Refining Uncertainty Management in Machine Learning: An Interval-Valued Fuzzy Set Approach to Logistic Regression

Jarosław Szkoła[1], Barbara Pękala[1,2], and Krzysztof Dyczkowski[3]

[1] University of Rzeszów, Poland jszkola,bpekala@ur.edu.pl
[2] University of Information Technology and Management, Rzeszów, Poland
[3] Adam Mickiewicz University, Poznań, Poland chris@amu.edu.pl

Abstract. The article addresses the ubiquitous challenge of uncertainty in decision-making, with a particular focus on medical decision-making, through the innovative application of logistic regression enhanced by interval-valued fuzzy set theory. Traditional logistic regression relies on a linear combination of variables and a uniform set of regression coefficients, which can inaccurately represent the variability and uncertainty inherent in real-world data. Our proposed methodology differs in that it incorporates weights with interval values into the logistic regression model, allowing for a more nuanced and flexible representation of the data. This approach allows the model to adjust the weights independently in terms of values, offering a fit to interval data and improving the precision of predictions. By developing a specialized algorithm to calculate weighted coefficients adjusted to specific inputs or attributes, we demonstrate the practical effectiveness of our method in dealing with uncertainty. Experimental results highlight the potential of interval-valued fuzzy sets in improving machine learning techniques and enhancing the accuracy of decision-making models in complex, uncertain environments.

1 Introduction and motivation

Our research direction has been inspired by the ubiquitous issue of uncertainty in decision-making, especially in daily medical practice. Uncertainty or imprecision, as it is also known, finds a robust representation through interval-based fuzzy sets. This method has been validated in many applications, especially when adopting an epistemic approach (see [7]) in which the intervals cover a single desired value. The source of uncertainty is often a lack of precise knowledge. This is particularly true in the medical field, where descriptions can be inherently vague or ambiguous. Such inaccuracy can result from a variety of factors, including the type of medical equipment used or the subjective interpretation of the healthcare professional.

The current situation demands the adoption of non-traditional methods for data modeling and inference that can adequately account for imprecision. Although extensive research has been conducted in this area (e.g. [3, 1]), there

remains a lack of efficient methods to manage such imprecision within the medical and industrial sectors. In response, this paper proposes the utilization of logistic regression within a machine learning framework, combined with interval fuzzy set theory, as a means to effectively represent uncertainty and facilitate decision-making in uncertain contexts. This is particularly relevant for the representation and processing of parameters that have been learned through the model. Machine learning is an ever-evolving field that continues to uncover new practical applications, including in the domain of medicine, where the integration of artificial intelligence offers promising new avenues for exploration and implementation (see [18]).

At its core, machine learning is a technique for constructing and updating model weights, a process essential to the effectiveness of these models. This article introduces a novel method for developing and modifying weights in a logistic regression model by using interval representations. This approach specifically addresses data uncertainty, a critical factor in predictive modeling. Logistic regression is particularly effective in estimating the probability of diseases such as breast cancer, diabetes and ischemic heart disease by analyzing patient characteristics, including age, gender, body mass index and various blood test results (see for example [12, 11]). The model has proven effective in predicting the risk of common chronic diseases based on simple clinical indicators, as detailed in previous studies. Such predictive performance underscores the potential of logistic regression compared to other machine learning models, highlighting its utility in healthcare applications (see [19] or [15])

The basic learning algorithm traditionally updates scalar weights based on the difference between predicted model values and actual results for individual data points. Our innovative proposal changes this approach by introducing a new coding technique for the weighting factor update mechanism. The technique uses a structure of intervals and distinct values represented by these intervals. Within this framework, a single weight can correspond to multiple scalar values, each associated with input values that fall within a designated interval. We intend to introduce an algorithm designed to calculate weighted coefficients adopted to specific inputs or attributes. The driving force behind our research was to improve the conversion of data into learning coefficients, while taking into account the inherent uncertainty in the data. This is intended to improve the adaptability and accuracy of machine learning models by providing a more nuanced representation of real-world data variability. In particular, we consider the use of interval calculus to represent uncertainty in two aspects:

1. representation of imprecise data;
2. enhancing the flexibility of the sigmoidal model in logistic regression (model parameters in the interval form, see Fig. 2).

In traditional models, such as logistic regression, parameters are typically represented as scalar values, as demonstrated in Fig. 1. Our approach challenges this convention by advocating for the use of intervals, which allows for a more nuanced and flexible representation of model parameters. This innovation is designed to improve the model's performance by enabling it to more effectively

process and interpret data that is characterized by uncertainty. Our methodol-

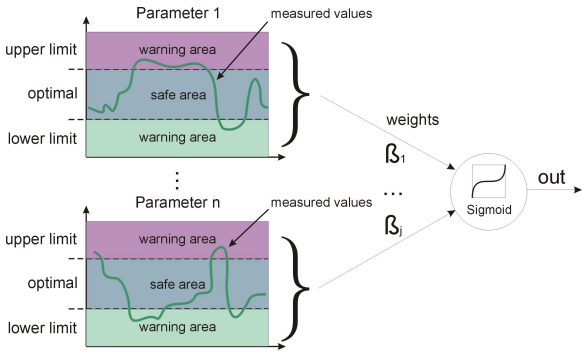

Fig. 1. Classical logistic regression

ogy is centered on the principle of constructing parameter models in a deliberate manner, grounded on the specific ranges of data that necessitate representation through subintervals. This leads to the conceptualization of representing these ranges as a sequence of subintervals, akin to an array, as illustrated in Fig. 2. This approach ensures that model parameters are not randomly assigned but are instead systematically derived from the inherent data structure, offering a more accurate and nuanced understanding of the data's underlying patterns and uncertainties.

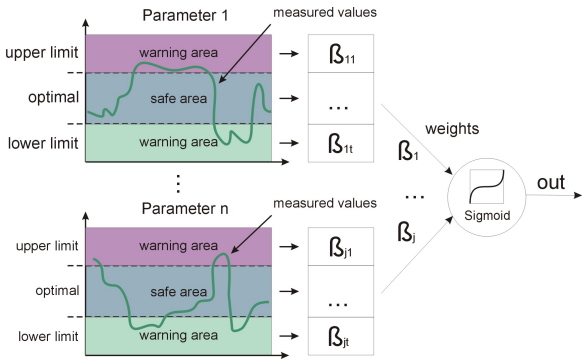

Fig. 2. Multi-interval-weights logistic regression

2 Uncertainty, fundamentals of interval-valued theory

Since the introduction of fuzzy sets by Zadeh [26], many approaches and theories to study and model uncertainty have been proposed. Especially, interval-valued fuzzy sets [23, 27] are a strong tool for uncertainty modeling in many practical issues.

2.1. Interval-valued fuzzy setting

Needed in our approach will be the notation of intervals representing uncertainty. Thus a family of intervals belonging to the unit interval is $L^I = \{[\underline{p}, \overline{p}] : \ \underline{p}, \overline{p} \in [0,1], \ \underline{p} \leq \overline{p}\}$. We also have to recall the definition of **interval-valued fuzzy set** (IVFS) [27, 23, 24, 10], S in X as a mapping $S : X \to L^I$ such that for each $x \in X$, $X \neq \emptyset$, and $S(x) = [\underline{S}(x), \overline{S}(x)]$ 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, because of the application aspect that $X = \{x_1, \ldots, x_n\}$ is a finite set. The IVFSs spend so useful for the uncertainty of information because in opposite to fuzzy sets, in IVFSs the membership of an element x is not exactly indicated. Specified an upper and lower bound of the possible membership. For any fixed $x \in X$ we assume $S(x) = [\underline{S}(x), \overline{S}(x)] = [\underline{s}, \overline{s}]$.

In L^I we may use the best-known and often-used partial order

$$[\underline{s}, \overline{s}] \leq_2 [\underline{t}, \overline{t}] \Leftrightarrow \underline{s} \leq \underline{t} \text{ and } \overline{s} \leq \overline{t}. \tag{1}$$

In practical scenarios, it is frequently necessary to compare data that is represented as intervals, subsequently requiring a mechanism to establish some form of linear order. This need arises because we aim to transcend the limitations posed by incomparability, often encountered in interval data. By extending the partial order, denoted as $\leq_2$, to a linear and admissible order, we enable a more straightforward comparison of interval-represented data. This adjustment facilitates the alignment and analysis of such data within our proposed models, enhancing their applicability and effectiveness in handling real-world problems where data uncertainty is a significant factor [4, 28].

2.2. Aggregation process

The integration of uncertain operators and data plays a crucial role in accurately modeling reality through mathematics. In particular, the concept of an aggregation function within the context of L^I proves to be essential across a wide array of applications (e.g., [9, 20] or [2, 21]). Such functions are pivotal for compiling high-quality, precise summaries of data, which in turn, facilitate the generation of reliable outcomes in decision-making scenarios. Aggregation, in essence, is the methodology employed to synthesize and represent data cohesively. When dealing with input data represented as interval-valued fuzzy sets, it becomes possible to define aggregation processes that adhere to specific, adequate orders, such as $\leq_2$ or $\leq_{Adm}$. This allows for a structured and effective way to combine and interpret uncertain data, enhancing the precision and relevance of the conclusions drawn from such analyses.

Definition 1 ([13]). *Let $n \in \mathbf{N}$, $n \geq 2$. 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.* $\forall x_i, y_i \in L^I \quad x_i \leq y_i \Rightarrow \mathcal{A}(x_1, ..., x_n) \leq \mathcal{A}(y_1, ..., y_n)$ *and* $\mathcal{A}(\underbrace{[0,0], ..., [0,0]}_{n}) = [0,0]$, $\mathcal{A}(\underbrace{[1,1], ..., [1,1]}_{n}) = [1,1]$.

Since their introduction by Yager in 1988, Ordered Weighted Averaging (OWA) operators have become a widely discussed and applied concept in various practical applications. OWA represents a further generalization of the arithmetic mean, enabling aggregation across different orders. Significantly, OWA operators are a subset of a broader category of aggregation functions known as Choquet integrals. This conceptual expansion allows for the adaptation of OWA operators to interval-valued fuzzy settings by extending their definition to accommodate linear or admissible orders on L^I. This adaptation highlights the operators' flexibility and their capacity to provide nuanced aggregations within the realm of interval-valued fuzzy sets theory, offering a powerful tool for handling complex data in a more structured and interpretable manner.

In the work [5], the authors have expanded the definition of Ordered Weighted Averaging (OWA) operators to encompass linear or admissible orders within the context of L^I:

Definition 2 ([5]). *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* $w \cdot [\underline{x}, \overline{x}] = [w\underline{x}, w\overline{x}]$*,* $[\underline{x}_1, \overline{x}_1] + [\underline{x}_2, \overline{x}_2] = [\underline{x}_1 + \underline{x}_2, \overline{x}_1 + \overline{x}_2]$.

Given that $IVOWA_{\leq,w}$ does not serve as an aggregation function under the order $\leq_2$ as noted in [5], we opt to employ a linear order when defining uncertain OWA operators.

2.3. Moore's calculus

In the field of interval arithmetic, Moore's arithmetic is recognized as the most universally accepted and frequently applied method, as documented in references [16, 17]. Moore's arithmetic outlines basic operations on intervals, such as addition, subtraction, and multiplication, with specific formulas for each operation. For two intervals $X = [\underline{x}, \overline{x}]$ and $Y = [\underline{y}, \overline{y}]$, the operations are defined as follows: Addition: $[\underline{x}, \overline{x}] + [\underline{y}, \overline{y}] = [\underline{x} + \underline{y}, \overline{x} + \overline{y}]$; Subtraction: $[\underline{x}, \overline{x}] - [\underline{y}, \overline{y}] = [\underline{x} - \overline{y}, \overline{x} - \underline{y}]$; Scalar multiplication for a positive real number $a \in R^+$ is:$a * [\underline{x}, \overline{x}] = [a\underline{x}, a\overline{x}]$ and for a negative real number $a \in R^-$, it is: $a * [\underline{x}, \overline{x}] = [a\overline{x}, a\underline{x}]$. The product of two intervals is calculated as: $[\underline{x}, \overline{x}] * [\underline{y}, \overline{y}] = [\min(\underline{x} * \underline{y}, \overline{x} * \overline{y}, \underline{x} * \overline{y}, \overline{x} * \underline{y}), \max(\underline{x} * \underline{y}, \overline{x} * \overline{y}, \underline{x} * \overline{y}, \overline{x} * \underline{y})]$

These operations apply to real numbers within the intervals where $\underline{x} \leq \overline{x}$ and $\underline{y} \leq \overline{y}$, providing a solid foundation for arithmetic operations involving intervals.

Some limitations and drawbacks have been identified in Moore's interval arithmetic, notably the issue of excess width in the results. An alternative approach, known as multidimensional interval arithmetic, has been proposed to

overcome these challenges. This concept, introduced by A. Piegat [22], offers a novel way to represent values within an interval $X = [\underline{x}, \overline{x}]$. Specifically, any value x from interval X is characterized using a variable γ_x, where $\gamma_x \in [0, 1]$, as shown in the following representation formula:

$$Rep_\gamma(x) = \underline{x} + \gamma_x(\overline{x} - \underline{x}). \tag{2}$$

With this approach, the interval $X = [\underline{x}, \overline{x}]$ is described more dynamically as:

$$X = \{Rep_\gamma(x) : Rep_\gamma(x) = \underline{x} + \gamma(\overline{x} - \underline{x}), \gamma \in [0, 1]\}. \tag{3}$$

The variable γ thus allows for the retrieval of any value between the lower bound $\underline{x}$ and the upper bound $\overline{x}$ of interval X, offering a more nuanced and flexible method for managing interval data.

3 Structure of the dataset

The effectiveness of the proposed learning model, which incorporates uncertainty (as described in Section 5), was evaluated using medical diagnostic data. Specifically, we applied our methodology to the Wisconsin (diagnostic) breast cancer dataset available from the UCI Machine Learning Repository [6]. This dataset is derived from digitized images of fine needle aspirates (FNA) of breast masses, with features characterizing the cell nuclei present in the images.

For each cell nucleus, ten real-valued features are calculated. For each feature, both the standard deviation and the mean value of the measurements for the patient are considered. From these values, an interval is constructed to represent each feature accurately:
$[\mathbf{mean - standard\ deviation}, \mathbf{mean + standard\ deviation}]$.
This interval is then normalized/fuzzified for each value: "mean - standard deviation" and "mean + standard deviation," providing a comprehensive representation of the data in terms of intervals.

The decision outcome within this dataset indicates the diagnosis: malignant (denoted by “0”) or benign (denoted by “1”). The dataset encompasses a total of 569 patients/objects, which includes 212 malignant cases and 357 benign cases. Given the dichotomous nature of the decision values (0 and 1), logistic regression emerged as the optimal choice for predicting the diagnosis.

4 Proposed new methodology

In the initial phase of our research, we selected logistic regression with stochastic gradient descent as our foundational model due to its simplicity and widespread applicability. Our experimental modification involves adapting this model to handle interval data effectively. We utilize the following notation for our dataset $\{Y_i, x_{i1}, ...x_{ip}\}$ and $x_{ip} \in L^I$, $Y_i \in \{0, 1\}$ for $i = 1, ..., n$, n representing the total number of instances and p the number of attributes.

The model was trained on a dataset consisting of n_k observations through a predefined number of internal iterations. The outcome of this training phase is encapsulated in a result vector comprising the trained parameters β and ϵ, expressed as:

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

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

Logistic regression finds utility across a diverse range of domains, including but not limited to machine learning, various medical fields, and the social sciences. It is particularly effective for predicting the likelihood of disease development, such as diabetes or ischemic heart disease, by analyzing patient-specific characteristics like age, sex, body mass index, and blood test results (see [12]).

The fundamental principle of logistic regression involves employing a linear combination of explanatory variables alongside a set of model-specific regression coefficients, which remain consistent across all instances. For a given data point i, the linear predictor function y_i incorporates parameters $\beta_0, \ldots, \beta_m$, with each β coefficient reflecting the degree of influence exerted by corresponding explanatory variables on the outcome.

The core of the learning algorithm lies in the adjustment of scalar β values, which is based on the discrepancy between the predicted outcomes and the actual observations for each data entry.

Our approach introduces a novel encoding method and an updating mechanism for the weighting factors. Unlike traditional methods where weighting factors are assigned singular real values post-training, our method represents these factors as a sequence of intervals and distinct values, segmented into compartments. This adjustment allows for a more nuanced interpretation and application of the regression coefficients, enhancing the model's ability to capture and reflect the complexity of the data.

In this novel approach, a single weighting factor is capable of possessing multiple scalar values, each of which is precisely assigned to input values falling within a specific range.

Procedure based on a new approach

The proposed procedure consists of the following three main steps describing the Construction of the initial parameters β - sequence composed of subintervals for training data, optimization of the width of the intervals method, and Learning procedure, which collectively form the backbone of our approach.
Description of the new procedure:

I. Construction of the initial values of the model parameters. For each attribute, we build parameters $\beta_j = \{\beta_{jt} : v_t \to \beta_{jt},\ \beta_{jt} \in L_I \ and \ v_t \in R\}$ as a sequence of intervals based on the input data (see below step 1 and step 2 (sequences of intervals with optimization of their width)). Therefore, the jth parameter

$$\beta_j = \{\beta_{jt\uparrow_{v_t}}\} = \{[\eta_1, \eta_2)_{\uparrow_{v_1}}, ..., [\eta_z, \eta_{z+1}]_{\uparrow_{v_z}}\},$$

$1 \leq j \leq p$ for p attributes and $1 \leq t \leq z$. Note that $\beta_0 = [0, 1]$: with $v_0 = 1$.

II. Learning process based on a new form: a sequence of intervals with corresponding representatives. During training (in individual iterations - Step 3), individual representatives are modified based on the error value determined based on the distance of the exact value Y_i from the approximate y_i of the ith object. In particular, in each iteration, we select for updating that v_t, $1 \leq t \leq z$ for which the intervals associated with them have a part in common with the interval corresponding to the input data of a given attribute (selection method described in Step 3.3.) and we use in the process of updating Moore's operation. Returning updated values in new ranges β_0' and β_j' for $1 \leq j \leq p$.

In particular, our procedure includes the following key steps:

Step 1 Construction of the initial value of model parameters β - sequence composed of subintervals for training data.
In our approach, a critical aspect is the determination of the number and width of subintervals, which significantly enhances the performance of traditional linear regression. Various tactics can be employed, such as selecting the number and width of intervals randomly, setting a fixed number of intervals of a specified width, or initializing the segmentation based on a preliminary analysis of the training data's structure. For the first two strategies, finding the most effective subdivision into subintervals requires experimental validation across different datasets. With the latter method, the process of dividing into a specific number of subintervals and determining their width is automated based on the initial data analysis. It's important to note that this initial division of weights into sub-ranges is just the beginning; there's room for further optimization of these sub-ranges, especially if the initial ranges are deemed too narrow. This optimization level is controlled by a threshold parameter set within the (0,1) range. When this parameter is set to 1, no optimization takes place.
The procedure of automatically building adequate subintervals based on data we observe in Fig. 3.

Step 2 Optional optimization of the width of the intervals.
We use the following algorithm to eliminate of to narrow subintervals of each β at the assumed significance threshold ψ (we choose optimal 0,1).
In Algorithm 1 - Modification of Parameters, we propose to use the following weight aggregation A :

$$A(a,b) = \begin{cases} \frac{1}{2}(\frac{w_a}{w_b}a + b), \text{ if } \ w_b \geq w_a, \\ \frac{1}{2}(a + \frac{w_b}{w_a}b), \text{ otherwise,} \end{cases} \tag{4}$$

where w_a, w_b are weights of intervals with assumed values a and b, respectively.
Thus, formally we present j-th parameter in the following way:

$$\beta_j = \{\beta_{jt_{\uparrow_{v_j}}}\} = \{[\eta_1, \eta_2)_{\uparrow_{v_1}}, ..., [\eta_z, \eta_{z+1}]_{\uparrow_{v_z}}\},$$

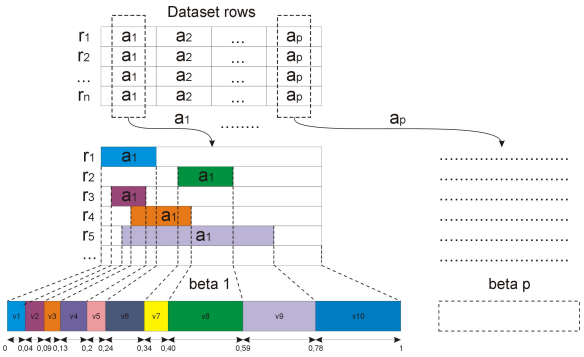

Fig. 3. Parameters initialization of β

Algorithm 1: Algorithm Modification of Parameters

input : n element data set instances, p - number of attributes, $\{\beta_j\}_{j=1}^{p}$, $\beta_j = \{[\eta_1, \eta_2], ..., [\eta_z, \eta_{z+1}]$ for p attributes and with assumed values for each interval: v_1,...,v_z, $z \leq n-1$; A satisfy (4).

output: $\{\beta_j\}_{j=1}^{p}$, $\beta_j = \{[\delta_1, \delta_2], ..., [\delta_m, \delta_{m+1}]\}$ for p attributes and with assumed values for each interval: d_1,...,d_m, $m \leq z$ and for all k $\delta_{k+1} - \delta_k \leq \psi$

for $j = 1, ..., p$ **do**
  for *k=1,...,z* **do**
    if $\eta_j - \eta_{j+1} \leq \psi$ **then**
      if $|v_j - v_{j-1}| \leq |v_j - v_{j+1}|$ **then**
        return $d_i \leftarrow A(v_j, v_{j-1}) \leftarrow [\delta_{j-1}, \delta_j]$
      return $d_j \leftarrow A(v_j, v_{j+1}) \leftarrow [\delta_j, \delta_{j+1}]$

$1 \leq j \leq p$ for p attributes and $1 \leq t \leq z$ and we assumed values for each intervals: $v_1, ...v_z$. $\beta_{jt} \in L^I$ and $v_j \in R$, where $v_j \rightarrow \beta_{jt}$. Note that $\beta_0 = [0, 1]$: with $v_0 = 1$.

Step 3 Learning procedure.
One iteration of the 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 defined *Rep* function in (2). 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 the computation of an error (loss function) between the computed value and the actual value we assumed for Y_i - actual output value:

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

3. Finally, we update of learning coefficients in the steps:

$$\beta_{jt_{\uparrow v_j}} = \beta_{jt_{\uparrow v_j}} + \alpha \cdot \bigtriangledown_{\beta_{j\uparrow v_j}} \mathcal{L}(y_i) \cdot x_{ij},$$

$$\beta_{0\uparrow v_0} = \beta_{0\uparrow v_0} + \alpha \cdot \bigtriangledown_{\beta_{0\uparrow v_0}} \mathcal{L}(y_i),$$

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

Individual iterations are associated with the following method of modifying β parameters along with their weight representatives $v_1, ..., v_z$.
A simplified diagram of the learning mechanism, using an assigned value, indexed by interval matching, is graphically present in Fig. 4.

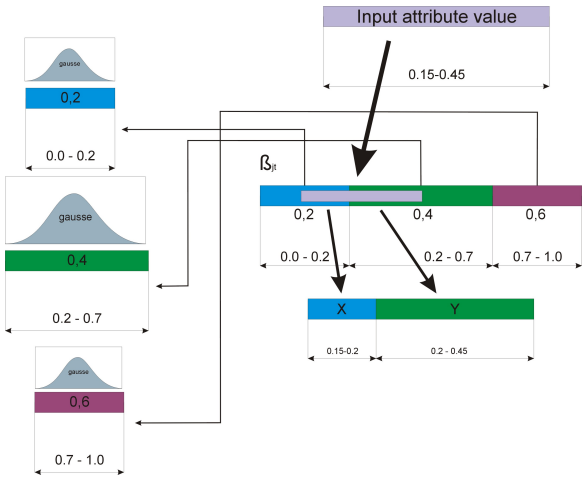

Fig. 4. Change/choose the representative value of intervals

For determining the weighted beta coefficients for excitation from a given input/attribute, we define the following procedure based on the weight aggregation method.
For input value $x_{ij} \in L^I$, $i = 1, ..., p$. If $\emptyset \neq x_{ij} \cap [\eta_k, \eta_{k+l}]$, $k + l \leq z$, then

$$v_j = \sum_{t=k}^{k+l-1} \int_{t+\theta_1}^{t+1-\theta_2} g(t) * v_j,$$

with θ_1, θ_2 as offsets input attribute with respect to some β subintervals (see Fig. 4) and the Gaussian function $g(t)$ described on $[t, t + 1]$

subinterval of β_j such that: $g(t) = \int_t^{t+1} e^{-t^2*\lambda}$ and for λ satisfy $\int_t^{t+1} e^{-t^2*\lambda} = 1$.

Each step of learning is based on each time checking which sub-intervals within β_j a given interval of input values of the considered attribute have in common part, to build a new one based on their representatives and there we also use the presented algorithm to eliminate to narrow subintervals of each β by Alg. 1. Finally, we obtain optimal for each $1 \leq j \leq p$ $\beta_{jm_{\uparrow_{v_j}}}$ and $\beta_{0_{\uparrow_{v_0}}}$ for $1 \leq m \leq z$.

Please note that for the testing process that takes place on 10% of the randomly selected data, we do not use Step 1 of the above learning process.

5 Experimental results and discussion

For the training process, input data is divided into training and testing subsets in a 90% to 10% percent ratio. The learning algorithm is based on iterative learning using a 10-fold cross-validation method. The test subset does not participate in the training process.

We checked our model in various real-life scenarios, dealing with uncertain and real weights (baseline model), and compared it with the Interval weights, i.e., the Model Multi-interval-weights logistic regression. During the research of cases 5.1-.5.2 we assumed $\epsilon_i = 0$, $\alpha = 0.01$, and $\gamma = 0.5$ of the algorithm described in Section 4 with 100 learning epochs.

We compare the application of the proposed new approach "Interval weights" to the classical "Baseline model".

5.1. Baseline model

As a Baseline model with real values of weights, we decided to use a situation in which we have uncertain interval data with no missing values (complete uncertain data), and in the learning procedure, we use real values of weights. We present the effectiveness of the studied algorithm used in the training and tested set (Tab. 1).

Table 1. Classification results for Baseline model on training and test set by accuracy (ACC), sensitivity (SENS), specificity (SPEC), precision (PREC), and F1 measure

	ACC	SENS	SPEC	PREC	F1
training set	0.962	0.979	0.930	0.965	0.972
test set	0.913	0.916	0.909	0.942	0.929

5.2. Interval weights model

As the Interval weights Model, we called the model multi-interval-weights logistic regression, the performance of which we also present for the training and test subsets of data (Table 2).

From Tables 1 and 2, it's evident that there's a notable enhancement in the classification efficiency when utilizing regression combined with our proposed method for updating weighting factors, which are structured as a series of intervals for test datasets. Despite a rise in computational complexity compared to the baseline model, this improvement represents a significant advantage, particularly in applications such as enhancing breast cancer diagnosis accuracy. Additionally, it's important to highlight that this model improvement, facilitated by interval weights, surpasses the efficiency parameters of the baseline model even with fewer epochs of training. Moreover, because the F1 score combines precision and sensitivity using their harmonic mean, our very high F1 score implies simultaneously maximizing both precision and sensitivity and confirms the effectiveness of the proposed method. The study encompassed both the baseline model—that is, a

Table 2. Classification results for interval weights model on training and test set by accuracy, sensitivity, specificity, precision, and F1 measure

	ACC	SENS	SPEC	PREC	F1
training set	0.968	0.980	0.950	0.968	0.974
test set	0.931	0.916	0.954	0.970	0.942

regression-based training model utilizing standard real-valued learning parameters—and a model employing multi-interval-weighted logistic regression, considering various factors. This included examining the effects of varying the number of epochs and different values of the ψ parameter. The ψ parameter serves as a threshold for regulating the width of intervals within the individual β parameters. It was found that optimal results are typically achieved when ψ falls within the range of $[0.1, 0.5]$

In real-world datasets, each attribute signifies a distinct characteristic of the input set, and the correlations between the values of individual records play a crucial role in accurately classifying samples. Most machine learning architectures assign specific weights to each input to signify the importance of an attribute in determining the membership of a sample in a particular decision class. A significant challenge with this method is the assumption that an attribute linearly influences the model's final decision, and a single scalar weighting factor suffices to interpret the values of an attribute correctly. This approach tends to diminish the significance of an attribute which, particularly in medical datasets, can have vastly different implications depending on the measured value. For instance, extreme values of blood pressure, temperature, or heart rate are interpreted differently than average values. Employing linear weighting factors for such data could lead to overemphasizing low-value data or underrepresenting high-value data. The remedy to this issue involves utilizing separate weighting factors for different value ranges, tailored to their actual significance. This strategy enables the precise determination of weighting factors for specific sub-ranges, significantly enhancing the data representation accuracy and thereby crafting a model

that aligns closely with the input data. In conclusion, the innovative learning method employing multi-interval weights in logistic regression not only boosts classification efficiency but also shows promising prospects for further advancements, such as its application in federated learning models.

6 Summary and future works

In this paper, we introduced an efficient algorithm for calculating the weighted beta coefficients based on inputs or attributes. Logistic regression fundamentally relies on a linear combination of explanatory variables along with a consistent set of regression coefficients tailored to the model but identical across all instances.

Looking ahead, we aim to propose methods for weight aggregation alongside the algorithm for determining weighted coefficients in their new form, addressing various practical challenges, notably in federated learning scenarios as discussed in [8]. Our future efforts will focus on enhancing the work in the realm of federated learning. The application of interval-valued logistic regression is particularly suited to the federated learning framework, as it allows local models to adapt to diverse parameter ranges, a flexibility afforded by our proposed method.

It's important to underline that federated learning emerges as a paradigm designed to tackle data governance and privacy issues. It achieves this by enabling collaborative algorithm training without the need to share the data itself, as highlighted in [25] and [14]. This approach not only preserves privacy but also opens new avenues for the application of our interval-valued logistic regression methodology in distributed learning environments.

References

1. Asiain, M.J., Bustince, H., Bedregal, B., Takáč, Z., Baczyński, M., Paternain, D., Dimuro, G.: About the Use of Admissible Order for Defining Implication Operators. In: IPMU'2016, pp. 353–362. Springer International Publishing, Cham (2016)
2. Beliakov, G., Sola, H.B., Sánchez, T.C.: A practical guide to averaging functions, *Studies in Fuzziness and Soft Computing*, vol. 329. Springer (2016)
3. Bustince, H.: Indicator of inclusion grade for interval-valued fuzzy sets. application to approximate reasoning based on interval-valued fuzzy sets. Internat. J. Appr. Reas. **23**, 137 – 209 (2000)
4. Bustince, H., Fernandez, J., Kolesárová, A., Mesiar, R.: Generation of linear orders for intervals by means of aggregation functions. Fuzzy Sets Syst. **220**, 69–77 (2013)
5. Bustince, H., Galar, M., Bedregal, B., Kolesárová, A., Mesiar, R.: A new approach to interval-valued choquet integrals and the problem of ordering in interval-valued fuzzy sets applications. IEEE Trans. on Fuzzy Syst. **21**(6), 1150–1162 (2013)
6. Dua, D., Graff, C.: UCI machine learning repository (2017)
7. Dubois, D., Liu, W., Ma, J., Prade, H.: The basic principles of uncertain information fusion. an organised review of merging rules in different representation frameworks. Information Fusion **32**, 12–39 (2016)
8. Dyczkowski, K., Pękala, B., Szkoła, J., Wilbik, A.: Federated learning with uncertainty on the example of a medical data. 2022 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE) (2002)

9. Dyczkowski, K., Wójtowicz, A., Żywica, P., Stachowiak, A., Moszyński, R., Szubert, S.: An Intelligent System for Computer-Aided Ovarian Tumor Diagnosis. In: Intelligent Systems'2014, pp. 335–343. Springer International Publishing, Cham (2015)
10. Gorzałczany, M.B.: A method of inference in approximate reasoning based on interval-valued fuzzy sets. Fuzzy Sets and Systems **21**(1), 1–17 (1987)
11. Hosmer, D.W.: Applied Logistic Regression, 3nd edn. John Wiley and Sons Inc (2013)
12. Kologlu, M., Elker, D., Altun, H., Sayek, I.: Validation of mpi and pia ii in two different groups of patients with secondary peritonitis. Hepato-Gastroenterology **48**, 141–151 (2001)
13. Komorníková, M., Mesiar, R.: Aggregation functions on bounded partially ordered sets and their classification. Fuzzy Sets and Systems **175**(1), 48–56 (2011)
14. Li, T., Sahu, A.K., Talwalkar, A., Smith, V.: Federated learning: Challenges, methods, and future directions. IEEE Signal Processing Magazine **37**(3), 50–60 (2020)
15. Lynam, A.L., Dennis, J.M., Owen, K.R., Oram, R.A., Jones, A.G., Shields, B.M., Ferrat, L.A.: Logistic regression has similar performance to optimised machine learning algorithms in a clinical setting: application to the discrimination between type 1 and type 2 diabetes in young adults. Diagn Progn Res. **4**, 6 (2020)
16. Moore, R.E.: Interval analysis. Prentice Hall (1966)
17. Moore, R.E.: Methods and applications of interval analysis. SIAM (1979)
18. Niewęgłowski, K., Wilczek, N., Madoń, B., Palmi, J., Wasyluk, M.: Applications of artificial intelligence (ai) in medicine. Med Og Nauk Zdr. **27**(3), 213–219 (2021)
19. Nusinovici, S., Tham, Y.C., Chak Yan, M.Y., Wei Ting, D.S., Li, J., Sabanayagam, C., Wong, T.Y., Cheng, C.Y.: Logistic regression was as good as machine learning for predicting major chronic diseases. J. Clin. Epidem. **122**, 56–69 (2020)
20. Pękala, B.: Uncertainty Data in Interval-Valued Fuzzy Set Theory: Properties, Algorithms and Applications, *Studies Fuzz. Soft Comp.*, vol. 367. Springer (2019)
21. Pękala, B., Bentkowska, U., Bustince, H., Fernandez, J., Galar, M.: Operators on intuitionistic fuzzy relations. In: 2015 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), pp. 1–8. IEEE (2015)
22. Piegat, A., Landowski, M.: Multidimensional approach to interval uncertainty calculations. In: K. Atanassov, et al. (eds.) New Trends in Fuzzy Sets, Intuitionistic: Fuzzy Sets, Generalized Nets and Related Topics, p. 137–151. IBS PAN-SRI PAS, Warsaw (2013)
23. Sambuc, R.: Fonctions ϕ-floues: Application á l'aide au diagnostic en pathologie thyroidienne. Ph.D. thesis, Faculté de Médecine de Marseille (1975). (in French)
24. Türksen, I.B.: Interval valued fuzzy sets based on normal forms. Fuzzy Sets and Systems **20**(2), 191–210 (1986)
25. Yang, Q., Liu, Y., Chen, T., Tong, Y.: Federated machine learning: Concept and applications. ACM Trans. Intell. Syst. Technol. **10**(2) (2019)
26. Zadeh, L.A.: Fuzzy sets. Information and Control **8**(3), 338–353 (1965)
27. Zadeh, L.A.: The concept of a linguistic variable and its application to approximate reasoning–i. Information Sciences **8**(3), 199–249 (1975)
28. Zapata, H., Bustince, H., Montes, S., Bedregal, B., Dimuro, G.P., Takáč, Z., Baczyński, M., Fernandez, J.: Interval-valued implications and interval-valued strong equality index with admissible orders. Int. J. Appr. Reas. **88**, 91–109 (2017)