Federated Similarity-Based Learning with Incomplete Data

Federated similarity-based learning with incomplete data

1st Barbara Pekala
University of Rzeszów,
University of Information Technology and Management
Rzeszów, Poland
bpekala@ur.edu.pl

2nd Jarosław Szkoła
University of Rzeszów
Rzeszów, Poland
jszkola@ur.edu.pl

3th Krzysztof Dyczkowski
Adam Mickiewicz University
Poznań, Poland
chris@amu.edu.pl

4rd Anna Wilbik
Maastricht University
Maastricht, The Netherlands
a.wilbik@maastrichtuniversity.nl

***Abstract*—In the analysis of social, medical, and business issues, the problem of incomplete data often arises. In addition, in situations where privacy policy makes it difficult to share data with organizations conducting related activities, it is necessary to exchange knowledge instead of data, that is, to use federated learning. In this scenario there are several private data clients, whose models are improved through the aggregation of model components. Here, we propose a methodology for training local models to deal well with missing data, with an algorithm using similarity measures that take into account the uncertainty present in many types of data, such as medical data. Therefore, this paper describes a federated learning model capable of processing imprecise and missing data. Federated learning is a technique to overcome limitations resulting from data governance and privacy by training algorithms without exchanging the data itself. The performance of the proposed method is demonstrated using medical data on breast cancer cases. Results for different data loss scenarios and corresponding measures of classification quality are presented and discussed.**

I. Introduction

In the activities of many organizations, the use of information from other analogous sources – namely, horizontal federated learning – enables improvement of the individual predictions made by their models, and consequently, better business results. At present, issues related to competition or privacy make it impossible to combine data. Federated learning (FL) allows organizations to bypass these problems and to train models efficiently without sharing data. The general FL training process consists of five steps:

- Client selection: Select the clients participating in the training;
- Broadcast: The central server initializes the global model and makes it available to clients;
- Client calculation: Each client updates the global model using the training protocol and makes the updates available to the central server;
- Aggregation: The central server uses the aggregation function to update its model;
- Model update: The updated global model is made available to customers.

This protocol can be repeated many times until a convergence criterion is met. Such a process of model training has been successfully applied in various use cases, especially involving data collaboration in the medical sector [1], [2]. An international group of hospitals and medical imaging centers recently evaluated NVIDIA Clara Federated Learning software, and found that AI models for mammography evaluation trained with federated learning techniques outperform neural networks trained on data from a single institution. Other fields of application include next-word prediction [3] (Apple is using privacy machine learning and FL to improve its voice assistant while protecting data on users' phones), vehicle image classification [4], IoT data analysis [5], and social research [6]. Comprehensive surveys can be found in [7] and [8]. In many cases of practical FL application, some data are missing, which limits the amount of data that can be used in the model training process. This, along with the problem of privacy, is one of the most important challenges facing federated learning.

Therefore, the main aim of this work is to propose an optimal learning model that performs well in FL with the problem of missing or incomplete data. In practice, we propose a new technique using a similarity measure that reflects the uncertainty implied by lack of data. Here, the gradient-based approach is replaced by a properly selected function based on the interval similarity measure. In this paper, we concentrate on the representation of data and operations, especially similarity measures with respect to uncertainty relating to data and decisions. Thus, we will study the result of applying interval-valued fuzzy set theory in the federated learning process, because the interval calculus well reflects various types of uncertainties contained in the data, here resulting from its incompleteness. Simulations are carried out which confirm an improvement in effectiveness compared with the methods used in the literature, in particular regression.

II. Background

A. Federated learning (FL)

Federated learning enables collaboration between multiple parties for the purpose of jointly training a machine learning model without exchanging the local data [7]. The federated learning model was originally proposed by Google researchers [9], [10], [11]. Their main idea was to build machine learning models based on datasets that are distributed across multiple devices (cf. [12], [13] or [14]).

Federated learning is a learning paradigm seeking to address the problem of data governance and privacy by training algorithms collaboratively without exchanging the data itself [12], [15]. The core challenges associated with solving the optimization problem during federated learning make the federated setting distinct from other classical problems, such as distributed learning in data center settings or traditional private data analyses. These challenges are communication, heterogeneity, and privacy. Generally, FL can be divided into different scenarios based on how the data is partitioned or distributed among the data owners, that is, horizontally or vertically. Horizontal federated learning is used when different parties collect the same features but from different subjects. A common example of horizontal federated learning is a group of hospitals collaborating to build a model that can predict a health risk for their patients, based on agreed data, as in [1], [2]. Vertical federated learning is used when multiple parties share not the features, but the subjects; for example, when a telecom company collaborates with a home entertainment company (cable television provider) or an airline collaborates with a car rental agency.

In this paper, we consider a horizontal federated learning scenario. Figure 1 shows the general architecture of the federated model. The assumption is that all clients have the same local data structure and use a common machine learning model. They exchange with the server only coefficients describing the learned local models and parameters describing the classification quality, which is used only to determine when to stop the iteration process. The server performs model aggregation, that is, the appropriate aggregation of coefficients. The server then returns the new coefficients to the clients.

In federated settings, optimization methods that allow flexible local updating and low client participation play a key role. The most commonly used method for federated learning is a method based on averaging local stochastic gradient descent (SGD) updates for the primal problem ([11], [16], [17] or more in [15]). Our approach presents a different concept in this respect, namely, the use in local updating of a function based on similarity measures, in particular using the calculus of interval-valued fuzzy sets, i.e. interval similarity measures.

B. Interval-valued fuzzy set theory

Since Zadeh proposed the fuzzy set in 1965 [18], and particularly since 1975, when research on extensions of fuzzy sets began [19], [20], the effective modeling of uncertainty and imprecision in data has been possible. Thus, we may describe the data in terms of interval calculations. Specifically, $L^I = \{[\underline{p}, \overline{p}] : \ \underline{p}, \overline{p} \in [0,1], \ \underline{p} \leq \overline{p}\}$ denotes a family of intervals belonging to the unit interval.

*1) **Interval operations**:* Many applications of AI require data aggregation to summarize information from data. Aggregate functions take, as input sets, multisets (bags) from an input range, and produce outputs as members of an output range. A definition of aggregation for input data in the form of interval-valued fuzzy values – that is, with uncertainty – can be found in [21], [22], [23]. Certain aggregate functions in L^I, namely interval-valued fuzzy aggregation functions, are important concepts in many applications (e.g. [24], [25] or [22]).

Moreover, interval arithmetic came to be considered necessary with the development of the theory of uncertainty. It was realized that the use of uncertain parameters and uncertain data is very important for the description of reality in the form of a mathematical model. The most common and most frequently used interval arithmetic is Moore arithmetic [26], [27]. In Moore arithmetic, basic operations on intervals $X = [\underline{x}, \overline{x}]$ and $Y = [\underline{y}, \overline{y}]$ are realized by formulae for sum, difference, and product:

$$[\underline{x}, \overline{x}] + [\underline{y}, \overline{y}] = [\underline{x} + \underline{y}, \overline{x} + \overline{y}]$$

$$[\underline{x}, \overline{x}] - [\underline{y}, \overline{y}] = [\underline{x} - \overline{y}, \overline{x} - \underline{y}]$$

$$a * [\underline{x}, \overline{x}] = [a\underline{x}, a\overline{x}], \quad a \in R^+$$

$$a * [\underline{x}, \overline{x}] = [a\overline{x}, a\underline{x}], \quad a \in R^-$$

$$[\underline{x}, \overline{x}] * [\underline{y}, \overline{y}] =$$

$$[\min(\underline{x} * \underline{y}, \overline{x} * \overline{y}, \underline{x} * \overline{y}, \overline{x} * \underline{y}), \max(\underline{x} * \underline{y}, \overline{x} * \overline{y}, \underline{x} * \overline{y}, \overline{x} * \underline{y})]$$

for $\underline{x}, \overline{x}, \underline{y}, \overline{y} \in R$ and $\underline{x} \leq \overline{x}, \underline{y} \leq \overline{y}$.

Some limitations and drawbacks have been found in the Moore interval arithmetic scheme, such as the excess width effect problem. Hence, as an alternative to Moore arithmetic we may use multidimensional interval arithmetic. The idea of multidimensional arithmetic was developed by A. Piegat [28], where a given value x from the interval $X = [\underline{x}, \overline{x}]$ is described using the variable γ_x, where $\gamma_x \in [0,1]$, as follows:

$$Rep_\gamma(x) = \underline{x} + \gamma_x(\overline{x} - \underline{x}). \tag{1}$$

In this notation the interval $X = [\underline{x}, \overline{x}]$ is described in the form:

$$X = \{Rep_\gamma(x) : Rep_\gamma(x) = \underline{x} + \gamma_x(\overline{x} - \underline{x}), \gamma_x \in [0,1]\}.$$

The variable γ_x provides the possibility of obtaining any value between the left boundary $\underline{x}$ and the right boundary $\overline{x}$ of the interval X.

*2) **Interval measures**:* Crucial to our methodology is the similarity measure for interval-valued fuzzy sets (IVFS). We define an **interval-valued fuzzy set** (IVFS) S in X [20], [19] as a mapping $S : X \to L^I$ such that for each $x \in X$

$$S(x) = [\underline{S}(x), \overline{S}(x)]$$

means the degree of membership of an element x in S. The family of all IVFSs in X is denoted by IVFS(X). We assume that this reflects the aspect of applications on a finite non-empty set $X = \{x_1, \ldots, x_n\}$.

Definition 1 ([29], cf. [30])**.** Let $A_1 : [0,1]^n \to [0,1]$ be an aggregation function. Then a function $SIM : IVFS(X) \times IVFS(X) \to L^I$ which meets the conditions: leftmargin=.4in

(S_1) $SIM(S,T) = SIM(T,S)$ for $S,T \in IVFS(X)$;

(S_2) $SIM(S,S) = [1 - A_1(w_S(x_1), ..., w_S(x_n)), 1]$;

(S_3) $SIM(S,T) = [0,0]$, if $\{S(x_i), T(x_i)\} = \{[0,0],[1,1]\}$;

(S_4) if $S \preceq T \preceq U$, then $SIM(S,U) \leq SIM(S,T)$ and $SIM(S,U) \leq SIM(T,U)$

is called a similarity measure for $i = 1, ..., n$..

To construct interval-valued similarity, we need interval aggregation functions ([21], [22], [23]) and an inclusion measure (precedence indicator) ([29], cf. [31]) that take into account the width of the intervals, that is, the uncertainty.

Proposition 1 ([29])**.** *Let* Prec *be a precedence indicator. If* $\mathcal{A} = [A_1, A_2]$, $\mathcal{B} = [B_1, B_2]$ *are representable interval-valued fuzzy aggregation functions for which* A_1 *is self-dual,* $\mathcal{B}$ *is symmetric with the neutral element* $[1,1]$ *and* B_1 *is an idempotent aggregation function, then the function* $SIM : IVFS(X) \times IVFS(X) \rightarrow L^I$*:*

$$SIM(S,T) = \mathcal{A}_{i=1}^{n}(\mathcal{B}(\text{Prec}(S(x_i), T(x_i)), \text{Prec}(T(x_i), S(x_i))))$$

is a similarity measure.

The following example presents direct conclusions from the above theorem.

Example 1. The function $SIM : IVFS(X) \times IVFS(X) \rightarrow L^I$:

$$SIM(S,T) = \mathcal{A}_{i=1}^{n}(\text{Prec}_{\mathcal{A}}(S(x_i), T(x_i) \wedge \text{Prec}_{\mathcal{A}}(T(x_i), S(x_i))))$$

is a similarity measure, where $\mathcal{A} \in \{\mathcal{A}_{mean}, \mathcal{A}_{meanpow}, \mathcal{A}_{meanmax}\}$ with respective precedence indicators $\text{Prec}_{\mathcal{A}_{mean}}$, $\text{Prec}_{\mathcal{A}_{meanpow}}$ and $\text{Prec}_{\mathcal{A}_{meanmax}}$, where

$$\text{Prec}_{\mathcal{A}_{mean}}(x,y) = \begin{cases} [1 - w(x), 1], & x = y, \\ [1,1], & x <_2 y, \\ [\frac{1-\overline{x}+\underline{y}}{2}, \frac{1-\underline{x}+\overline{y}}{2}], & \text{otherwise}, \end{cases}$$

$$\text{Prec}_{\mathcal{A}_{meanpow}}(x,y) = \begin{cases} [1 - w(x), 1], & x = y, \\ [1,1], & x <_2 y, \\ [\frac{1-\overline{x}+\underline{y}}{2}, \sqrt{\frac{(1-\underline{x})^2+\overline{y}^2}{2}}], & \text{otherwise}, \end{cases}$$

$$\text{Prec}_{\mathcal{A}_{meanmax}}(x,y) = \begin{cases} [1 - w(x), 1], & x = y, \\ [1,1], & x <_2 y, \\ [\frac{1-\overline{x}+\underline{y}}{2}, \max(1-\underline{x}, \overline{y})], & \text{otherwise}, \end{cases}$$

and

$\mathcal{A}_{mean}([\underline{x}, \overline{x}], [\underline{y}, \overline{y}]) = [\frac{\underline{x}+\underline{y}}{2}, \frac{\overline{x}+\overline{y}}{2}]$,

$\mathcal{A}_{meanpow}([\underline{x}, \overline{x}], [\underline{y}, \overline{y}]) = [\frac{\underline{x}+\underline{y}}{2}, \sqrt{\frac{\overline{x}^2+\overline{y}^2}{2}}]$,

$\mathcal{A}_{meanmax}([\underline{x}, \overline{x}], [\underline{y}, \overline{y}]) = [\frac{\underline{x}+\underline{y}}{2}, \max(\overline{x}, \overline{y})]$ and where $\leq_2$ is a partial order in L^I:

$$x \leq_2 y \ \ iff \ \ \underline{x} \leq \underline{y}, \overline{x} \leq \overline{y}.$$

III. Proposed Method

We consider a horizontal federated learning scenario, where each client has its own independent data set $z_i \in \{Y_i, x_{i1}, ...x_{ip}\}$ and $x_{ip} \in L^I$, $Y_i \in \{0,1\}$ for $i = 1, ..., n$, n is the number of instances, and p is the number of attributes.

Each client trains a set model on its data (n_k observations) in a specified number of internal iterations, and provides the training result in the form of a result vector of the trained parameters β and ϵ,

$$y_i = \beta_0 + \beta_1 x_{i1} + ... + \beta_p x_{ip} + \epsilon_i$$

for $i = 1, ...n_k$ and $\beta_k \in R$ for $k = 1, ...p$.

In our earlier paper [32], we proposed a federated learning approach that could deal with missing data (Figure 1).

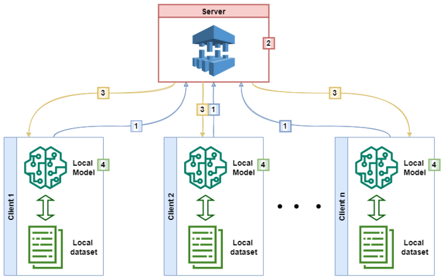

Fig. 1. Proposed federated model

There we used the arithmetic average in the aggregation process, and the process of training models was based on the gradient method.

However, in this paper, we investigate another method in the process of training models, based on the similarity measure. The new learning model, after the server initializes the model sent to local models, consists of iterative executions (following initialization) of the following steps:

1) Each client performs a few steps of training of its own model on its local data and passes it to the server. The training mechanism makes use of the similarity measure with respect to uncertainty;
2) The server aggregates the models;
3) The server returns the new model to the clients;
4) Local models are updated if the new one is better.

The process continues until the acquired quality of local models is high enough and it is impossible to improve them, that is to say, subsequent iterations do not reduce the error of the model. In other words, we propose that the federated learning scheme be extended to include the sensitivity threshold Q as a stop function in the validation process for the error difference for the also fixed multiplicity correction of model parameters β.

As mentioned above, the federated learning scheme thus constructed is independent of the choice of a particular machine learning model. We chose logistic regression with the use of similarity measures in the process of updating the parameters of the model, in contrast to the classically used stochastic gradient descent. Moreover, for the experiment, we modify it to operate on interval data, as in [32].

Then one iteration of the local learning process follows the scheme:

1) calculation of the model response for each training sample according to the sigmoid function:

$$f(y_i) = \frac{1}{1 + e^{-Rep_{\gamma}(\beta_0 + \beta_1 \cdot x_{i1} + ... + \beta_p \cdot x_{ip} + \epsilon_i)}}$$

for $\gamma \in [0,1]$ and $f : L^I \rightarrow R$.

From this step, in every single iteration, we switch from the interval calculus to the real model using the Rep function defined in (1). This allows us to operate on data in the form of interval-valued fuzzy sets while obtaining the model in the form of a vector of real numbers.

2) For the computation of an error (loss function) between the computed value and the actual value, we take

$$\mathcal{L}(y_i) = -\log(f(y_i)) \cdot Y_i - \log(1 - f(y_i)) \cdot (1 - Y_i),$$

where Y_i is the actual output value for a given object z_i.

3) Finally, we update the learning coefficients in the steps:

$$\beta_l(k+1) = \beta_l(k) + \alpha \cdot \mathcal{L}(y_i) \cdot \frac{\frac{1}{t-1}\sum_{j=1}^{t-1} S(z_i, z_j)}{\frac{1}{n}\sum_{j=1}^{n} S(z_i, z_j)} \cdot$$

$$\max_{j=1}^{t-1}\{S(x_{il}, x_{jl})\},$$

$$\beta_0(k+1) = \beta_0(k) + \alpha \cdot \mathcal{L}(y_i) \cdot \frac{\frac{1}{t-1}\sum_{j=1}^{t-1} S(z_i, z_j)}{\frac{1}{n}\sum_{j=1}^{n} S(z_i, z_j)},$$

where α is the learning coefficient, t is the number of objects with the same decision for z_i, $i = 1, .., n_k$, the number of a given attribute is $l \in \{1, .., p\}$, and S is the similarity measure.

IV. Experiment and Results

In this section, we describe our initial evaluation of the proposed method.

A. Structure of dataset

The dataset used is a Wisconsin (diagnostic) breast cancer dataset. This is one of the popular datasets from the UCI Machine Learning Repository [33]. The data contain information on 569 medical cases. Features are calculated from a digitized image of a fine needle aspirate (FNA) of a breast mass. They describe the characteristics of the cell nuclei present in the image.

Ten real-valued features are computed for each cell nucleus:

- radius (mean of distances from center to points on the perimeter),
- texture (standard deviation of gray-scale values),
- perimeter,
- area,
- smoothness (local variation in radius lengths),
- compactness,
- concavity (severity of concave portions of the contour),
- concave points (number of concave portions of the contour),
- symmetry,
- fractal dimension ($coastline\ approximation - 1$).

For each value of an attribute, the standard deviation and mean value of the trait measurements for the patient are given. On the basis of both of these values, the value of the interval is constructed:

$$[mean - standard\ deviation,\ mean + standard\ deviation]$$

Later we fuzzified both values ”mean–standard deviation” and ”mean+standard deviation”, separately, by normalization.

The decision attribute stores information about the diagnosis: malignant (0) or benign (1). The dataset consists of 212 malignant objects and 357 benign objects. Since the dependent variable (explained variable) takes two dichotomous values of 0 and 1, the optimal model choice for decision prediction turned out to be the logistic regression model, which determines the probability of a given event occurring for the values of the predictors entered into the model.

To simulate the datasets of a group of clients (three in this case), the data were randomly divided into three groups, with decision-balanced and unbalanced behavior. The data for each client were then randomly split into a training set and a test set in a ratio of 90% to 10%.

In our model, we allow the data to be in interval form. First, for gaps in the data, we create suitable intervals to reflect uncertainty. We simulate data gaps using the loss method. We assume that the data are normalized, and the missing data are presented in the form of intervals $[0, 1]$.

B. Experimental results for different real problem scenarios

We checked our model in various real-life scenarios involving uncertain and missing data, and compared it with the crisp model (benchmark). We assumed, for each iteration of the algorithm described in section III, $\epsilon_i = 0$, $\alpha = 0.01$, and $\gamma = 0.5$ (the optimal results). Moreover, we simulated a number of cases of the number of epochs in local learning and the number of aggregations, respectively: (100, 5), (5, 100), (10, 10), and others. We obtained very similar results (indicating the stability of the similarity-based method), and therefore we present only the results for the efficiency of the model for 100 learning epochs and 5 aggregations in FL. We assess the effectiveness of the tested models using the following measures: accuracy (ACC), sensitivity (SENS), specificity (SPEC), and precision (PREC).

Validation. In a federated model, we wish to achieve the best possible global model, that is, one that achieves high decision performance across all clients. Therefore, models should be analyzed not only on the local data of a given client, but also using the data of other clients (although without direct access to them). Our proposed federated learning model enables this exchange of model quality information. Validation is carried out in two stages: during each local learning phase and also after model aggregation, so that we check that the new parameters do not make the model worse. Finally, the client decides whether to update its model and strive for the highest quality global model. Moreover, the use of a new error sensitivity threshold is a new approach in the validation process.

1) Model 1 – benchmark model: As a benchmark model, we chose a centralized model in which the data are complete and lack uncertainty. The model was trained on a 90% training set (sum of customer sets) and tested on a 10% test set. To ensure the correctness of the learning process, we conducted a 10-fold cross-validation. That is, the modifying logistic regression model was used without a federated learning model.

The reference performance of the benchmark model is presented in Table I.

TABLE I
PERFORMANCE OF BENCHMARK MODEL

	ACC	SENS	SPEC	PREC
Complete data	0.955	0.972	0.901	0.935

2) Model 2 – baseline model (local models): As a baseline, we decided to consider a situation in which both clients have uncertain interval data with no missing values (complete uncertain data) and without FL – that is, a complete interval-valued dataset without data gaps (performance based on 10-fold cross-validation); see Table II.

TABLE II
PERFORMANCE OF BASELINE MODEL

Dataset	ACC	SENS	SPEC	PREC
Client 1	0.901	0.954	0.816	0.893
Client 2	0.960	0.978	0.903	0.947

3) Model 3 – federated model on full data: This model was based on full data and was trained with federated averaging as proposed in section III. The learning rate was set to 0.01, there were 100 local learning epochs, and the stopping criterion was set to 5 aggregation cycles. The results are given in Table III.

TABLE III
PERFORMANCE OF BASE FEDERATED MODEL

Dataset	ACC	SENS	SPEC	PREC
Client 1	0.914	0.964	0.845	0.899
Client 2	0.964	0.978	0.913	0.951

4) Model 4 – with missing random values in different attributes: This model reflects two situations when random data are missing to some degree, potentially in all attributes: without and with federated averaging. In real-world conditions, this may result from measurement equipment malfunctions, improper testing, or human error. In our experiment, we simulated situations where we had random missing data distributed one per record and different levels of missing data: from 10% to 50% of values (records). Also, in this case, data from only one client were deleted; the other had full data.

In this scenario, the first client has a prepared dataset with different percentages of missing data (10–50%), and the second client has full data. Table IV gives the results for the first client's local model.

Results for the federated learning model (using the standard mean for aggregation) are presented in Table V. Calculations were performed on Client 1's test data.

TABLE IV
PERFORMANCE OF CLIENT 1'S LOCAL MODEL

% of missing data	ACC	SENS	SPEC	PREC
10	0.894	0.949	0.807	0.888
20	0.880	0.944	0.794	0.847
30	0.867	0.904	0.776	0.809
40	0.848	0.899	0.744	0.788
50	0.795	0.874	0.732	0.756

TABLE V
PERFORMANCE OF FL AGGREGATED MODEL FOR CLIENT 1

% of missing data	ACC	SENS	SPEC	PREC
10	**0.898**	**0.954**	**0.882**	**0.889**
20	**0.883**	**0.947**	**0.877**	**0.870**
30	**0.877**	**0.939**	**0.845**	**0.861**
40	**0.872**	**0.923**	**0.831**	**0.847**
50	**0.854**	**0.883**	**0.826**	**0.839**

C. Discussion

In this paper, we have extended our earlier approach to federated learning that could handle missing data using the classical logistic regression method [16]. In [16], using the classical logistic regression model and the method of learning model parameters with the use of a gradient, the decrease in efficiency (e.g. ACC) at only 10% missing data in one local model was 0.056, and when FL was used it was 0.033. However, using the new method of updating parameters proposed here, we observe a decrease of only 0.007 (without FL) and 0.003 (with FL). In addition, the decrease in efficiency as the amount of missing data increases from 10% to 50% is slower by about 0.03 in the case of the new method, as illustrated in Fig. 2.

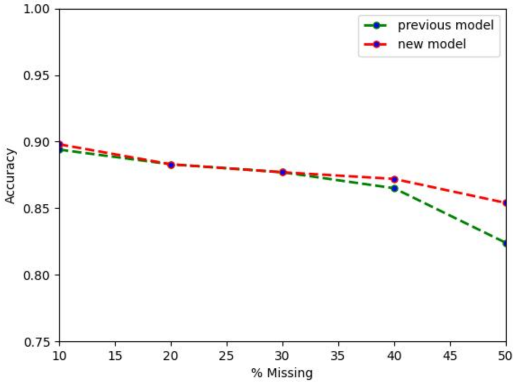

Fig. 2. Comparison of previous (green) and new (red) models for ACC and different percentages of missing values

This indicates a much more stable classification method and thus a more effective FL model.

V. CONCLUDING REMARKS

This paper describes a federated learning model able to process imprecise data with the problem of missed data. Medical data on breast cancer cases demonstrate the performance of the proposed method. Results for different data loss scenarios and corresponding measures of classification quality are presented and discussed. We observed that the proposed method used to learn the parameters of the federated models was more stable with respect to missing data. In future research, we plan to concentrate on two aspects: determination of the stop condition for local learning and model aggregation in the case of different data types and problem specifics, and selection of the parameter for the level of similarity of objects used in the algorithm.

References

[1] T. M. Deist, A. Jochems, J. van Soest, G. Nalbantov *et al.*, "Infrastructure and distributed learning methodology for privacy-preserving multicentric rapid learning health care: euroCAT," *Clinical and Translational Radiation Oncology*, vol. 4, pp. 24–31, 2017.

[2] T. S. Brisimi, R. Chen, T. Mela, A. Olshevsky, I. C. Paschalidis, and W. Shi, "Federated learning of predictive models from federated electronic health records," *International Journal of Medical Informatics*, vol. 112, pp. 59–67, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S138650561830008X

[3] A. Hard, K. Rao, R. Mathews, S. Ramaswamy, F. Beaufays, S. Augenstein, H. Eichner, C. Kiddon, and D. Ramage, "Federated learning for mobile keyboard prediction," 2018. [Online]. Available: https://arxiv.org/abs/1811.03604

[4] D. Ye, R. Yu, M. Pan, and Z. Han, "Federated learning in vehicular edge computing: A selective model aggregation approach," *IEEE Access*, vol. 8, pp. 23 920–23 935, 2020.

[5] X. Wang, Y. Han, C. Wang, Q. Zhao, X. Chen, and M. Chen, "In-edge AI: Intelligentizing mobile edge computing, caching and communication by federated learning," *IEEE Network*, vol. 33, no. 5, pp. 156–165, sep 2019.

[6] K. G. Liakos, P. Busato, D. Moshou, S. Pearson, and D. Bochtis, "Machine learning in agriculture: A review," *Sensors*, vol. 18, no. 8, 2018. [Online]. Available: https://www.mdpi.com/1424-8220/18/8/2674

[7] P. Kairouz, B. McMahan *et al.*, "Advances and open problems in federated learning," *Foundations and Trends® in Machine Learning*, vol. 14, pp. 1–210, 2021.

[8] L. U. Khan, W. Saad, Z. Han, E. Hossain, and C. S. Hong, "Federated learning for internet of things: Recent advances, taxonomy, and open challenges," 2020. [Online]. Available: https://arxiv.org/abs/2009.13012

[9] J. Konecný, H. B. McMahan, D. Ramage, and P. Richtárik, "Federated optimization: Distributed machine learning for on-device intelligence," *ArXiv*, vol. 1610.02527, 2016.

[10] J. Konečný, H. B. McMahan, F. X. Yu, P. Richtárik, A. T. Suresh, and D. Bacon, "Federated learning: Strategies for improving communication efficiency," *ArXiv*, vol. 1610.05492, 2017.

[11] H. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *AISTATS 2017*, 2017.

[12] Q. Yang, Y. Liu, T. Chen, and Y. Tong, "Federated machine learning: Concept and applications," *ACM Trans. Intell. Syst. Technol.*, vol. 10, no. 2, 2019.

[13] A. Wilbik and P. Grefen, "Towards a federated fuzzy learning system." IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), 2021, pp. 1–6.

[14] H. Yan, L. Hu, X. Xiang, Z. Liu, and X. Yuan, "Privacy-preserving collaborative learning for mitigating indirect information leakage," *Information Sciences*, vol. 548, pp. 423–437, 2021.

[15] T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, "Federated learning: Challenges, methods, and future directions," *IEEE Signal Processing Magazine*, vol. 37, no. 3, pp. 50–60, 2020.

[16] K. Dyczkowski, B. Pekala, J. Szkoła, and A. Wilbik, "Federated learning with uncertainty on the example of a medical data," in *2022 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*. IEEE, 2022, pp. 1–8.

[17] A. Wilbik, B. Pekala, K. Dyczkowski, and J. Szkoła, "A comparison of client weighting schemes in federated learning," in *IWIFSG'2022, Springer*, to appear.

[18] L. A. Zadeh, "Fuzzy sets," *Information and Control*, vol. 8, no. 3, pp. 338–353, 1965.

[19] R. Sambuc, "Fonctions ϕ-floues: Application á l'aide au diagnostic en pathologie thyroidienne," Ph.D. dissertation, Faculté de Médecine de Marseille, 1975, (in French).

[20] L. A. Zadeh, "The concept of a linguistic variable and its application to approximate reasoning—I," *Information Sciences*, vol. 8, no. 3, pp. 199–249, 1975.

[21] H. Zapata, H. Bustince, S. Montes, B. Bedregal, G. Dimuro, Z. Takáč, M. Baczyński, and J. Fernandez, "Interval-valued implications and interval-valued strong equality index with admissible orders," *International Journal of Approximate Reasoning*, vol. 88, pp. 91–109, 2017.

[22] G. Beliakov, H. B. Sola, and T. C. Sánchez, *A practical guide to averaging functions*, ser. Studies in Fuzziness and Soft Computing. Springer, 2016, vol. 329.

[23] M. Komorníková and R. Mesiar, "Aggregation functions on bounded partially ordered sets and their classification," *Fuzzy Sets and Systems*, vol. 175, no. 1, pp. 48–56, 2011.

[24] K. Dyczkowski, A. Wójtowicz, P. Żywica, A. Stachowiak, R. Moszyński, and S. Szubert, "An Intelligent System for Computer-Aided Ovarian Tumor Diagnosis," in *Intelligent Systems'2014*. Cham: Springer International Publishing, 2015, pp. 335–343.

[25] B. Pekala, *Uncertainty Data in Interval-Valued Fuzzy Set Theory: Properties, Algorithms and Applications*, ser. Studies in Fuzziness and Soft Computing. Springer, 2019, vol. 367.

[26] R. E. Moore, *Interval analysis*. Prentice Hall, 1966.

[27] ——, *Methods and applications of interval analysis*. SIAM, 1979.

[28] A. Piegat and M. Landowski, "Multidimensional approach to interval uncertainty calculations," in *New Trends in Fuzzy Sets, Intuitionistic: Fuzzy Sets, Generalized Nets and Related Topics, Volume II: Applications*, K. Atanassov *et al.*, Eds. Warsaw: IBS PAN - SRI PAS, 2013, p. 137–151.

[29] B. Pekala, D. Kosior, K. Dyczkowski, and J. Szkoła, "Application of entropy measures with uncertainty in classification methods with missing data problem," in *IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, 2021, pp. 1–8.

[30] H. Bustince, C. Marco-Detchart, J. Fernandez, C. Wagner, J. Garibaldi, and Z. Takáč, "Similarity between interval-valued fuzzy sets taking into account the width of the intervals and admissible orders," *Fuzzy Sets and Systems*, vol. 390, no. 1, pp. 23–47, 2020.

[31] B. Pekala, U. Bentkowska, M. Sesma-Sara, J. Fernandez, J. Lafuente, A. Altalhi, M. Knap, H. Bustince, and J. M. Pintor, "Interval subsethood measures with respect to uncertainty for the interval-valued fuzzy setting," *International Journal of Computational Intelligence Systems*, vol. 13, pp. 167–177, 2020.

[32] B. Pekala, T. Mroczek, D. Gil, and M. Kepski, "Application of fuzzy and rough logic to posture recognition in fall detection system," *Sensors*, vol. 22, no. 4, 2022. [Online]. Available: https://www.mdpi.com/1424-8220/22/4/1602

[33] "UCI machine learning repository," 2017. [Online]. Available: http://archive.ics.uci.edu/ml