Federated learning with uncertainty on the example of a medical data

Federated learning with uncertainty on the example of a medical data

1st Krzysztof Dyczkowski
Adam Mickiewicz University
Poznań, Poland
chris@amu.edu.pl

2nd Barbara Pękala
University of Rzeszów,
University of Information Technology and Management
Rzeszów, Poland
bpekala@ur.edu.pl

3rd Jarosław Szkoła
University of Rzeszów
Rzeszów, Poland
jszkola@ur.edu.pl

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

***Abstract*—This paper describes a federated learning model capable to process imprecise and missing data. Federation learning is a technique to solve the problem of data governance and privacy by training algorithms without exchanging the data itself. The performance of the proposed method is demonstrated on medical data of breast cancer cases. Results for different data loss scenarios and corresponding measures of classification quality are presented and discussed.**

I. Introduction

Successful training of models in machine learning requires data of high quality and in large quantity. In many research areas, collecting such data is difficult. This is very often due to the specifics of the problem at hand, legal and organizational constraints. For example, many research centers conducting medical research collect data about their patients which are used to build models that allow them to diagnose a given disease or prepare treatment procedures. In many cases, however, the data in these centers are too small to generate a sufficiently universal model or to be reliable for a wider group of cases. The natural idea would be to centralize the data and train models on common data, but this is impossible in many cases due to the sensitivity of the data, lack of patients' consent for their further transfer, or simply lack of trust in other research centers. The answer to this kind of problem is a federated learning model (see [1], [2]). The key idea of federated learning is to share not data but learned local models, which, appropriately aggregated and returned to individual units, would significantly increase the global quality of the resulting decision systems, while preserving data privacy.

One challenge of machine learning, but also federated learning is how to process imprecise incomplete data in such a model, resulting e.g. from lack of data in particular units. The incomplete data may be missing at random, caused e.g., by a random loss of some parameters due to failure or improperly collected data. Alternatively, they can be structurally incomplete, because of e.g., the inability to perform certain tests in one of the centers, working according a different protocol. In those cases, one could either exclude that center from the federation, or exclude that missing variable. Both solutions are not desired, as they exclude potentially useful data from the distributed data set.

In this paper, we extend the federated learning model with interval-valued fuzzy sets to use imprecise and/or incomplete data for training the models. We test our method on medical data on breast cancer diagnosis.

Representations and investigations in modeling of imprecision and uncertainty are still continued since the fuzzy set theory was introduced [3]. One of the many extensions of fuzzy sets (FSs), i.e. interval-valued fuzzy sets (IVFSs) ([4], [5]) similar to intuitionistic fuzzy sets (AIFSs) [6] occurred very useful of their flexibility [7]. Interval arithmetic is successfully used in scientific fields such as uncertainty theory, or fuzzy systems, to determine the uncertain data and modeling of uncertain systems. So, diverse applications of IVFSs for solving real-life problems involving, for example, pattern recognition, medical diagnosis, or image thresholding were successfully proposed.

The discussed problem is not limited to medical data but equally applies to industry, business, etc., where we meet a similar need for modeling global phenomena and processes without creating a centralized data set. Specific real problems pose new challenges for scientists related to the problem of data disclosure that can only be met by federated learning, which is a combination of traditional learning methods (supervised and unsupervised), data fusion, and aggregation, granular calculations, collaborative learning, compressing updates, asynchronous communication, natural language processing or respect of unbalanced data, non-id data, and so on.

Thus, the paper is focused on two important aspects of contemporary decision support systems, i.e. Federated Learning and modeling various forms of uncertainty. Specifically, the purpose of our research is to:

1) Propose a federated learning method that can handle uncertain and/or incomplete data (in the form of interval fuzzy sets);
2) Improve the performance of local models with missing data using fusion of third-party models;
3) Propose a method for cross-validating models using data from different clients.

978-1-6654-4407-1/21/$31.00 © 2021 European Union

II. Federated learning

A. Federated Model

The federated learning model was originally proposed by google researchers [8], [9], [10]. Their main idea was to build machine learning models based on datasets that are distributed across multiple devices. (cf. [1], [11] or [2]).

Federated learning (FL) is a learning paradigm seeking to address the problem of data governance and privacy by training algorithms collaboratively without exchanging the data itself [1], [12]. 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. Necessary to develop communication-efficient methods that iteratively send small messages or model updates as part of the training process, as opposed to sending the entire dataset over the network. Thus key aspects to consider are: (i) reducing the total number of communication rounds, or (ii) reducing the size of transmitted messages at each round;
- Heterogeneity. The preserving, computational, and communication capabilities of each device in a federated net may differ due to variability in hardware, network connectivity, and power. Federated learning methods that are developed and analyzed must, therefore: (i) anticipate a low amount of participation, (ii) tolerate heterogeneous hardware, and (iii) be robust to dropped devices in the network. Moreover, statistical Heterogeneity is the challenge;
- Privacy. Mentioned privacy is often a major concern in federated learning applications. Federated learning makes a step towards protecting data generated on each device by sharing model updates, e.g., gradient information, instead of the raw data.

Different types of federated learning are distinguished in the literature depending on the characteristics of the local data used. The most popular ones include:

- Horizontal federated learning, or sample-based federated (see Figure 1) is used in scenarios where datasets have the same attributes in each local dataset but different records (samples);
- Vertical federated learning or feature-based federated learning (see Figure 2) is used when local datasets have records that describe the same samples (identifiers) but have different (partially or completely) different sets of attributes. Vertical Federated Learning is also referred to as Heterogeneous Federated Learning, on account of differing feature sets.

In this paper, we will focus on the horizontal model of federated learning, although as we will show later in this paper, in the case of a model with missing whole attributes, such a model could be considered to have some characteristics of a vertex model. Figure 3 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. The server performs model aggregation, i.e. appropriate aggregation of coefficients. The server then returns the new coefficients to the clients. The process proceeds as follows (see Figure 3):

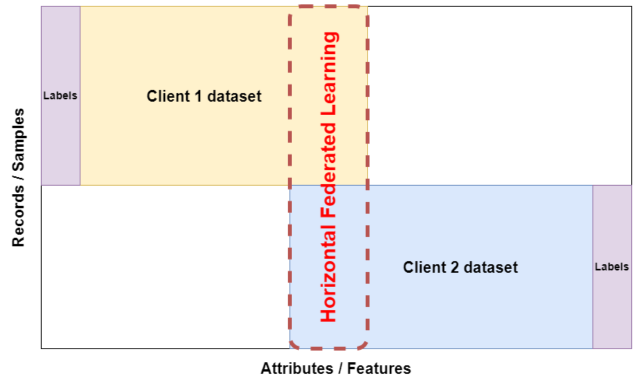

Fig. 1. Horizontal federated learning ([1])

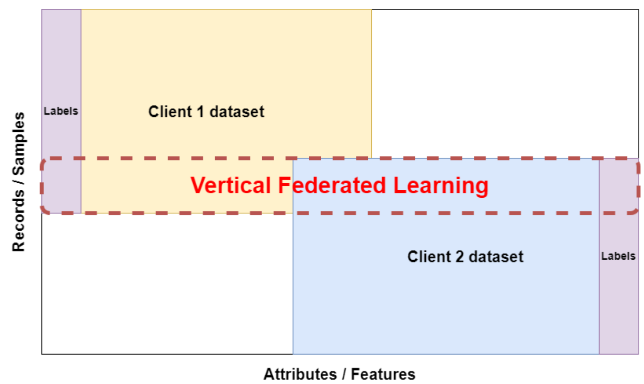

Fig. 2. Vertical Federated Learning (see [1])

1) Each client trains its own model on its local data and passes it to the server;
2) The server aggregates the models;
3) The server returns the new model to the clients;
4) Local models are updated.

The process continues until the obtained quality of local models is high enough and it is not possible to improve them (subsequent iterations do not reduce the error function). The model described in this way is independent of the particular machine learning algorithm chosen and is based only on the replacement of the vector/matrix of parameters describing the model.

In many areas of life, e.g. in industry, medicine, or economics, a problem arises when an organization does not have a sufficiently large set of data to construct a decision / diagnostic system of appropriate quality. In this case, data from different organizations must be used, which is related to the problem of data sharing. To overcome these problems, federated learning

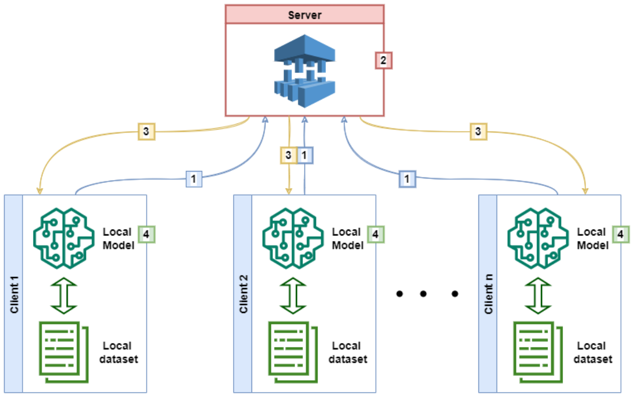

Fig. 3. Federated model architecture (see [1])

is becoming more and more popular, enabling automatic learning in distributed partners' distributed networks without sharing raw data. Federated Learning (FL) is a particular approach to training Machine Learning (ML) algorithms in a way that means data remains private. In particular, federated learning techniques are designed to train machine learning algorithms on multiple distributed devices or servers, each with its own local and private data.

This collaborative approach contrasts with traditional machine learning techniques, which are centralized in nature by collecting all data samples in one unique dataset before using them. It also differs from parallel computing techniques that have been developed to optimize ML computation on multiple CPUs using a centralized dataset that is broken down into identically distributed computation subsets.

The research aims to improve the efficiency of systems supporting federated learning, guaranteeing the privacy of members, while at the same time resisting attacks and ensuring fairness between members of the system.

In this paper, we will use the following notation. Let us assume we have K clients, each client has its own independent data set $\{y_i, x_{i1}, ...x_{ip}\}$ for $i = 1, ..., n_k$, where n_k is the number of instances in client k, and p 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.

III. Fundamentals of interval-valued theory

Many approaches and theories for investigating and modeling imprecision have been proposed since fuzzy sets were originally introduced by Zadeh [3]. For example, interval-valued fuzzy sets [5], [4] are an effective tool for uncertainty modeling in many practical issues.

A. Interval-valued fuzzy setting

By $L^I = \{[\underline{p}, \overline{p}] : \underline{p}, \overline{p} \in [0, 1], \underline{p} \leqslant \overline{p}\}$ we denoted a family of intervals belonging to the unit interval. If $X \neq \emptyset$, then according to the following papers [4], [5], [13] and [14] we define an **interval-valued fuzzy set** (IVFS) S in X 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 into S. The family of all IVFSs in X we denoted by IVFS(X). We assume, reflect an aspect of applications on a finite set $X = \{x_1, \ldots, x_n\}$. In opposite to fuzzy sets, in the case of the IVFSs, the membership of an element x is not exactly indicated. We only specify an upper and lower bound of the possible membership. This is cause the IVFSs spend so useful for the uncertainty of information. Of course, each fuzzy set S could be treated as the IVFS such that $\underline{S}(x) = \overline{S}(x)\ \forall x \in X$. Thus, FS($X$) $\subset$ IVFS(X), where FS(X) is a family of fuzzy sets on X.

In this section, we concentrate on relations between arbitrary IVFSs $S(x)$ and $T(x)$ for any fixed $x \in X$, so let us assume the following notation
$S(x) = [\underline{S}(x), \overline{S}(x)] = [\underline{s}, \overline{s}]$, $T(x) = [\underline{T}(x), \overline{T}(x)] = [\underline{t}, \overline{t}]$.

The best known and often used partial order in L^I it is

$$[\underline{s}, \overline{s}] \leq_2 [\underline{t}, \overline{t}] \Leftrightarrow \underline{s} \leqslant \underline{t} \quad \text{and} \quad \overline{s} \leqslant \overline{t}. \qquad (1)$$

In real-life problems we need often to be able to compared data with uncertainty, e.g. intervals, use some a linear order. So we must extended the partial order $\leq_2$ to a linear one, $\leq_{Adm}$, called admissible [15], [16].

B. Aggregation functions

Data aggregation is the process of gathering data and presenting it in a summarized format. The data may be gathered from multiple data sources with the intent of combining these data sources into a summary for data analysis. This is a crucial step since the accuracy of insights from data analysis depends heavily on the amount and quality of data used. It is important to gather high-quality accurate data and a large enough amount to create relevant results. An aggregate function takes as input a set, a multiset (bag), or a list from some input domain and outputs an element of an output domain.

The notion of an aggregation function on L^I being a significant concept in numerous applications (e.g. [17], [18] or [19]). What follows is the description of aggregation functions connected with $\leq_2$ and $\leq_{Adm}$. For the input data in the form of interval-valued fuzzy sets, we can define aggregations as follows

Definition 1 ([16], [19], [20])**.** Let $n \in \mathbf{N}$, $n \geqslant 2$. An operation $\mathcal{A} : (L^I)^n \to L^I$ is called an interval-valued (I-V) aggregation function if it is increasing with regard to the order $\leq$ (partial or linear), i.e.

$$\forall x_i, y_i \in L^I \quad x_i \leq y_i \Rightarrow \mathcal{A}(x_1, ..., x_n) \leq \mathcal{A}(y_1, ..., y_n) \quad (2)$$

$$\text{and} \quad \mathcal{A}(\underbrace{[0,0], ..., [0,0]}_{n\times}) = [0,0], \quad \mathcal{A}(\underbrace{[1,1], ..., [1,1]}_{n\times}) = [1,1].$$

The particular case of operation of I-V aggregations is a representable I-V aggregation function with regard to $\leq_2$.

Definition 2 ([21], [22])**.** The I-V aggregation function $\mathcal{A} : (L^I)^n \to L^I$ is coined representable in a situation when there exist aggregation functions $A_1, A_2 : [0,1]^n \to [0,1]$ as follows

$$\mathbf{A}(x_1, ..., x_n) = [A_1(\underline{x}_1, ...\underline{x}_n), A_2(\overline{x}_1, ...\overline{x}_n)]$$

for all $x_1, ..., x_n \in L^I$.

Example 1. The fundamental cases of representable I-V aggregation functions on L^I include two operations $\wedge$ as well as $\vee$ on L^I with $A_1 = A_2 = \min$ and $A_1 = A_2 = \max$, respectively (as for the order $\leq_2$, however not the lexicographical or Xu-Yager order).
The other examples of representable I-V aggregation functions concerning $\leq_2$ are the following:

- the representable arithmetic mean
  $\mathcal{A}_{mean}([\underline{x}, \overline{x}], [\underline{y}, \overline{y}]) = [\frac{\underline{x}+\underline{y}}{2}, \frac{\overline{x}+\overline{y}}{2}]$,
- the α mean
  $\mathcal{A}_\alpha(x, y) = [\alpha\underline{x} + (1-\alpha)\underline{y}, \alpha\overline{x} + (1-\alpha)\overline{y}]$
  is an I-V aggregation function on L^I with regard to the lexicographical and Xu-Yager order [16] for $x, y \in L^I$.

Often discussed in the literature and applied in practice, are OWA operators introduced by Yager in 1988. The concept of OWA has been extended to the interval-valued setting (or more generally, to the type-2 fuzzy sets setting or in the real set), which was called the uncertain OWA operator and which is the next generalization of $\mathcal{A}_{mean}$ aggregation. The constructions may be different from point of view used orders. OWA operators are a particular case of more general aggregation functions called Choquet integrals. In [23] was introduced discrete interval-valued Choquet integrals of interval-valued fuzzy sets based on admissible orders.

In [23], the class of linear orders on L^I is used to extend the definition of OWA operators for interval-valued fuzzy setting in following way.

Definition 3 ([23])**.** Let $\leq$ be an admissible order on L^I, and $w = (w_1, \ldots, w_n) \in [0,1]^n$, with $w_1 + \cdots + w_n = 1$. The interval-valued ordered weighted averaging (OWA) operator (IVOWA) associated with $\leq$ and w is a mapping $IVOWA_{\leq,w} : (L^I)^n \to L^I$, given by

$$IVOWA_{\leq,w}([\underline{x}_1, \overline{x}_1], \ldots, [\underline{x}_n, \overline{x}_n]) = \sum_{i=1}^{n} w_i \cdot [\underline{x}_{(i)}, \overline{x}_{(i)}],$$

where $[\underline{x}_{(i)}, \overline{x}_{(i)}]$, $i = 1, \ldots, n$, denotes the i-th greatest of the inputs with respect to the order $\leq$ and $w \cdot [\underline{x}, \overline{x}] = [w\underline{x}, w\overline{x}]$, $[\underline{x}_1, \overline{x}_1] + [\underline{x}_2, \overline{x}_2] = [\underline{x}_1 + \underline{x}_2, \overline{x}_1 + \overline{x}_2]$.

Because $IVOWA_{\leq,w}$ is not an aggregation function with respect to $\leq_2$ ([23]), this is the way in which we prefer to use the linear order to definition.

If $L^I = [0,1]$ or R, then we call IVOWA operator by OWA operator with natural order.

C. Moore's calculus

Interval arithmetic was deemed as 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 [24], [25]. In Moore arithmetic basic operations on intervals $X = [\underline{x}, \overline{x}]$ and $Y = [\underline{y}, \overline{y}]$ are realized by formulas

- 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 such as the excess width effect problem so the alternative for Moore arithmetic we may used multidimensional interval arithmetic. The idea of multidimensional arithmetic was developed by A. Piegat [26], where given value x from interval $X = [\underline{x}, \overline{x}]$ is described using variable γ_x, $\gamma_x \in [0,1]$, as shown:

$$Rep_\gamma(x) = \underline{x} + \gamma_x(\overline{x} - \underline{x}). \tag{3}$$

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 gives the possibility to obtain any value between the left border $\underline{x}$ and right border $\overline{x}$ of interval X.

IV. A FEDERATED LEARNING MODEL WITH UNCERTAINTY

In line with the main challenges in FL we propose a federated model, where each client has its own independent data set $\{Y_i, x_{i1}, ...x_{ip}\}$ and $x_{ip} \in L^I$, $Y_i \in \{0,1\}$ for $i = 1, ..., n$, n is the number of instances, and p 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 $k = 1, ...p$.

The new learning model consists of the following steps (see Figure 4):

1) Each client trains its own model on its local data and passes it to the server;
2) Server aggregates the models;
3) Server returns the new model to the clients;
4) Local models are updated;
5) For initial validation, new models are distributed to other clients on the server where they are tested and model

quality metrics are returned to the client in question. The client decides whether to continue updating its model for better quality.

The process continues until the obtained quality of local models is high enough and it is not possible to improve them (subsequent iterations do not reduce the error function).

As can be seen, the federated learning scheme has been extended to include a cross-validation step for local models. This allows the use of foreign data sets to make decisions in subsequent iterations of the model without having direct access to such data and avoiding model overfitting.

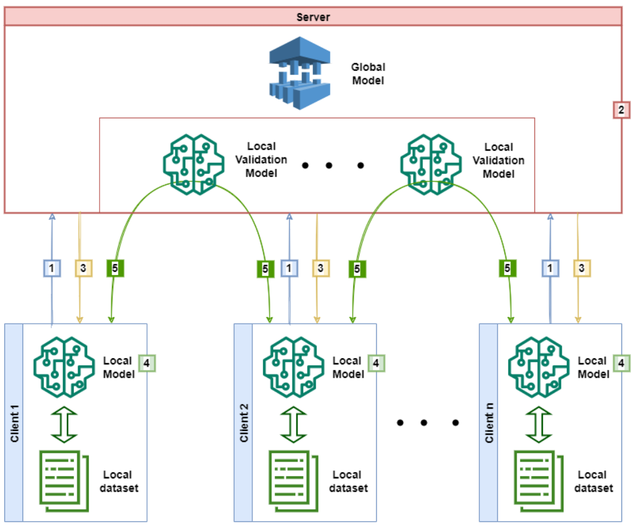

Fig. 4. Proposed federated model

In our model, we allow the data to be in interval form and for there to be gaps in the data. We assume that the data is normalized and the missing data are presented in the form of intervals $[0, 1]$.

A. Logistic regression for interval data

As mentioned in an earlier section, the federated learning scheme thus constructed is independent of the choice of a particular machine learning model. In the first stage of the research, we chose one of the simplest models, which is logistic regression with stochastic gradient descent. For the experiment, we modify it to operate on interval data.

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:

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

for $\gamma \in [0, 1]$ and $f : L^I \to R$.
From this step, in each single iteration, we switch from the interval calculus to the real model using the defined Rep function in (3). Which allows us to operate on data in the form of interval-valued fuzzy sets, while receiving the model in the form of a vector of real numbers.

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

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

where Y_i - actual output value,

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

$$\beta_j = \beta_j + \alpha \cdot \bigtriangledown_{\beta_j} \mathcal{L}(y_i) \cdot x_{ij},$$
$$\beta_0 = \beta_0 + \alpha \cdot \bigtriangledown_{\beta_0} \mathcal{L}(y_i),$$

α is learning coefficient and $\bigtriangledown$ is gradient for $i = 1, .., n_k,\ j = 1, .., p$.

V. Experiment and results

The effectiveness of the proposed federation learning model with uncertainty was tested on medical diagnostic data.

A. Structure of dataset

The dataset is a Wisconsin (diagnostic) breast cancer dataset. This is one of the popular datasets from UCI Machine Learning Repository [27]. Data contains 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, the standard deviation and the 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:

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

after earlier fuzzyfing both values: "mean-standard deviation" and "mean+standard deviation", 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, the 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 data sets of a group of clients (two in this case), the data were randomly divided into two groups with the decision balance behavior. Then, the data in each client is then randomly split into a training set and a test set at a ratio of 90% to 10%.

B. Validation problem

In a federated model, we want to achieve the best possible global model, i.e., one that achieves high decision performance across all clients. Therefore, models should be tested not only on the local data of a given client, but also using the data of other clients (however without direct access to them). Therefore, our proposed federated learning model provides this exchange of model quality information.

Validation is carried out in two stages: during each local learning phase and also after models aggregation. This allows each client to decide whether to use the new incoming model and whether their local model is also effective on foreign data. Such a scheme allows local models to cross-validate effectively. The parameters of the learned model are distributed to other clients across the server where they are tested, and model quality indicators are returned to the client. Finally, the client decides whether to update its model strive for the highest quality global model.

The use of other clients' data (without direct access to this data) in the process of validating the effectiveness of a given model is a proposed new aspect of our federated learning model.

C. Experimental results of different scenarios of real problems

We checked our model in various real-life scenarios, dealing with uncertain and missing data, and compared it with the crisp model (benchmark). During the research of the cases 2)–4) we assumed $\epsilon_i = 0$, $\alpha = 0.01$ and $\gamma = 0.5$ (optimal results) of the algorithm described in Section IV with 100 learning epochs and in cases 3) and 4) 100 aggregation cycles.

1) Benchmark model: As a benchmark model, we chose a situation in which the data is complete and without 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, a standard logistic regression model without a federated learning model was used.

Reference performance of Benchmark model are presented on Table I

	ACC	SENS	SPEC	PREC
Complete data	0,965	0,972	0,935	0,965

TABLE I
PERFORMANCE OF BENCHMARK MODEL

2) Baseline model: As a baseline, we decided to use a situation in which both clients have uncertain interval data with no missing values (complete uncertain data).

That is, a complete interval-valued dataset (no partitioned for clients) without data gaps (performance based on 10 fold cross validation) see Table II.

Dataset	ACC	SENS	SPEC	PREC
Client 1	0,976	0,977	0,976	0,977
Client 2	0,927	0,982	0,849	0,905

TABLE II
PERFORMANCE OF BASELINE MODEL

3) Scenario 1 with missing whole attribute in the data: These types of missing data are very common in medical data and may be due, for example, to the medical unit not having the equipment to perform the measurements or not having access to a particular test (e.g. a blood marker). In our experiments, we assumed that we would perform the calculation in the absence of one attribute in one client (in our case client 2). We want to check whether use of the FL model will allow us to improve the quality of the local model using the information coming from the second model which has complete data.

In Table III and in Table IV performance measures for client 1 (with complete data) and client 2 (with missing 1 attribute), respectively, are presented.

	ACC	SENS	SPEC	PREC
Complete data	0,976	0,977	0,976	0,977

TABLE III
PERFORMANCE OF CLIENT 1 LOCAL MODEL (SCENARIO 1)

Missed attribute	ACC	SENS	SPEC	PREC
1	0,824	0,953	0,690	0,759
2	0,824	0,930	0,714	0,769
3	0,824	0,953	0,690	0,759
4	0,824	0,953	0,690	0,759
5	0,824	0,953	0,690	0,759
6	0,824	0,930	0,714	0,769
7	0,847	0,953	0,738	0,788
8	0,835	0,953	0,714	0,774
9	0,835	0,953	0,714	0,774
10	0,824	0,930	0,714	0,769

TABLE IV
PERFORMANCE OF CLIENT 2 LOCAL MODEL (SCENARIO 1)

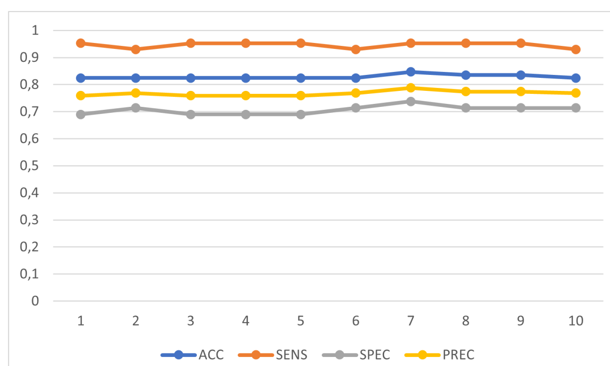

Fig. 5. Performance of client 2 local model (Scenario 1)

4) Scenario 2 with missing random values in different attributes: This scenario reflects the situation when some degree of random data is missing, potentially in all attributes. In real-world conditions, this may result in measurement equipment malfunctions, improper testing, or human error. In our experiment, we simulated situations where we have random missing data distributed in up to five attributes per record and different levels of missing data: 10%-50% of

Missed attribute	ACC	SENS	SPEC	PREC
1	0,895	0,954	0,833	0,862
2	0,895	0,954	0,833	0,862
3	0,900	0,954	0,845	0,873
4	0,895	0,954	0,833	0,862
5	0,894	0,954	0,833	0,866
6	0,894	0,954	0,833	0,866
7	0,900	0,965	0,833	0,865
8	0,906	0,965	0,845	0,876
9	0,900	0,965	0,833	0,868
10	0,900	0,954	0,845	0,8739

TABLE V
PERFORMANCE OF FL AGGREGATED MODEL (SCENARIO 1)

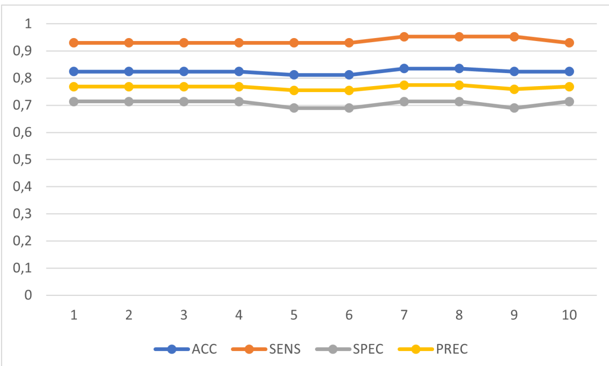

Fig. 6. Performance of aggregated model on client 2 test data (Scenario 1)

values). Also in this case, data was only deleted in one client, the other had full data

As in the previous scenario first client have full data (in Table VI result of it's local model is presented) and second client have prepared data with different percentage of missing data (10% – 50%), in Table VII and Figure 7 are presented results for it's local model.

% of missing data	ACC	SENS	SPEC	PREC
Complete data	0,976	0,977	0,976	0,977

TABLE VI
PERFORMANCE OF CLIENT 1 LOCAL MODEL (SCENARIO 2)

% of missing data	ACC	SENS	SPEC	PREC
10	0,871	0,953	0,786	0,820
20	0,776	0,837	0,714	0,750
30	0,788	0,744	0,833	0,821
40	0,753	0,953	0,548	0,683
50	0,729	0,884	0,571	0,679

TABLE VII
PERFORMANCE OF CLIENT 2 LOCAL MODEL (SCENARIO 2)

Results for federated learning model (using standard mean as aggregation) are presented in Table VIII and Figure 8. Calculations was made on Client 2 test data.

5) Conclusions: The baseline uncertified model is very close in performence to the benchmark model which shows that the methodology used performs very well for full data (without data gaps).

In scenario one, it is noted that the local model is sensitive to a small degree to the absence of a single attribute. This may of course be due to the high correlation of some attributes and the distribution of samples in the test data. In this case, the federated model slightly improves the classification quality.

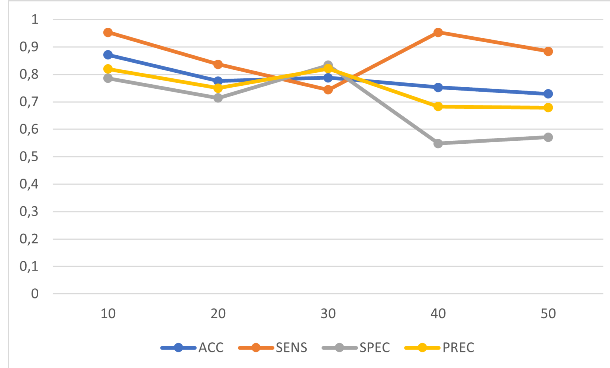

Fig. 7. Performance of client 2 local model (Scenario 2)

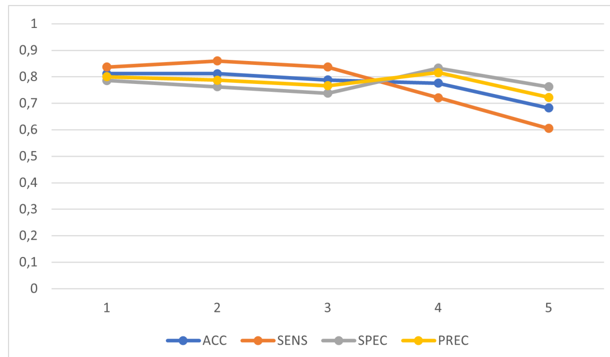

Fig. 8. Performance of FL model on client 2 test data (Scenario 2)

% of missing data	ACC	SENS	SPEC	PREC
10	0,894	0,907	0,881	0,889
20	0,883	0,907	0,857	0,870
30	0,877	0,919	0,834	0,851
40	0,865	0,849	0,881	0,875
50	0,824	0,803	0,846	0,829

TABLE VIII
PERFORMANCE OF FL AGREGATED MODEL (SCENARIO 2)

The most interesting results were obtained for the second scenario. The client 2 local model performed very poorly when the amount of missing data increased. The use of a federated model allowed for a significant improvement in classification quality using knowledge from another client.

The use of the federated learning model in both scenarios allowed the classification quality scores to approach the baseline model despite the loss of a large portion of the client's data in some cases.

VI. SUMMARY AND FUTURE WORKS

This paper presents the first attempt to construct and analyze the effectiveness of a federated learning model taking into account uncertainty. This problem still needs a lot of research and computational problems to be solved, but preliminary results are very promising. We can mention here the following conclusions:

1) Very good results were shown for binary medical decisions using logistic regression. Improved/stabilized classification quality of the model with missing data using knowledge propagated from other models.
2) Very good comparative results, for federated learning for a multitude of smaller sets, relative to the results for learning a uniform set without splitting.
3) Conducted a deeper analysis of the ability to detect inconsistent (faulty, contradictory) data-sets, based on an analysis of the adaptability of the local model relative to the global model (how often the local model is updated based on an assessment of classification errors, a measure of the frequency of change can be a measure of the quality of fit to the global model).
4) Future research should be done on the impact of different learning methods on different types of data and related specific problems, e.g., applying classification or combining classical machine learning with soft computational methods such as similarity measures, fuzzy clustering, approximate inference, etc.

   In addition, other aggregations such as OWA or IVOWA for the interval case will be considered in aggregating model parameters.

References

[1] 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.

[2] 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.

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

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

[5] 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).

[6] K. T. Atanassov, "Intuitionistic fuzzy sets," *Fuzzy Sets and Systems*, vol. 20, no. 1, pp. 87–96, 1986.

[7] H. Bustince, E. Barrenechea, M. Pagola, J. Fernandez, Z. Xu, B. Bedregal, J. Montero, H. Hagras, F. Herrera, and B. De Baets, "A historical account of types of fuzzy sets and their relationships," *IEEE Transactions on Fuzzy Systems*, vol. 24, no. 1, pp. 179–194, 2016.

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

[9] 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.

[10] 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.

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

[12] 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.

[13] I. B. Turksen, "Interval valued fuzzy sets based on normal forms," *Fuzzy sets and systems*, vol. 20, no. 2, pp. 191–210, 1986.

[14] M. B. Gorzałczany, "A method of inference in approximate reasoning based on interval-valued fuzzy sets," *Fuzzy sets and systems*, vol. 21, no. 1, pp. 1–17, 1987.

[15] H. Bustince, J. Fernandez, A. Kolesárová, and R. Mesiar, "Generation of linear orders for intervals by means of aggregation functions," *Fuzzy Sets and Systems*, vol. 220, pp. 69–77, 2013.

[16] 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.

[17] 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.

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

[19] 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.

[20] 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.

[21] G. Deschrijver and E. E. Kerre, "Implicators based on binary aggregation operators in interval-valued fuzzy set theory," *Fuzzy Sets and Systems*, vol. 153, no. 2, pp. 229–248, 2005.

[22] G. Deschrijver, "Quasi-arithmetic means and owa functions in interval-valued and atanassov's intuitionistic fuzzy set theory," in *Proceedings of the 7th conference of the European Society for Fuzzy Logic and Technology*. Atlantis Press, 2011, pp. 506–513.

[23] H. Bustince, M. Galar, B. Bedregal, A. Kolesárová, and R. Mesiar, "A new approach to interval-valued choquet integrals and the problem of ordering in interval-valued fuzzy sets applications," *IEEE Transactions on Fuzzy Systems*, vol. 21, no. 6, pp. 1150–1162, 2013.

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

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

[26] 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 and et al., Eds. Warsaw: IBS PAN - SRI PAS, 2013, p. 137–151.

[27] D. Dua and C. Graff, "UCI machine learning repository," 2017. [Online]. Available: http://archive.ics.uci.edu/ml