *e-Infrastructure and e-Services for Developing Countries - 11th EAI International Conference, AFRICOMM 2019, Porto-Novo, Benin, December 3-4, 2019, Proceedings* vol. 311 of *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*. Springer (2019):(169–179). doi:10.1007/978-3-030-41593-8_12.

17. Rosenbloom, S. T., Smith, J. R. L., Bowen, R., Burns, J., and other (2019). Updating HIPAA for the electronic medical record era. J. Am. Medical Informatics Assoc. *26*, 1115–1119. doi:10.1093/jamia/ocz090.

18. McMahan, B., Moore, E., Ramage, D., Hampson, S., and y Arcas, B. A. Communication-efficient learning of deep networks from decentralized data. In: Singh, A., and Zhu, X. J., eds. *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, AISTATS 2017, 20-22 April 2017, Fort Lauderdale, FL, USA* vol. 54 of *Proceedings of Machine Learning Research*. PMLR (2017):(1273–1282).

19. Li, L., Fan, Y., Tse, M., and Lin, K.-Y. (2020). A review of applications in federated learning. Computers & Industrial Engineering *149*, 106854. doi:10.1016/j.cie.2020.106854.

20. Polato, M., Esposito, R., and Aldinucci, M. Boosting the federation: Cross-silo federated learning without gradient descent. In: *Intl. Joint Conference on Neural Networks (IJCNN)*. Padua, Italy: IEEE ():(1–10). doi:10.1109/IJCNN55064.2022.9892284.

21. Arfat, Y., Mittone, G., Colonnelli, I., D'Ascenzo, F., Esposito, R., and Aldinucci, M. Pooling critical datasets with federated learning. In: *Proc. of 31st Euromicro Intl. Conference on Parallel Distributed and network-based Processing (PDP)*. Napoli, Italy: IEEE (2023):.

22. Yang, Q., Liu, Y., Cheng, Y., Kang, Y., Chen, T., and Yu, H. Federated Learning. Synthesis Lectures on Artificial Intelligence and Machine Learning Morgan & Claypool Publishers (2019). doi:10.2200/S00960ED2V01Y201910AIM043.

23. Sheller, M. J., Edwards, B., Reina, G. A., Martin, J., Pati, S., Kotrotsou, A., Milchenko, M., Xu, W., Marcus, D., Colen, R. R. et al. (2020). Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data. Scientific reports *10*, 1–12.

24. Casella, B., Esposito, R., Cavazzoni, C., and Aldinucci, M. Benchmarking fedavg and fedcurv for image classification tasks. In: *Proc. of th 1st Italian Conference on Big Data and Data Science, ITADATA22, Milan, September 20-21, 2022* (2022):.

25. Li, Q., Diao, Y., Chen, Q., and He, B. Federated learning on non-iid data silos: An experimental study. In: *38th IEEE International Conference on Data Engineering, ICDE 2022, Kuala Lumpur, Malaysia, May 9-12, 2022*. IEEE (2022):(965–978). doi:10.1109/ICDE53745.2022.00077.

26. Fu, X., Zhang, B., Dong, Y., Chen, C., and Li, J. (2022). Federated graph machine learning: A survey of concepts, techniques, and applications. SIGKDD Explor. *24*, 32–47. doi:10.1145/3575637.3575644.

27. Joshi, M., Pal, A., and Sankarasubbu, M. (2022). Federated learning for healthcare domain - pipeline, applications and challenges. ACM Trans. Comput. Heal. *3*, 40:1–40:36. doi:10.1145/3533708.

28. Mårtensson, G., Ferreira, D., Granberg, T., Cavallin, L., Oppedal, K., Padovani, A., Rektorová, I., Bonanni, L., Pardini, M., Kramberger, M. G., Taylor, J., Hort, J., Snædal, J., Kulisevsky, J., Blanc, F., Antonini, A., Mecocci, P., Vellas, B., and Westman, E. (2020). The reliability of a deep learning model in clinical out-of-distribution MRI data: A multicohort study. Medical Image Anal. *66*, 101714. doi:10.1016/j.media.2020.101714.

29. Schmidt, A. M., Desai, A. D., Watkins, L. E., Crowder, H. A., Black, M. S., Mazzoli, V., Rubin, E. B., Lu, Q., MacKay, J. W., Boutin, R. D. et al. (2022). Generalizability of deep learning segmentation algorithms for automated assessment of cartilage morphology and mri relaxometry. Journal of Magnetic Resonance Imaging. doi:10.1002/jmri.28365.

30. Liang, X., Nguyen, D., and Jiang, S. B. (2021). Generalizability issues with deep learning models in medicine and their potential solutions: illustrated with cone-beam computed tomography (CBCT) to computed tomography (CT) image conversion. Mach. Learn. Sci. Technol. *2*, 15007. doi:10.1088/2632-2153/abb214.

31. Lai, C., Zhou, S., and Trayanova, N. A. (2021). Optimal ecg-lead selection increases generalizability of deep learning on ecg abnormality classification. Philosophical Transactions of the Royal society A *379*, 20200258. doi:10.1098/rsta.2020.0258.

32. Nguyen, D., Kay, F., Tan, J., Yan, Y., Ng, Y. S., Iyengar, P., Peshock, R., and Jiang, S. B. (2021). Deep learning-based COVID-19 pneumonia classification using chest CT images: Model generalizability. Frontiers Artif. Intell. *4*, 694875. doi:10.3389/frai.2021.694875.

33. Cruciani, F., Altmann, A., Lorenzi, M., Menegaz, G., and Galazzo, I. B. What PLS can still do for imaging genetics in alzheimer's disease. In: *IEEE-EMBS International Conference on Biomedical and Health Informatics, BHI 2022, Ioannina, Greece, September 27-30, 2022*. IEEE (2022):(1–4). doi:10.1109/BHI56158.2022.9926813.

34. Venkateswaran, P., Isahagian, V., Muthusamy, V., and Venkatasubramanian, N. (2022). Fedgen: Generalizable federated learning. CoRR *abs/2211.01914*. doi:10.48550/arXiv.2211.01914. arXiv:2211.01914.

35. Sarma, K. V., Harmon, S. A., Sanford, T., Roth, H. R., Xu, Z., Tetreault, J., Xu, D., Flores, M. G., Raman, A. G., Kulkarni, R., Wood, B. J., Choyke, P. L., Priester, A., Marks, L. S., Raman, S. S., Enzmann, D., Turkbey, B., Speier, W., and Arnold, C. W. (2021). Federated learning improves site performance in multicenter deep learning without data sharing. J. Am. Medical Informatics Assoc. *28*, 1259–1264. doi:10.1093/jamia/ocaa341.

36. Silva R., S. S., Gutman, B. A., Romero, E., Thompson, P. M., Altmann, A., and Lorenzi, M. Federated learning in distributed medical databases: Meta-analysis of large-scale subcortical brain data. In: *16th IEEE International Symposium on Biomedical Imaging, ISBI 2019, Venice, Italy, April 8-11, 2019*. IEEE (2019):(270–274). doi:10.1109/ISBI.2019.8759317.

37. Poria, S., Cambria, E., Bajpai, R., and Hussain, A. (2017). A review of affective computing: From unimodal analysis to multimodal fusion. Inf. Fusion *37*, 98–125. doi:10.1016/j.inffus.2017.02.003.

38. Sleeman, W. C., Kapoor, R., and Ghosh, P. (2023). Multimodal classification: Current landscape, taxonomy and future directions. ACM Comput. Surv. *55*, 150:1–150:31. doi:10.1145/3543848.

39. Yan, X., Hu, S., Mao, Y., Ye, Y., and Yu, H. (2021). Deep multi-view learning methods: A review. Neurocomputing *448*, 106–129. doi:10.1016/j.neucom.2021.03.090.

40. Pan, L., Ji, B., Wang, H., Wang, L., Liu, M., Chongcheawchamnan, M., and Peng, S. (2022). MFDNN: multi-channel feature deep neural network algorithm to identify COVID19 chest x-ray images. Health Inf. Sci. Syst. *10*, 4. doi:10.1007/s13755-022-00174-y.

41. Spairani, E., Daniele, B., Signorini, M. G., and Magenes, G. (2022). A deep learning mixed-data type approach for the classification of fhr signals. Frontiers in Bioengineering and Biotechnology *10*.

42. Lee, J., and Ebrahimi, T. Two-level bimodal association for audio-visual speech recognition. In: Blanc-Talon, J., Philips, W., Popescu, D. C., and Scheunders, P., eds. *Advanced Concepts for Intelligent Vision Systems, 11th International Conference, ACIVS 2009, Bordeaux, France, September 28 - October 2, 2009. Proceedings* vol. 5807 of *Lecture Notes in Computer Science*. Springer (2009):(133–144). doi:10.1007/978-3-642-04697-1_13.

43. Thomee, B., Shamma, D. A., Friedland, G., Elizalde, B., Ni, K., Poland, D., Borth, D., and Li, L. (2016). YFCC100M: the new data in multimedia research. Commun. ACM *59*, 64–73. doi:10.1145/2812802.

44. Tran, T., Le, T., Pham, D., Hoang, V., Khong, V., Tran, Q., Nguyen, T., and Pham, C. A multi-modal multi-view dataset for human fall analysis and preliminary investigation on modality. In: *24th International Conference on Pattern Recognition, ICPR 2018, Beijing, China, August 20-24, 2018*. IEEE Computer Society (2018):(1947–1952). doi:10.1109/ICPR.2018.8546308.

45. Sanabria, R., Caglayan, O., Palaskar, S., Elliott, D., Barrault, L., Specia, L., and Metze, F. (2018). How2: A large-scale dataset for multimodal language understanding. CoRR *abs/1811.00347*. arXiv:1811.00347.

46. Tripathi, S., and Beigi, H. S. M. (2018). Multi-modal emotion recognition on IEMOCAP dataset using deep learning. CoRR *abs/1804.05788*. arXiv:1804.05788.

47. Menze, B. H., Jakab, A., Bauer, S., Kalpathy-Cramer, J. et al. (2015). The multimodal brain tumor image segmentation benchmark (BRATS). IEEE Trans. Medical Imaging *34*, 1993–2024. doi:10.1109/TMI.2014.2377694.

48. Bekker, A. J., Greenspan, H., and Goldberger, J. A multi-view deep learning architecture for classification of breast microcalcifications. In: *13th IEEE International Symposium on Biomedical Imaging, ISBI 2016, Prague, Czech Republic, April 13-16, 2016*. IEEE (2016):(726–730). doi:10.1109/ISBI.2016.7493369.

49. Liu, S., Liu, S., Cai, W., Che, H., Pujol, S., Kikinis, R., Feng, D., and Fulham, M. J. (2015). Multimodal neuroimaging feature learning for multiclass diagnosis of alzheimer's disease. IEEE Trans. Biomed. Eng. *62*, 1132–1140. doi:10.1109/TBME.2014.2372011.

50. Dhungel, N., Carneiro, G., and Bradley, A. P. Fully automated classification of mammograms using deep residual neural networks. In: *14th IEEE International Symposium on Biomedical Imaging, ISBI 2017, Melbourne, Australia, April 18-21, 2017*. IEEE (2017):(310–314). doi:10.1109/ISBI.2017.7950526.

51. Singh, R., and Khare, A. (2014). Fusion of multimodal medical images using daubechies complex wavelet transform - A multiresolution approach. Inf. Fusion *19*, 49–60. doi:10.1016/j.inffus.2012.09.005.

52. Messner, E., Fediuk, M., Swatek, P., Scheidl, S., Smolle-Jüttner, F., Olschewski, H., and Pernkopf, F. (2020). Multi-channel lung sound classification with convolutional recurrent neural networks. Comput. Biol. Medicine *122*, 103831. doi:10.1016/j.compbiomed.2020.103831.

53. Tong, T., Gray, K. R., Gao, Q., Chen, L., and Rueckert, D. (2017). Multi-modal classification of alzheimer's disease using nonlinear graph fusion. Pattern Recognit. *63*, 171–181. doi:10.1016/j.patcog.2016.10.009.

54. Li, W., Qi, F., Tang, M., and Yu, Z. (2020). Bidirectional LSTM with self-attention mechanism and multi-channel features for sentiment classification. Neurocomputing *387*, 63–77. doi:10.1016/j.neucom.2020.01.006.

55. Sun, S. (2013). A survey of multi-view machine learning. Neural Comput. Appl. *23*, 2031–2038. doi:10.1007/s00521-013-1362-6.

56. Xu, C., Tao, D., and Xu, C. (2013). A survey on multi-view learning. CoRR *abs/1304.5634*. arXiv:1304.5634.

57. Dolci, G., Rahaman, M. A., Chen, J., Duan, K., Fu, Z., Abrol, A., Menegaz, G., and Calhoun, V. D. A deep generative multimodal imaging genomics framework for alzheimer's disease prediction. In: *22nd IEEE International Conference on Bioinformatics and Bioengineering, BIBE 2022, Taichung, Taiwan, November 7-9, 2022*. IEEE (2022):(41–44). doi:10.1109/BIBE55377.2022.00017.

58. Kang, H., Xia, L., Yan, F., Wan, Z., Shi, F., Yuan, H., Jiang, H., Wu, D., Sui, H., Zhang, C., and Shen, D. (2020). Diagnosis of coronavirus disease 2019 (COVID-19) with structured latent multi-view representation learning. IEEE Trans. Medical Imaging *39*, 2606–2614. doi:10.1109/TMI.2020.2992546.

59. Huang, S., Shi, W., Xu, Z., Tsang, I. W., and Lv, J. (2022). Efficient federated multi-view learning. Pattern Recognit. *131*, 108817. doi:10.1016/j.patcog.2022.108817.

60. Kang, Y., Liu, Y., and Liang, X. (2022). Fedcvt: Semi-supervised vertical federated learning with cross-view training. ACM Trans. Intell. Syst. Technol. *13*, 64:1–64:16. doi:10.1145/3510031.

61. Yang, Y., Ye, X., and Sakurai, T. Multi-view federated learning with data collaboration. In: *ICMLC 2022: 14th International Conference on Machine Learning and Computing, Guangzhou, China, February 18 - 21, 2022*. ACM (2022):(178–183). doi:10.1145/3529836.3529904.

62. Qayyum, A., Ahmad, K., Ahsan, M. A., Al-Fuqaha, A. I., and Qadir, J. (2022). Collaborative federated learning for healthcare: Multi-modal COVID-19 diagnosis at the edge. IEEE Open J. Comput. Soc. *3*, 172–184. doi:10.1109/OJCS.2022.3206407.

63. Che, S., Kong, Z., Peng, H., Sun, L., Leow, A. D., Chen, Y., and He, L. (2022). Federated multi-view learning for private medical data integration and analysis. ACM Trans. Intell. Syst. Technol. *13*, 61:1–61:23. doi:10.1145/3501816.

64. Al-Jarrah, O. Y., Yoo, P. D., Muhaidat, S., Karagiannidis, G. K., and Taha, K. (2015). Efficient machine learning for big data: A review. Big Data Res. *2*, 87–93. doi:10.1016/j.bdr.2015.04.001.

65. Alam, M. U., and Rahmani, R. (2023). Fedsepsis: A federated multi-modal deep learning-based internet of medical things application for early detection of sepsis from electronic health records using raspberry pi and jetson nano devices. Sensors *23*. URL: https://www.mdpi.com/1424-8220/23/2/970.

66. Zhao, Y., Barnaghi, P. M., and Haddadi, H. Multimodal federated learning on iot data. In: *Seventh IEEE/ACM International Conference on Internet-of-Things Design and Implementation, IoTDI 2022, Milano, Italy, May 4-6, 2022*. IEEE (2022):(43–54). doi:10.1109/IoTDI54339.2022.00011.

67. Bernecker, T., Peters, A., Schlett, C. L., Bamberg, F., Theis, F. J., Rueckert, D., Weiß, J., and Albarqouni, S. (2022). Fednorm: Modality-based normalization in federated learning for multi-modal liver segmentation. CoRR *abs/2205.11096*. doi:10.48550/arXiv.2205.11096. arXiv:2205.11096.

68. Aggarwal, P., Mishra, N. K., Fatimah, B., Singh, P., Gupta, A., and Joshi, S. D. (2022). COVID-19 image classification using deep learning: Advances, challenges and opportunities. Comput. Biol. Medicine *144*, 105350. doi:10.1016/j.compbiomed.2022.105350.

69. Ozkaya, U., Öztürk, S., and Barstugan, M. (2020). Coronavirus (COVID-19) classification using deep features fusion and ranking technique. CoRR *abs/2004.03698*. arXiv:2004.03698.

70. Barstugan, M., Ozkaya, U., and Öztürk, S. Coronavirus (covid-19) classification using CT images by machine learning methods. In: Xhina, E., and Hoxha, K., eds. *Proceedings of the 4th International Conference on Recent Trends and Applications in Computer Science and Information Technology, Tirana, Albania, May 21st - to - 22nd, 2021* vol. 2872 of *CEUR Workshop Proceedings*. CEUR-WS.org (2021):(29–35).

71. Sadat, S. U., Shomee, H. H., Awwal, A., Amin, S. N., Reza, M. T., and Parvez, M. Z. Alzheimer's disease detection and classification using transfer learning technique and ensemble on convolutional neural networks. In: *2021 IEEE International Conference on Sys-*

tems, Man, and Cybernetics, SMC 2021, Melbourne, Australia, October 17-20, 2021. IEEE (2021):(1478–1481). doi:10.1109/SMC52423.2021.9659179.

72. Liu, F., Zhou, L., Shen, C., and Yin, J. (2014). Multiple kernel learning in the primal for multimodal alzheimer's disease classification. IEEE J. Biomed. Health Informatics *18*, 984–990. doi:10.1109/JBHI.2013.2285378.

73. Zhang, D., Wang, Y., Zhou, L., Yuan, H., and Shen, D. (2011). Multimodal classification of alzheimer's disease and mild cognitive impairment. NeuroImage *55*, 856–867. doi:10.1016/j.neuroimage.2011.01.008.

74. Gao, X., Shi, F., Shen, D., and Liu, M. (2022). Task-induced pyramid and attention GAN for multimodal brain image imputation and classification in alzheimer's disease. IEEE J. Biomed. Health Informatics *26*, 36–43. doi:10.1109/JBHI.2021.3097721.

75. Zhang, D., and Shen, D. Semi-supervised multimodal classification of alzheimer's disease. In: *Proceedings of the 8th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, ISBI 2011, March 30 - April 2, 2011, Chicago, Illinois, USA*. IEEE (2011):(1628–1631). doi:10.1109/ISBI.2011.5872715.

76. He, K., Zhang, X., Ren, S., and Sun, J. Deep residual learning for image recognition. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016*. IEEE Computer Society (2016):(770–778). doi:10.1109/CVPR.2016.90.

77. Aldinucci, M., Rabellino, S., Pironti, M., Spiga, F., Viviani, P., Drocco, M., Guerzoni, M., Boella, G., Mellia, M., Margara, P., Drago, I., Marturano, R., Marchetto, G., Piccolo, E., Bagnasco, S., Lusso, S., Vallero, S., Attardi, G., Barchiesi, A., Colla, A., and Galeazzi, F. HPC4AI, an AI-on-demand federated platform endeavour. In: *ACM Computing Frontiers* (2018):doi:10.1145/3203217.3205340.

78. Reina, G. A., Gruzdev, A., Foley, P., Perepelkina, O., Sharma, M., Davidyuk, I. et al. (2021). Openfl: An open-source framework for federated learning. CoRR *abs/2105.06413*.

79. O'Shea, K., and Nash, R. (2015). An introduction to convolutional neural networks. CoRR *abs/1511.08458*. arXiv:1511.08458.

80. Clark, J. Y., and Warwick, K. (1998). Artificial keys for botanical identification using a multilayer perceptron neural network (MLP). Artif. Intell. Rev. *12*, 95–115. doi:10.1023/A:1006544506273.

81. Weiner, M. W., Veitch, D. P., Aisen, P. S., Beckett, L. A., Cairns, N. J., Green, R. C., Harvey, D., Jack, C. R., Jagust, W., Liu, E. et al. (2013). The alzheimer's disease neuroimaging initiative: a review of papers published since its inception. Alzheimer's & Dementia *9*, e111–e194.

82. Weiner, M. W., Veitch, D. P., Aisen, P. S., Beckett, L. A., Cairns, N. J., Green, R. C., Harvey, D., Jack Jr, C. R., Jagust, W., Morris, J. C. et al. (2017). The alzheimer's disease neuroimaging initiative 3: Continued innovation for clinical trial improvement. Alzheimer's & Dementia *13*, 561–571.

83. Jenkinson, M., Beckmann, C. F., Behrens, T. E., Woolrich, M. W., and Smith, S. M. (2012). Fsl. Neuroimage *62*, 782–790.