# Chapter 1
# Influence of interval-valued measures on classification methods with missing values

Kosior Dawid[1] and Barbara Pękala[1,2]

**Abstract** .
In this paper, the basic version of the k-nearest neighbors (k-NN) algorithm has been extended to include similarity and indistinguishability measures that reflect uncertainty. The measures used take into account the uncertainty in the decision-making process when classifying objects. The algorithm has been adapted to the processing of attribute values stored as interval values. The algorithm was tested on data with missing values. Three approaches were tested to determine the missing value of individual objects.

**Key words:** IV-kNN, interval-valued fuzzy set, aggregation function, precedence indicator, indistinguishability, similarity measure.

## 1.1 Introduction

Since the introduction of fuzzy sets by Zadeh [1], many new approaches and theories have been proposed for the study and modeling of uncertainty. Fuzzy sets with interval values (Zadeh and Sambuc 1975) or intuitionistic fuzzy sets (Atanassov 1986) have proved particularly useful for representing uncertainty and modeling knowledge with uncertainty (see [2] for more details). Besides, various applications of fuzzy sets with interval values have been successfully proposed to solve real-world problems including pattern recognition, medical diagnostics, and decision making. The above results are possible due to significant advances in the fuzzy set theory with interval values. Because many researchers have proposed different distances, measures of inclusion, equiv-

1. College of Natural Sciences, Institute of Computer Science, University of Rzeszów, Rzeszów, Poland
2. Department of Artificial Intelligence, University of Information Technology and Management, Rzeszów, Poland

alence measures, or similarity measures for fuzzy sets with interval values and investigated different types of relationships between them with reflecting dependences to interval aggregation operators ([3] - [12]).

The main theme of the article is the use of the impact of the lack of data and methods of their supplementation on diagnostic methods, in particular, based on the degree of inclusion and the measure of equivalence between interval values in the classification problem, especially in the k-nearest neighbors (k-NN) algorithm. Here we consider the concept of a fuzzy equivalence in an interval-valued fuzzy setting which may be treated as a special kind of lattice with its underlying interpretation and semantics. Interval-valued fuzzy sets represent uncertainty observed in many fields of applications. In the proposed definition of interval-valued equivalence, we use the usual axioms for an equivalence relation, i.e. reflexivity, symmetry, and transitivity. Moreover, we study a reference to the methods using the similarity measure. The fundamental novelty of the proposed approach is the use of the different measures between intervals and the width of these intervals representing both the uncertainty of information and the imprecision of the membership function.

The paper is organized as follows. In Section 2 basic information on interval-valued fuzzy sets are recalled. Inclusion, similarity, and indistinguishability degree measures for interval-valued fuzzy sets are presented in Section 3. Finally, the algorithm of application mentioned measures in k-NN classifiers by missing data problem is considered (Section 4).

## 1.2 Interval-valued fuzzy set theory

Firstly, we recall definition of an interval-valued fuzzy set and the classically applied order for this setting.

**Definition 1 (cf. [13, 14])** An interval-valued fuzzy set IVFS $\widetilde{A}$ in $X$ is a mapping $\widetilde{A} : X \to L^I$ such that $\widetilde{A}(x) = [\underline{A}(x), \overline{A}(x)] \in L^I$ for $x \in X$, where

$$\widetilde{A} \cap \widetilde{B} = \big\{\langle x, \big[\min\{\underline{A}(x), \underline{B}(x)\}, \min\{\overline{A}(x), \overline{B}(x)\}\big]\rangle : x \in X\big\},$$

$$\widetilde{A} \cup \widetilde{B} = \big\{\langle x, \big[\max\{\underline{A}(x), \underline{B}(x)\}, \max\{\overline{A}(x), \overline{B}(x)\}\big]\rangle : x \in X\big\} \ \ and$$

$$L^I = \{[x_1, x_2] : \ x_1, x_2 \in [0,1], \ x_1 \leq x_2\}.$$

The well-known classical monotonicity (partial order) for intervals is of the form

$$[x_1, y_1] \leq_2 [x_2, y_2] \Leftrightarrow x_1 \leq x_2, \ y_1 \leq y_2 \quad and$$

$[\underline{x}, \overline{x}] <_2 [\underline{y}, \overline{y}] \Leftrightarrow [\underline{x}, \overline{x}] \leq_2 [\underline{y}, \overline{y}] \ \ and \ \ (\underline{x} < \underline{y} \ \ or \ \ \overline{x} < \overline{y}).$

The family of all interval-valued fuzzy sets on a given universe $X$ with $\leq_2$ is partially ordered and moreover it is a lattice. The operations joint and meet are defined respectively

$$[\underline{x},\overline{x}]\vee[\underline{y},\overline{y}]=[\max(\underline{x},\underline{y}),\max(\overline{x},\overline{y})],\ [\underline{x},\overline{x}]\wedge[\underline{y},\overline{y}]=[\min(\underline{x},\underline{y}),\min(\overline{x},\overline{y})].$$

Note that the structure $(L^I,\vee,\wedge)$ is a complete lattice, with the partial order $\leq_2$, where $\mathbf{1}=[1,1]$ and $\mathbf{0}=[0,0]$ are the greatest and the smallest element of $(L^I,\leq_2)$, respectively.

Since in many real-life problems we need a linear order to be able to compare any two intervals, we are interested in extending the partial order $\leqslant_2$ to a linear one. The concept of, so called, admissible order would be of useful there.

**Definition 2 ([15])** An order $\leq$ in $L^I$ is called admissible if it is linear and satisfies that for all $x,y\in L^I$, such that $x\leq_{L^I} y$, then $x\leq y$.

Simply, an order $\leq$ on $L^I$ is admissible, if it is linear and refines the order $\leq_{L^I}$.

**Proposition 1 ([15])** *Let $B_1,B_2:[0,1]^2\to[0,1]$ be two continuous aggregation functions, such that, for all $x=[\underline{x},\overline{x}],y=[\underline{y},\overline{y}]\in L^I$, the equalities $B_1(\underline{x},\overline{x})=B_1(\underline{y},\overline{y})$ and $B_2(\underline{x},\overline{x})=B_2(\underline{y},\overline{y})$ hold if and only if $x=y$. If the order $\leq_{B_{1,2}}$ on $L^I$ is defined by $x\leq_{B_{1,2}} y$ if and only if $B_1(\underline{x},\overline{x})<B_1(\underline{y},\overline{y})$ or $(B_1(\underline{x},\overline{x})=B_1(\underline{y},\overline{y})$ and $B_2(\underline{x},\overline{x})\leq B_2(\underline{y},\overline{y}))$, then $\leq_{B_{1,2}}$ is an admissible order on $L^I$.*

*Example 1* Admissible (linear) orders on $L^I$:
• Xu and Yager order: $[\underline{x},\overline{x}]\leq_{XY}[\underline{y},\overline{y}]$ if and only if

$$\underline{x}+\overline{x}<\underline{y}+\overline{y}\ or\ (\overline{x}+\underline{x}=\overline{y}+\underline{y}\ and\ \overline{x}-\underline{x}\leq\overline{y}-\underline{y}).$$

• The first lexicographical order (with respect to the first variable), $\leq_{lex1}$ defined as: $[\underline{x},\overline{x}]\leq_{lex1}[\underline{y},\overline{y}]$ if and only if

$$\underline{x}<\underline{y}\ or\ (\underline{x}=\underline{y}\ and\ \overline{x}\leq\overline{y}).$$

• The second lexicographical order (with respect to the second variable), $\leq_{lex2}$ defined as: $[\underline{x},\overline{x}]\leq_{lex2}[\underline{y},\overline{y}]$ if and only if

$$\overline{x}<\overline{y}\ or\ (\overline{x}=\overline{y}\ and\ \underline{x}\leq\underline{y}).$$

*Remark 1* In the later part we will use the notation $\leq$ both for the partial or admissible linear order, with $\mathbf{0}$ and $\mathbf{1}$ as minimal and maximal element of $L^I$, respectively. Notation $\leq_2$ will be used while the results for the admissible linear orders will be used with the notation $\leq_{Adm}$.

Now we recall the concept of an aggregation function on $L^I$ which is an important notion in many applications. We consider aggregation functions both with respect to $\leqslant_2$ and $\leqslant_{Adm}$.

**Definition 3 ([16, 17])** An operation $\mathcal{A} : (L^I)^n \to L^I$ is called an interval-valued aggregation function if it is increasing with respect to the order $\leq$ (partial or total) and

$$\mathcal{A}(\underbrace{\mathbf{0}, ..., \mathbf{0}}_{n\times}) = \mathbf{0}, \quad \mathcal{A}(\underbrace{\mathbf{1}, ..., \mathbf{1}}_{n\times}) = \mathbf{1}.$$

The special case of interval-valued aggregation operation is a representable interval-valued aggregation function with respect to $\leq_2$.

**Definition 4 ([18, 19])** An interval-valued aggregation function $\mathcal{A} : (L^I)^n \to L^I$ is called representable if there exist aggregation functions $A_1, A_2 : [0,1]^n \to [0,1]$ such that

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

The next result shows the characterization of representable aggregation functions on $L^I$.

**Theorem 1 ([20])** *An operation $\mathcal{A} : (L^I)^n \to L^I$ is a representable interval-valued aggregation function with respect to $\leq_2$ if and only if there exist aggregation functions $A_1, A_2 : [0,1]^n \to [0,1]$ such that for all $x_1, ..., x_n \in L^I$ and $A_1 \leqslant A_2$*

$$\mathcal{A}(x_1, ..., x_n) = [A_1(\underline{x}_1, ...\underline{x}_n), A_2(\overline{x}_1, ...\overline{x}_n)]. \tag{1.1}$$

*Example 2* Lattice operations $\wedge$ and $\vee$ on $L^I$ are representable aggregation functions on $L^I$ with $A_1 = A_2 = \min$ in the first case and $A_1 = A_2 = \max$ in the second one. It holds true with respect to the order $\leq_2$, but not $\leq_{Lex1}$, $\leq_{Lex2}$ or $\leq_{XY}$. Moreover, many other examples of representable aggregation functions with respect to $\leq_2$ may be considered, such as:

- 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 representable mean-power mean $\mathcal{A}_{meanpow}([\underline{x}, \overline{x}], [\underline{y}, \overline{y}]) = [\frac{\underline{x}+\underline{y}}{2}, \sqrt{\frac{\overline{x}^2+\overline{y}^2}{2}}]$,
- the representable mean-max $\mathcal{A}_{meanmax}([\underline{x}, \overline{x}], [\underline{y}, \overline{y}]) = [\frac{\underline{x}+\underline{y}}{2}, \max \overline{x}, \overline{y}]$

for $[\underline{x}, \overline{x}], [\underline{y}, \overline{y}] \in L^I$.

## 1.3 Interval-valued fuzzy measures

In this section, we recall three measures that play a crucial role in our studies, i.e. precedence indicator, similarity, and equivalence measures with respect to uncertainty.

### 1.3.1 Precedence indicator

Before we get to the main issue that interests us, i.e. the similarity and equivalence measures we recall the notion of a precedence indicator that use linear or partial order and uncertainty measure [21]. Which pay attention to the width of the intervals, as this width is seen as a way to measure the uncertainty related to the interval, and which we denote by $w$, where
$w(\alpha) = \overline{\alpha} - \underline{\alpha}$ for $\alpha \in L^I$.

The precedence indicators can be used to build subsethood or inclusion measures which have been discussed in axiomatic or constructive ways. We remind that we use the notation $\leq$ both for linear and partial orders.

**Definition 5 ([21])** A function Prec : $(L^I)^2 \to L^I$ is said to be a **precedence indicator** if it satisfies the following conditions for any $a, b, c \in L^I$

P1 if $a = 1_{L^I}$ and $b = 0_{L^I}$, then $\mathrm{Prec}(a, b) = 0_{L^I}$;
P2 if $a < b$, then $\mathrm{Prec}(a, b) = 1_{L^I}$ for any $a, b \in L^I$;
P3 $\mathrm{Prec}(a, a) = [1 - w(a), 1]$ for any $a \in L^I$;
P4 if $a \leq b \leq c$ and $w(a) = w(b) = w(c)$, then $\mathrm{Prec}(c, a) \leq \mathrm{Prec}(b, a)$ and $\mathrm{Prec}(c, a) \leq \mathrm{Prec}(c, b)$ for any $a, b, c \in L^I$.

Here is the example of the construction of precedence indicator.

**Proposition 2 ([21])** *The operation*

$$\mathrm{Prec}_w(a, b) = \begin{cases} 1_{L^I}, & a < b, \\ [1 - max(w(a), r(a, b)), 1 - r(a, b)], & else \end{cases}$$

*is the precedence indicator with respect to* $\leq$, *where*
$r(a, b) = \max\{|\underline{a} - \underline{b}|, |\overline{a} - \overline{b}|\}$ *for* $a, b \in L^I$.

### 1.3.2 Similarity measure

In this section, we recall a class of similarity measures between interval-valued fuzzy sets. To construct this interval-valued similarity, interval-valued aggregation functions and interval-valued inclusion measure which take into account the width of the intervals are needed [22].

**Definition 6 ([23])** Let $A_1 : [0, 1]^n \to [0, 1]$ be an aggregation function. Then operation $S : IVFS(X) \times IVFS(X) \to L^I$, which satisfies the following items:
(SIMV1) $S(A, B) = S(B, A)$ for $A, B \in IVFS(X)$;
(SIMV2) $S(A, A) = [1 - A_1(w_A(x_1), ..., w_A(x_n)), 1]$;
(SIMV3) $S(A, B) = 0_{L^I}$, if $\{A(x_i), B(x_i)\} = \{0_{L^I}, 1_{L^I}\}$;
(SIMV4) if $A \preceq B \preceq C$ and $w_A(x_i) = w_B(x_i) = w_C(x_i)$, then $S(A, C) \leq$

$S(A,B)$ and $S(A,C) \leq S(B,C)$
is called a similarity measure for $i = 1, ..., n$, where

$$A \preceq B \Leftrightarrow a_i \leq b_i \text{ for each } i \text{ and } a_i = A(x_i),\ b_i = B(x_i).$$

**Proposition 3** *Let* Prec *be a precedence indicator. If* $\mathcal{A} = [A_1, A_2]$, $\mathcal{B} = [B_1, B_2]$ *are representable interval-valued aggregation functions for which* $A_1$ *is as in previous definition and self-dual,* $\mathcal{B}$ *is symmetric with the neutral element* $1_{L^I}$ *and* $B_1$ *is idempotent aggregation function, then the operation* $S : IVFS(X) \times IVFS(X) \to L^I$*:*

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

*is a similarity measure, where* $w_A(x_i) = w_B(x_i)$*.*

*Example 3* The operation $S : IVFS(X) \times IVFS(X) \to L^I$:

$$S(A,B) = \mathcal{A}_{i=1}^{n}(\mathrm{Prec}_w(A(x_i), B(x_i)) \wedge \mathrm{Prec}_w(B(x_i), A(x_i))))$$

is a similarity measure, where $\mathcal{A} \in \{\mathcal{A}_{mean}, \mathcal{A}_{meanpow}, \mathcal{A}_{meanmax}\}$.

### 1.3.3 Equivalence measures

In this section, we discuss some of the new types of interval-valued equivalence, partially presented in [24], which pays attention to the width of the input intervals and is defined both for partial order or the linear one.

**Definition 7** Let $\mathcal{B} : (L^I)^2 \to L^I$ be an IV aggregation (with respect to the order $\leq$). An operation $E : (L^I)^2 \to L^I$ is called a $\mathcal{B}$-equivalence measure if for $\alpha, \beta, \gamma \in L^I$ it holds:
(E1) $E(\alpha, \alpha) = [1 - w(\alpha), 1]$ (reflexivity);
(E2) $E(\alpha, \beta) = E(\beta, \alpha)$ (symmetry);
(E3) $\mathcal{B}(E(\alpha, \beta), E(\beta, \gamma)) \leq E(\alpha, \gamma)$ ($\mathcal{B}$-transitivity).

We use the following notation $EM = E : (L^I)^2 \to L^I$, where $E$ is an interval-valued $\mathcal{B}$-equivalence measure.

In the following proposition we may observe an interesting connection between the mentioned precedence indicators and equivalence measures.

**Proposition 4** *Let* $\mathcal{A}$ *be a symmetric, idempotent IV aggregation,* Prec *be* $\mathcal{B}$*-transitive precedence indicator, where* $\mathcal{B}$ *be a symmetric IV aggregation and* $\mathcal{A} \gg \mathcal{B}$ *(the dominance property and* $\mathcal{B}$*-transitivity are considered with the same order). Thus a function* $E_{\mathcal{A}} : (L^I)^2 \to L^I$,

$$E_{\mathcal{A}}(\alpha, \beta) = \mathcal{A}(\mathrm{Prec}(\alpha, \beta), \mathrm{Prec}(\beta, \alpha))$$

*is a* $\mathcal{B}$*-equivalence measure.*

### 1.3.4 Indistinguishability

Now we propose a class of a new type of measures, namely indistinguishability measures, between interval - valued fuzzy sets (IVFSs). The inspiration of this approach is the fact that we develop all the notions with respect to the total orders of intervals and that we consider the width of intervals. So the uncertainty of the output is strongly connected with the uncertainty of the inputs. For creating the new interval-valued indistinguishability, IV aggregation functions and IV equivalence measure which take into account the width of the intervals are needed.

**Definition 8** Let $A_1$ be a n-argument aggregation function and $\mathcal{B}$ be an IV aggregation function. Then $\mathbf{I} : IVFS(X) \times IVFS(X) \to L^I$, which satisfies the following items:
($\mathbf{I}1$) $\mathbf{I}(A, A) = [1 - A_1(w_A(\alpha_1), ..., w_A(\alpha_n)), 1]$;
($\mathbf{I}2$) $\mathbf{I}(A, B) = \mathbf{I}(B, A)$;
($\mathbf{I}3$) $\mathcal{B}(\mathbf{I}(A, B), \mathbf{I}(B, C)) \leq \mathbf{I}(A, C)$ for $A, B, C \in IVFS(X)$
is called a $\mathcal{B}$-indistinguishability measure between interval-valued fuzzy sets.

**Proposition 5** *Let for a representable IV aggregation function $\mathcal{A} = [A_1, A_2]$ $A_1$ be self-dual, $\mathcal{B}$ be an IV aggregation function and $\mathcal{A} \gg \mathcal{B}$ ($\mathcal{A}$, $\mathcal{B}$ and the dominance are in accordance with the order $\leq$). If $E \in EM$, then $\boldsymbol{I}_{\mathcal{A}}$ : $IVFS(X) \times IVFS(X) \to L^I$,*

$$\boldsymbol{I}_{\mathcal{A}}(A, B) = \mathcal{A}(E(A(\alpha_1), B(\alpha_1)), ..., E(A(\alpha_n), B(\alpha_n)))$$

*is a $\mathcal{B}$-indistinguishability measure for IVFSs.*

*Example 4* An operation

$$\mathbf{I}_{\mathcal{A}}(A, B) = \mathcal{A}_{mean}(E_{\mathcal{A}_{mean\,z}}(A(\alpha_1), B(\alpha_1)), ..., E_{\mathcal{A}_{mean\,z}}(A(\alpha_n), B(\alpha_n)))$$

is a $T_{L_{IV}}$-indistinguishability measure for IVFSs, where

$$E_{\mathcal{A}_{mean\,z}}(\alpha, \beta) = \begin{cases} [1 - w(\alpha), 1], & \alpha = \beta, \\ [\frac{1}{2}, \frac{1}{2}], & \alpha < \beta \text{ } or \text{ } \beta < \alpha, \\ \mathbf{0}, & otherwise \text{ } (\alpha || \beta) \end{cases}$$

is $T_{L_{IV}}$-equivalence measure with respect to $\leq_{L^I}$ created by $\mathcal{A} = \mathcal{A}_{mean}$ by $E_{\mathcal{A}_z}(\alpha, \beta) = \mathcal{A}(\mathrm{Prec}_z(\alpha, \beta), \mathrm{Prec}_z(\beta, \alpha))$, where

$$\mathrm{Prec}_z(\alpha, \beta) = \begin{cases} [1 - w(\alpha), 1], & \alpha = \beta, \\ \mathbf{1}, & \alpha < \beta, \\ \mathbf{0}, & otherwise \end{cases}$$

for $\alpha, \beta \in L^I$ ($\mathcal{A}_{mean} \gg T_{L_{IV}}$).

## 1.4 Application in K-NN classifiers

The classification problem consists of determining the class (category) to which a new, previously unknown object should be assigned. Among the most commonly used techniques for classification can be distinguished k-Nearest Neighbors classifier (k-NN) [25]. In this paper, we present the modified interval-valued fuzzy classifier for supporting decision-making processes based on imprecise (uncertain) data. The main goal was to develop a comprehensive and effective the approach that enables the modeling and processing of input data, and then the presentation of results, to be done in a way that preserves the valuable information concerning the amount of uncertainty at each stage of the process. Especially, the main problem will be to compare the effectiveness of the method based on the measure of similarity taking the measure of indistinguishability and, moreover, the impact of the lack of data.

### 1.4.1 Data set description

The dataset is a Wisconsin (diagnostic) breast cancer dataset. This is one of the popular datasets from UCI Machine Learning Repository [26]. Data are from November 1995. Data containing information on 569 instances. Features are computed 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 ($\frac{perimeter^2}{area-1.0}$)
- concavity (severity of concave portions of the contour)
- concave points (number of concave portions of the contour)
- symmetry
- fractal dimension ($coastline\ approximation - 1$)

The decision attribute stores information about the diagnosis: malignant or benign represented by values of 0 or 1. The dataset consists of 212 objects with malignant diagnosis and 357 objects with the benign diagnosis.

### 1.4.2 Proposed method

This steps list interval-valued k-Nearest Neighbors classifier (IV-kNN) shows the main steps of the proposed classification method.

**Algorithm.**
Step 1. Data double step fuzzyfication;
1.1. Fuzzyfication. The real data are normalized to $[0, 1]$ by classical formula based on minimal values $m_i$ and maximal $M_i$ for $i - th$ attribute with real value $b_i$

$$a_i = \frac{b_i - m_i}{M_i - m_i}.$$

1.2. 2-fuzzyfication. Interval-valued fuzzy set for each instance/object is built in the following way:

$$[x(1 - 0.25g(x)), x(1 - 0.25g(x)) + 0.25g(x)] \in L^I.$$

Note that the length of the intervals is equal to $g(x) = 2\min(x, 1 - x)$, where $g$ is ignorance function [27].
Step 2. Filling the missing value with one of the proposed methods:

1. For a missing value, insert a [0,1] interval.

2. Search for the most similar $k$ objects and select their value for the missing attribute and calculate for the missing value insert the minimum, maximum or average of the selected values on the missing attribute:

$$[min(a_1, ...a_k), max(a_1, ...a_k)], \quad [avg(a_1, ...a_k), avg(a_1, ...a_k)].$$

3. Find the most similar $l$ objects and choose their value for missing attribute and calculate for missing value insert minimum, maximum or average of selected values on missing attribute. For training objects, similar objects from the same decision class are selected:

$$[min(a_1, ..., a_l), max(a_1, .., a_l)], \quad [avg(a_1, ...a_l), avg(a_1, ...a_l)].$$

Step 3. Import new object (without decision);
Step 4. For k=1...n
Calculating similarities ($S_{\mathcal{M}}$) using similarity ($S$) or indistinguishability ($I$) measure for a new object;
Step 5. Ordered of a interval-valued fuzzy similarities;
Step 6. A selection of the $k$ most similar neighbors;
Step 7. Aggregation of individual classes among $k$ most like neighbors ($C_{\mathcal{A}}$);
Step 8. Class selection - selection of the largest interval for each class [28];
Step 9. Output: Class for the new object.

### 1.4.3 Results and discussion

The algorithm testing procedure was as follows:

- The data was divided into a training part and a test part. The training part have 70% of all and testing part have 30% from all objects.
- The selection of objects for both parts was random and made ten times. For each of the ten datasets, we test algorithm with different parameters were checked and accuracy, sensitivity, specificity and precision were calculated.
- From the ten values obtained in this way for accuracy, sensitivity, specificity and precision their arithmetic mean was calculated and presented as the value obtained by the algorithm with particular parameters.
- The algorithm was tested on data with 5%, 25%, 50% of missing values.

We will focus on presenting the conclusions of the algorithm analysis in three aspects:

1. modification of the value of $k$;
2. the selection of aggregation in the construction of the similarity measure and indistinguishability measure;
3. checking the different methods of filling in missing values and different sizes of missing values in the data;
4. the selection of ordering relations ($\leq_2, \leq_{XY}, \leq_{Lex1}, \leq_{Lex2}$).

In tables 1.1, 1.2 and 1.3 we present the results of our research. By $MS$ we denote the percent of missing values in data. For two methods of supplementing the missing values, two options for determining the interval value were used. The designated value was to replace the missing one. The first option constructed the range where the lower bound is the minimum of the lower bounds of ranges and the upper bound is the maximum of the upper bounds values of the ranges it was much better in the tests than the approach based on the mean value of the bounds of the range. However, this does not mean that it would not be otherwise for different data. In tables 1.2 and 1.3 we have only presented the results using the first option.

The tests showed that the methods of filling missing values studied by us by using our modified IV-kNN are characterized by a slower decrease in accuracy of the classification for the third method compare to the first and the second, with an increase in the number of missing values in the data. It is good to see this difference between the results on the set with 25% and 50% of the missing values. The use of the method with indistinguishability measure gave slightly better classification results, but the method based on similarity measure should not be considered worse, because the differences are less than 2%, these are slight differences that may result from the specificity of the data processed.

| $k$ | accuracy | sensitivity | specificity | precision | $C_{\mathcal{A}}$ | $S_{\mathcal{A}}$ | order | $MS$ | $S_{\mathcal{M}}$ |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 0,877295213 | 0,956354306 | 0,833021922 | 0,90946816 | $\mathcal{A}_{meanmax}$ | $\mathcal{A}_{mean}$ | $\leq_{lex2}$ | 5% | $I$ |
| | 0,855623164 | 0,952316431 | 0,785443313 | 0,888174475 | $\mathcal{A}_{mean}$ | $\mathcal{A}_{mean}$ | $\leq_{lex2}$ | 25% | |
| | 0,847856945 | 0,950069216 | 0,779792944 | 0,881347978 | $\mathcal{A}_{meanmax}$ | $\mathcal{A}_{mean}$ | $\leq_{lex2}$ | 50% | |
| 1 | 0,863293246 | 0,952064793 | 0,805510005 | 0,894909111 | $\mathcal{A}_{meanmax}$ | $\mathcal{A}_{meanmax}$ | $\leq_{lex1}$ | 5% | $S$ |
| | 0,841757424 | 0,946752024 | 0,758375556 | 0,874332808 | $\mathcal{A}_{meanmax}$ | $\mathcal{A}_{meanpow}$ | $\leq_{lex2}$ | 25% | |
| | 0,837122851 | 0,946521682 | 0,758785615 | 0,870113533 | $\mathcal{A}_{meanmax}$ | $\mathcal{A}_{meanpow}$ | $\leq_{lex2}$ | 50% | |
| 3 | 0,86671996 | 0,957699335 | 0,806798803 | 0,896665004 | $\mathcal{A}_{meanmax}$ | $\mathcal{A}_{mean}$ | $\leq_{lex2}$ | 5% | $I$ |
| | 0,846573346 | 0,953220951 | 0,761929092 | 0,877267935 | $\mathcal{A}_{mean}$ | $\mathcal{A}_{mean}$ | $\leq_{lex2}$ | 25% | |
| | 0,83858117 | 0,953411195 | 0,752340814 | 0,86839648 | $\mathcal{A}_{meanmax}$ | $\mathcal{A}_{mean}$ | $\leq_{lex2}$ | 50% | |
| 3 | 0,856294636 | 0,953290925 | 0,788916968 | 0,886932849 | $\mathcal{A}_{meanmax}$ | $\mathcal{A}_{meanpow}$ | $\leq_{lex2}$ | 5% | $S$ |
| | 0,83483665 | 0,9471642 | 0,741429156 | 0,866483001 | $\mathcal{A}_{meanmax}$ | $\mathcal{A}_{meanpow}$ | $\leq_{lex2}$ | 25% | |
| | 0,828317541 | 0,946351344 | 0,738225444 | 0,860137423 | $\mathcal{A}_{meanmax}$ | $\mathcal{A}_{meanpow}$ | $\leq_{lex2}$ | 50% | |
| 5 | 0,860619786 | 0,956844717 | 0,793790423 | 0,890199054 | $\mathcal{A}_{min}$ | $\mathcal{A}_{mean}$ | $\leq_{XY}$ | 5% | $I$ |
| | 0,836040603 | 0,950029196 | 0,741432911 | 0,867363151 | $\mathcal{A}_{min}$ | $\mathcal{A}_{mean}$ | $\leq_{lex2}$ | 25% | |
| | 0,832745543 | 0,949909305 | 0,744427506 | 0,863729177 | $\mathcal{A}_{min}$ | $\mathcal{A}_{mean}$ | $\leq_{XY}$ | 50% | |
| 5 | 0,861724323 | 0,956774869 | 0,79687821 | 0,890825269 | $\mathcal{A}_{mean}$ | $\mathcal{A}_{mean}$ | $\leq_{XY}$ | 5% | $S$ |
| | 0,835598403 | 0,948806975 | 0,742392271 | 0,867691953 | $\mathcal{A}_{mean}$ | $\mathcal{A}_{mean}$ | $\leq_{lex2}$ | 25% | |
| | 0,82592191 | 0,944076375 | 0,738009497 | 0,86004546 | $\mathcal{A}_{mean}$ | $\mathcal{A}_{meanmax}$ | $\leq_2$ | 50% | |

**Table 1.1** Table for the first method of filling missing values.

| $k$ | accuracy | sensitivity | specificity | precision | $C_{\mathcal{A}}$ | $S_{\mathcal{A}}$ | order | $MS$ | $S_{\mathcal{M}}$ |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 0,875164803 | 0,955569769 | 0,829275549 | 0,907565691 | $\mathcal{A}_{meanmax}$ | $\mathcal{A}_{mean}$ | $\leq_{lex2}$ | 5% | $I$ |
| | 0,862216862 | 0,951867453 | 0,801919554 | 0,895761817 | $\mathcal{A}_{mean}$ | $\mathcal{A}_{mean}$ | $\leq_{lex2}$ | 25% | |
| | 0,858307794 | 0,954335521 | 0,79649695 | 0,889915539 | $\mathcal{A}_{meanmax}$ | $\mathcal{A}_{mean}$ | $\leq_{lex2}$ | 50% | |
| 1 | 0,860383256 | 0,950900338 | 0,800522779 | 0,892336445 | $\mathcal{A}_{meanmax}$ | $\mathcal{A}_{meanpow}$ | $\leq_{lex2}$ | 5% | $S$ |
| | 0,848219814 | 0,95087641 | 0,769143143 | 0,879745901 | $\mathcal{A}_{mean}$ | $\mathcal{A}_{meanmax}$ | $\leq_{lex1}$ | 25% | |
| | 0,844206219 | 0,949049521 | 0,770200942 | 0,875996301 | $\mathcal{A}_{meanmax}$ | $\mathcal{A}_{meanpow}$ | $\leq_{lex2}$ | 50% | |
| 3 | 0,86430934 | 0,956728007 | 0,802580878 | 0,894504398 | $\mathcal{A}_{meanmax}$ | $\mathcal{A}_{mean}$ | $\leq_{lex2}$ | 5% | $I$ |
| | 0,852343376 | 0,952460141 | 0,777622187 | 0,883955726 | $\mathcal{A}_{mean}$ | $\mathcal{A}_{mean}$ | $\leq_{lex2}$ | 25% | |
| | 0,850073994 | 0,955676891 | 0,774964368 | 0,879456496 | $\mathcal{A}_{meanmax}$ | $\mathcal{A}_{mean}$ | $\leq_{lex2}$ | 50% | |
| 3 | 0,853131268 | 0,952682155 | 0,782661228 | 0,88381208 | $\mathcal{A}_{meanmax}$ | $\mathcal{A}_{meanpow}$ | $\leq_{lex2}$ | 5% | $S$ |
| | 0,841698444 | 0,948807864 | 0,756996901 | 0,873633797 | $\mathcal{A}_{mean}$ | $\mathcal{A}_{meanpow}$ | $\leq_{lex2}$ | 25% | |
| | 0,838060591 | 0,948662744 | 0,757630117 | 0,869702387 | $\mathcal{A}_{meanmax}$ | $\mathcal{A}_{meanpow}$ | $\leq_{lex2}$ | 50% | |
| 5 | 0,858213327 | 0,953285198 | 0,794692982 | 0,889330988 | $\mathcal{A}_{meanpow}$ | $\mathcal{A}_{mean}$ | $\leq_{XY}$ | 5% | $I$ |
| | 0,84107295 | 0,950628378 | 0,753018736 | 0,87232769 | $\mathcal{A}_{meanpow}$ | $\mathcal{A}_{mean}$ | $\leq_{lex2}$ | 25% | |
| | 0,840267842 | 0,950526542 | 0,760773919 | 0,871823934 | $\mathcal{A}_{min}$ | $\mathcal{A}_{mean}$ | $\leq_{XY}$ | 50% | |
| 5 | 0,853580999 | 0,952845669 | 0,784110124 | 0,885236216 | $\mathcal{A}_{mean}$ | $\mathcal{A}_{mean}$ | $\leq_{lex2}$ | 5% | $S$ |
| | 0,84107295 | 0,950628378 | 0,753018736 | 0,87232769 | $\mathcal{A}_{meanpow}$ | $\mathcal{A}_{mean}$ | $\leq_{lex2}$ | 25% | |
| | 0,842305378 | 0,951388131 | 0,764925667 | 0,873062717 | $\mathcal{A}_{mean}$ | $\mathcal{A}_{mean}$ | $\leq_{XY}$ | 50% | |

**Table 1.2** Table for the second method of filling missing values.

## 1.5 Conclusions

We applied the similarity measure and indistinguishability measure reflected uncertainty in the decision-making algorithm based on the interval-valued k-NN method with different methods of filling in missing values and different

| $k$ | accuracy | sensitivity | specificity | precision | $C_{\mathcal{A}}$ | $S_{\mathcal{A}}$ | order | $MS$ | $S_{\mathcal{M}}$ |
|---|---|---|---|---|---|---|---|---|---|
| | 0,876393879 | 0,955600931 | 0,83176004 | 0,908778205 | $\mathcal{A}_{meanmax}$ | $\mathcal{A}_{mean}$ | $\leq_{lex2}$ | 5% | |
| 1 | 0,863174599 | 0,952951685 | 0,802535894 | 0,896140721 | $\mathcal{A}_{mean}$ | $\mathcal{A}_{mean}$ | $\leq_{lex2}$ | 25% | $I$ |
| | 0,859022019 | 0,957291234 | 0,792617045 | 0,888400774 | $\mathcal{A}_{meanmax}$ | $\mathcal{A}_{mean}$ | $\leq_{lex2}$ | 50% | |
| | 0,861068713 | 0,951032444 | 0,801776804 | 0,89305433 | $\mathcal{A}_{meanmax}$ | $\mathcal{A}_{meanpow}$ | $\leq_{lex2}$ | 5% | |
| 1 | 0,849603856 | 0,952318698 | 0,770173382 | 0,880453221 | $\mathcal{A}_{mean}$ | $\mathcal{A}_{meanmax}$ | $\leq_{lex1}$ | 25% | $S$ |
| | 0,845324056 | 0,953290377 | 0,76583975 | 0,874440253 | $\mathcal{A}_{meanmax}$ | $\mathcal{A}_{meanmax}$ | $\leq_{lex1}$ | 50% | |
| | 0,864983644 | 0,956773312 | 0,803794131 | 0,895031392 | $\mathcal{A}_{meanmax}$ | $\mathcal{A}_{mean}$ | $\leq_{lex2}$ | 5% | |
| 3 | 0,85441494 | 0,954623268 | 0,779027873 | 0,88489549 | $\mathcal{A}_{mean}$ | $\mathcal{A}_{mean}$ | $\leq_{lex2}$ | 25% | $I$ |
| | 0,850464446 | 0,958875752 | 0,770235821 | 0,877601312 | $\mathcal{A}_{meanmax}$ | $\mathcal{A}_{mean}$ | $\leq_{lex2}$ | 50% | |
| | 0,853241271 | 0,952609413 | 0,783001422 | 0,883974259 | $\mathcal{A}_{meanmax}$ | $\mathcal{A}_{meanpow}$ | $\leq_{lex2}$ | 5% | |
| 3 | 0,843326753 | 0,950761731 | 0,757763476 | 0,874303137 | $\mathcal{A}_{mean}$ | $\mathcal{A}_{meanpow}$ | $\leq_{lex2}$ | 25% | $S$ |
| | 0,83960183 | 0,953539166 | 0,753201546 | 0,868218926 | $\mathcal{A}_{meanmax}$ | $\mathcal{A}_{meanpow}$ | $\leq_{lex2}$ | 50% | |
| | 0,856801942 | 0,951099929 | 0,794039717 | 0,888744803 | $\mathcal{A}_{meanpow}$ | $\mathcal{A}_{mean}$ | $\leq_{XY}$ | 5% | |
| 5 | 0,845019703 | 0,952937731 | 0,758746393 | 0,875059964 | $\mathcal{A}_{meanpow}$ | $\mathcal{A}_{mean}$ | $\leq_{lex2}$ | 25% | $I$ |
| | 0,83997864 | 0,955184783 | 0,751303913 | 0,86832831 | $\mathcal{A}_{meanpow}$ | $\mathcal{A}_{mean}$ | $\leq_{lex2}$ | 50% | |
| | 0,855589473 | 0,953187164 | 0,787761604 | 0,887139876 | $\mathcal{A}_{meanpow}$ | $\mathcal{A}_{mean}$ | $\leq_{lex2}$ | 5% | |
| 5 | 0,845019703 | 0,952937731 | 0,758746393 | 0,875059964 | $\mathcal{A}_{meanpow}$ | $\mathcal{A}_{mean}$ | $\leq_{lex2}$ | 25% | $S$ |
| | 0,83997864 | 0,955184783 | 0,751303913 | 0,86832831 | $\mathcal{A}_{meanpow}$ | $\mathcal{A}_{mean}$ | $\leq_{lex2}$ | 50% | |

**Table 1.3** Table for the third method of filling missing values.

sizes of missing values in the data. In the future, we will propose new methods of constructing similarity and indistinguishability, we will analyze the impact of using different aggregations, we will analyze another way to fill missing values, moreover, we will check whether and to what extent the conclusions obtained are affected by a small modification of the input parameters, e.g. the weak ignorance function or its parameters.

## References


1. L.A. Zadeh, "Fuzzy sets", Inf. Contr., 8, 338 - 353, 1965.
2. H. Bustince, E. Barrenechea, M. Pagola, J. Fernández, 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 Trans. Fuzzy Syst., 24(1), 174 - 194, 2016.
3. U. Bodenhofer, B. De Baets, J. Fodor, "A compendium of fuzzy weak orders: Representations and constructions", Fuzzy Sets Syst. 158, 811 - 829, 2007.
4. K. Bosteels, E.E. Kerre, "On a reflexivity-preserving family of cardinality-based fuzzy comparison measures", Inf. Sci., 179, 2342 - 2352, 2009.
5. H. Bustince, E. Barrenechea, M. Pagola, "Image thresholding using restricted equivalence functions and maximizing the measures of similarity", Fuzzy Sets Syst., 158, 496 - 516, 2007.
6. B. Pękala, "Operations on interval matrices", International Conference on Rough Sets and Intelligent Systems Paradigms 613-621 Springer, Berlin, Heidelberg, 2007.
7. B. De Baets, H. De Meyer, H. Naessens, "A class of rational cardinality-based similarity measures", J. Comp. Appl. Math., 132, 51 - 69, 2001.
8. B. De Baets, S. Janssens, H. De Meyer, "On the transitivity of a parametric family of cardinality-based similarity measures", Internat. J. Appr. Reas., 50, 104 - 116, 2009.

9. B. Jayaram, R. Mesiar, "I-Fuzzy equivalence relations and I-fuzzy partitions", Inf. Sci., 179, 1278 - 1297, 2009.
10. B. Pękala, "Preservation of Properties of Interval-Valued Fuzzy Relations", IFSA/EUSFLAT Conf., 1206-1210, 2009.
11. B. Pękala, K. Balicki, "Interval-valued intuitionistic fuzzy sets and similarity measure", Iranian Journal of Fuzzy Systems 14 (4), 87-98, 2017.
12. E.S. Palmeira, B. Bedregal, H. Bustince, D. Paternain, L. De Miguel, "Application of two different methods for extending lattice-valued restricted equivalence functions used for constructing similarity measures on L -fuzzy sets", Inf. Sci., 441, 95 - 112, 2018.
13. L.A. Zadeh, "The concept of a linguistic variable and its application to approximate reasoning", Inf. Sci., 8, Part I, 199 - 251, Part II, 301 - 357, Inf. Sci. 9, Part III, 43 - 80, 1975.
14. R. Sambuc, "Fonctions $\phi$-floues: Application á l'aide au diagnostic en pathologie thyroidienne", Ph.D. Thesis, Universit*é* de Marseille, France (in French), 1975.
15. H. Bustince, J. Fernandez, A. Kolesárová, R. Mesiar, "Generation of linear orders for intervals by means of aggregation functions", Fuzzy Sets Syst., 220, 69 - 77, 2013.
16. H. Zapata, H. Bustince, S. Montes, B. Bedregal, G.P. Dimuro, Z. Takáč, M. Baczyński, J. Fernandez, "Interval-valued implications and interval-valued strong equality index with admissible orders", Internat. J. Appr. Reas., 88, 91 - 109, 2017.
17. M. Komorníková, R. Mesiar, "Aggregation functions on bounded partially ordered sets and their classification", Fuzzy Sets Syst., 175, 48 - 56, 2011.
18. G. Deschrijver, E.E. Kerre, "Implicators based on binary aggregation operators in interval-valued fuzzy set theory", Fuzzy Sets Syst., 153(2), 229 - 248, 2005.
19. G. Deschrijver, "Quasi-arithmetic means and OWA functions in interval-valued and Atanassovs intuitionistic fuzzy set theory", In: EUSFLAT-LFA 2011, Aix-les-Bains, France, 506 - 513, 2011.
20. M. Elkano, J.A. Sanz, M. Galar, B. Pękala, U. Bentkowska, H. Bustince, "Composition of interval-valued fuzzy relations using aggregation functions", Inf. Sci., 369, 690 - 703, 2016.
21. B. Pękala, U. Bentkowska, M. Sesma-Sara, J. Fernandez, J. Lafuente, A. Altalhi, M. Knap, H. Bustince, "Interval subsethood measures with respect to uncertainty for interval-valued fuzzy setting", Int. J. Comp. Int. Syst., 3, 167–177, 2020.
22. B Pękala, K Dyczkowski, P Grzegorzewski, U Bentkowska, Inclusion and similarity measures for interval-valued fuzzy sets based on aggregation and uncertainty assessment, Information Sciences 547, 1182-1200, 2021.
23. H. Bustince, C. Marco-Detchart, J. Fernandez, C. Wagner, J.M. Garibaldi, Z. Takáč, "Similarity between interval-valued fuzzy sets taking into account the width of the intervals and admissible orders", Fuzzy Sets and Systems, 390, 23-47, 2020.
24. B Pękala, U Bentkowska, J Fernandez, H Bustince, "Equivalence measures for Atanassov intuitionistic fuzzy setting used to algorithm of image processing", 2019 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), 1-6, 2019.
25. T. M. Cover and P. E. Hart, "Nearest neighbor pattern classification", IEEE Transactions on Information Theory 13(1), 21–27, 1967.
26. https://archive.ics.uci.edu/ml/index.php
27. E. Barrenechea, J. Fernandez, M. Pagola, F. Chiclana, H. Bustince, "Construction of interval-valued fuzzy preference relations from ignorance functions and fuzzy preference relations: Application to decision making", Knowledge-Based Systems, 58, 33–44, 2014.
28. B. Pękala, E. Rak, D. Kosior, M. Mrukowicz, J.G. Bazan, "Application of similarity measures with uncertainty in classification methods", 2020 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), 1-8, 2020.