

Improving medical decisions under incomplete data using interval-valued fuzzy aggregation

Patryk Żywica, Andrzej Wójtowicz, Anna Stachowiak, Krzysztof Dyczkowski

Faculty of Mathematics and Computer Science, Adam Mickiewicz University in Poznań
Umultowska 87, 61-614 Poznań, Poland, e-mail: min@wmi.amu.edu.pl

Abstract

We state a problem concerning how to make an effective and proper decision in the presence of data incompleteness. As an example we consider a medical diagnostic system where the problem of missing data is commonly encountered. We propose and evaluate an approach that makes it possible to reduce the influence of missing data on the final result and to improve the quality of the decision. The process involves interval-valued fuzzy set modelling, uncertaintification of classical methods, and finally aggregation of the incomplete results. It was verified that the aggregation results in meaningful and accurate decisions despite the missing data.

Keywords: missing data, incomplete information, decision-making, uncertainty, aggregation, interval-valued fuzzy sets, supporting medical diagnosis

1. Introduction

It is almost inevitable that empirical observations and real-life research produce incomplete datasets. Incomplete, or missing, data can occur for many reasons, including malfunction of data-collecting equipment, a survey respondent's failure to respond to a question, insufficient resources (time, money) to collect all the data, and others. The important point is that these reasons are natural and unavoidable, and thus the desire for complete datasets is impossible to fulfil. Clearly, missing data can have a significant effect on the conclusions that can be drawn from the data, and so it becomes a crucial issue to deal properly with missingness. In recognition of this problem, missing data analysis and decision-making under incomplete information has recently become an important area of research [1, 2].

The present article is a contribution to the study of decision-making in the presence of incomplete information. The subject of our research is a method for supporting the medical diagnosis of ovarian tumors [3]. Since collecting all the data about a patient is often very difficult, it is crucial that the diagnostic system can give meaningful and accurate results even when some of the data is missing. We present here a novel approach that makes this possible. A key feature of our approach is that we do not use any of the known techniques for estimating missing data and data imputation, because these might

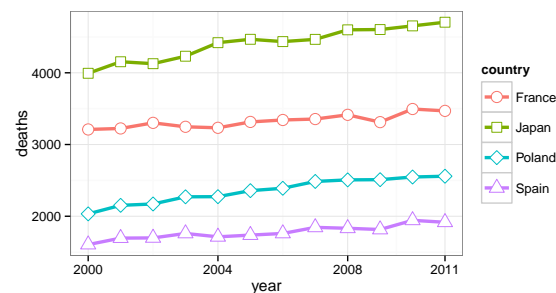


Figure 1: Mortality from ovarian cancer in France, Japan, Poland and Spain between 2000–2011 [4, 5].

significantly distort the final diagnosis. We also do not propose any new diagnostic model specifically for dealing with incomplete data, in order not to introduce any additional complexity to modelling and computation. Instead, we construct a general method that makes it possible to adapt existing and well-established diagnostic methods to make them usable with incomplete data. In the remainder of this article we describe particular components of our approach in detail.

In Section 2 we give a short introduction to the process of ovarian tumor diagnosis, and we state a question concerning how to make an effective diagnosis when some attributes of patients are not given. In Section 3 we propose a solution to this problem which involves IVFS modelling, uncertaintification of classical medical scales, and finally aggregation of the incomplete results obtained. In Section 4 we describe in detail all the tools and methods that were used, and we give a methodology for analysing and comparing different aggregating methods. Section 5 contains conclusions drawn from our results and indicates possible paths of further research.

2. Motivation

One of the major problems currently faced by medicine is that of ovarian tumors. Statistics from recent years show this to be a deadly disease among women. The annual numbers of deaths due to ovarian cancer in some countries are alarmingly high and are still increasing (see Fig. 1). A tumor may be either malignant or benign. We therefore have to deal with a differential diagnosis (binary classification) problem.

The correct classification of a tumor is particu-

larly important for two reasons. Firstly, the type of tumor determines whether the patient must undergo surgery. Secondly, incorrect indication of malignancy as benign causes deterioration of the patient's health in the longer term and results in a high risk of failure of surgery.

For this reason, a wide range of preoperative diagnostic models have been developed, where the goal is to predict the type of malignancy. They vary from basic scoring systems [6], through rule-based schemes [7] to machine learning techniques [8]. Both the sensitivity and specificity of some models exceeds 90% in external evaluation [9, 10].

The total number of attributes, which characterise a patient, amounts to a several dozen. They describe features ranging from basic medical history (such as menopausal status, age, contraception, tumors in family, etc.) to ultrasonographic examination and blood markers.

The Table 1 presents the six most common preoperative models (two based on scoring systems [6, 11] and four based on logistic regressions [8, 12, 13]) and a list of the attributes used. The attributes are divided into two groups: objective medical history and others (which comprise subjective medical history, ultrasound and blood markers). A detailed description of ultrasound features can be found in [14].

Our previous research indicated possible problems with the collection of data by a physician during examinations [15, 16]. It is common that some examinations might be omitted by a gynaecologist, either due to their unavailability or because of medical reasons. Data may be absent due to, for example, the technical limitations of the healthcare institution, the high costs of a medical examination, and the high risk of deterioration in a patient's health after a potential examination.

A common and rather straightforward approach to managing lack of data is to exploit certain well-established methods from the field of data imputation [17]. Undoubtedly in many research areas such an approach is sufficient. Although imputing missing values seems to be adequate from the statistical point of view, in medical applications such a solution must be applied with particular caution. When human life is concerned, it is not so clear whether we can consciously introduce new data which may be subject to error.

Such lack of data means that a physician is not able to calculate and use known diagnostic models, since these require all of the data in order to make a decision. As a result, the problem arises of objectification of the final diagnosis, which is particularly significant in the case of examinations performed by an inexperienced gynaecologist. For that reason, an urgent question is how to construct a diagnostic system that will be feasible in the absence of certain data. In the remainder of this paper we investigate, in particular, the following questions:

- Is it possible to integrate the knowledge that comes from many diagnostic scales to minimise the impact of incomplete data?
- How can a balance be maintained between effectiveness and decisiveness (diagnosability), with particular regard to minimising the risk of wrong decision due to the nature of the test problem (in our case, medical diagnosis)?

3. Proposed approach

3.1. Patient model

In a classical approach, a patient is modelled as a vector \mathbf{p} in a space P . Let D_1, D_2, \dots, D_n be real closed intervals denoting domains of attributes that describe patients. We define a set P in the following way $P = D_1 \times D_2 \times \dots \times D_n$. Then, a vector \mathbf{p} that describes a patient has the form $\mathbf{p} = (p_1, p_2, \dots, p_n)$, where $p_i \in D_i$.

In this basic approach we require a full description for each patient. In order to accept missing values for some attributes we have to add a special element to each domain which will represent them. We thus construct sets $D_i^* = D_i \cup \{\emptyset\}$ and a set P^* defined as $P^* = D_1^* \times \dots \times D_n^*$. Now a patient is modelled by a vector $\mathbf{p}^* = (p_1^*, \dots, p_n^*)$, where $p_i^* \in D_i^*$.

A major disadvantage of this approach is the need to introduce a new, separate value to represent missing values (which are, in practice, commonly denoted by NA). Therefore in this paper we suggest a different approach in which all the data are represented in the same, consistent way.

For each attribute D_i we introduce its interval version $\hat{D}_i = \mathcal{I}_{D_i}$. Analogously as before we define $\hat{P} = \hat{D}_1 \times \hat{D}_2 \times \dots \times \hat{D}_n$. Consequently, for each vector $\mathbf{p}^* \in P^*$ we can define its interval equivalent $\hat{\mathbf{p}} \in \hat{P}$ which has the form $\hat{\mathbf{p}} = ([\underline{p}_1, \bar{p}_1], \dots, [\underline{p}_n, \bar{p}_n])$, where

$$\underline{p}_i = \begin{cases} p_i & \text{if } p_i \neq \emptyset \\ \min_{d \in D_i} d & \text{if } p_i = \emptyset \end{cases}, \bar{p}_i = \begin{cases} p_i & \text{if } p_i \neq \emptyset \\ \max_{d \in D_i} d & \text{if } p_i = \emptyset \end{cases}. \quad (1)$$

The above definition of the vector $\hat{\mathbf{p}}$ allows one to describe each attribute of a patient by a numerical interval, regardless of whether or not the description of an attribute was given. If the value was not provided then the proposed representation has the form of a set containing all possible values for the attribute. If the value was given then it is represented by an interval reduced to a point. The main advantage of such an approach is that all patients can be described in the same, uniform way and can be processed with the same diagnostic model. This model is formally introduced in the next subsection.

3.2. Uncertaintification of diagnostic scales

A diagnostic scale can be formalised as a function $m : P \rightarrow [0, 1]$. The values returned by a function indicate the malignancy of a tumor and are interpreted in the following way:

group	attribute	diagnostic models					
		SM [11] m_1	Alc. [6] m_2	LR1 [8] m_3	LR2 [8] m_4	Timm. [12] m_5	RMI1 [13] m_6
objective medical history	age	-	-	✓	✓	-	✓
	menopausal status	✓	-	-	-	✓	✓
	ovarian cancer in family	-	-	✓	-	-	-
	hormonal therapy	-	-	✓	-	-	-
	hysterectomy	-	-	-	-	-	✓
other	pain during examination	-	-	✓	-	-	-
	lesion volume	✓	-	✓	-	-	-
	internal cyst walls	✓	-	✓	✓	-	-
	septum thickness	✓	-	-	-	-	-
	echogenicity	✓	✓	-	-	-	-
	localisation	✓	-	-	-	-	✓
	ascites	✓	-	✓	✓	-	✓
	papillary projections	-	✓	-	-	✓	-
	solid element size	-	✓	✓	✓	-	✓
	blood flow location	-	✓	✓	✓	-	-
	resistance index	-	✓	-	-	-	-
	acoustic shadow	-	-	✓	✓	-	-
	amount of blood flow	-	-	✓	-	✓	-
	CA-125 blood marker	-	-	-	-	✓	✓
	lesion quality class	-	-	-	-	-	✓

Table 1: Attributes used in the most common preoperative diagnostic models.

- $m(\mathbf{p}) > 0.5$ – diagnosis towards malignant;
- $m(\mathbf{p}) < 0.5$ – diagnosis towards benign;
- $m(\mathbf{p}) = 0.5$ – indicates the impossibility of determining the nature of malignancy.

As mentioned, the existing diagnostic scales operate on complete dataset only. The most straightforward way to generalise them to accept missing values is to extend P to P^* ; then $m : P^* \rightarrow [0, 1] \cup \{\emptyset\}$. Such an approach is simple but not interesting – a scale would not return any diagnosis whenever data were missing.

It seems to be much more appropriate to operate with the interval representation of a patient that was proposed in the previous section. We utilise a classical method of extending real functions to interval values [18] to obtain a new, uncertaintified diagnostic scale $\hat{m} : \hat{P} \rightarrow \mathcal{I}_{[0,1]}$ defined as:

$$\hat{m}(\hat{\mathbf{p}}) = \left\{ m(\mathbf{p}) : \forall_{1 \leq i \leq n} \quad p_i \leq p_i \leq \bar{p}_i \right\} = \left[\min_{\mathbf{p} \in \hat{\mathbf{p}}} m(\mathbf{p}), \max_{\mathbf{p} \in \hat{\mathbf{p}}} m(\mathbf{p}) \right]. \quad (2)$$

where by $\mathbf{p} \in \hat{\mathbf{p}}$ we denote that \mathbf{p} is an embedded vector of $\hat{\mathbf{p}}$.

The resultant interval represents all of the possible diagnoses that can be made based on a patient description in which every missing value was replaced with all possible values for that attribute. The more incomplete the description, the more uncertain the diagnosis. However, it is worth noting that in many cases it is still possible to make a proper diagnosis, since some amount of missing values is acceptable and will not affect the final result significantly. This is one of the basic advantages

that distinguishes our approach from simple P^* modelling.

3.3. Diagnosis modelling

The diagnostic model defined by (2) can conveniently be described within a framework of interval-valued fuzzy set (IVFS) theory. IVFSs were designed to represent uncertain (incomplete) information, in contrast to standard fuzzy sets, which model imprecision. Let X be a finite crisp set called the universe of discourse, and let $\mathcal{I}_{[a,b]}$ be the set of all closed sub-intervals of $[a, b]$, where a and b are real numbers. Then an IVFS is a mapping $\hat{A} : X \rightarrow \mathcal{I}_{[0,1]}$. An interval $\hat{A}(x) = [\underline{A}(x), \bar{A}(x)] \in \mathcal{I}_{[0,1]}$ is understood to contain the true, however incompletely known, membership degree of an element x in \hat{A} . This corresponds to the interpretation of an interval $\hat{m}(\hat{\mathbf{p}})$ which defines the boundaries of a possible diagnosis for a patient.

Thus a diagnostic model \hat{m} can be interpreted as an IVFS on the universe \hat{P} . A value $\hat{m}(\hat{\mathbf{p}})$ denotes an interval membership degree to which a given patient $\hat{\mathbf{p}}$ can be classified as a malignant case according to the model \hat{m} .

A diagnosis in the form of an interval has its advantages and drawbacks. An advantage is that such a model gives a diagnosis even in the presence of missing data. A drawback is that the diagnosis is often uncertain and not so easy for a physician to apply. A major problem is thus how to efficiently support a physician in making a final diagnosis under incomplete information.

In order to solve this problem we make the following observation. As Table 1 shows, different scales,

denoted by m_1, \dots, m_n , use different attributes describing the patient, and are therefore subject to different levels of uncertainty. The main idea is thus to improve the final diagnosis by taking advantage of the diversity of diagnostic scales. Given n scales $\hat{m}_1, \dots, \hat{m}_n$ we construct a function $\text{Agg} : \mathcal{IVFS}^n \rightarrow \mathcal{IVFS}$. Its result $\text{Agg}(\hat{m}_1, \dots, \hat{m}_n)$ is a new IVFS that gathers and integrates information from the input sets. Thanks to this interpretation we immediately see the relationship with the issues of group decision-making and information aggregation [19].

An n -argument interval-valued fuzzy aggregation operator is a mapping $\text{Agg} : \mathcal{I}_{[0,1]}^n \rightarrow \mathcal{I}_{[0,1]}$ with the following properties [20]:

1. if $\hat{y}_i \subseteq \hat{x}_i$ for all $i \in 1, \dots, n$, then $\text{Agg}(\hat{y}_1, \dots, \hat{y}_n) \subseteq \text{Agg}(\hat{x}_1, \dots, \hat{x}_n)$,
2. $\text{Agg}([1, 1], \dots, [1, 1]) = [1, 1]$,
3. $\text{Agg}([0, 0], \dots, [0, 0]) = [0, 0]$.

Recent research has led to the construction of many interval-valued fuzzy sets aggregation methods [21, 22, 20, 23]. The most commonly used aggregation methods in group decision making are based on the weighted arithmetic average [19]. We propose to use different IVFS aggregation methods to improve the quality of diagnosis as well as to minimise the impact of the lack of data and uncertainty in decision-making.

In the medical decision-making problem, the final diagnosis obtained from an aggregation operator must indicate whether a tumor is malignant or not. However, supporting a decision in a case where there is not enough information may lead to a wrong diagnosis. Thus we accept a situation in which no diagnosis recommendation is made. The conversion of interval diagnosis into final diagnosis (binarisation) is very important and may influence overall efficacy.

In view of the importance of matching the aggregation operator to a specific problem, in this paper we propose a methodology that allows one to investigate how different aggregation operators cope with those problems. It makes it possible to obtain the characteristics of the operator and to assess the degree of usefulness in a given decision problem. Since the main issue considered here is the possibility of making decisions under incomplete data, the proposed assessment methodology specifically focuses on this issue.

4. Results and discussion

4.1. Methodology description

The main goal of the proposed methodology is to simulate the decision-making process in a real decision problem. In supporting medical diagnosis, a very important problem is the possibility of missing data. Hence the proposed methodology is aimed at evaluating aggregation operators used to cope with

lack of data. For this purpose, an essential element of the methodology is to simulate different levels of missing data. In order to better reflect the reality we have divided the attributes that describe the patient into two separate groups: those that are subject to obscuration and those that are not. This separation naturally exists in many problems, including the problem of medical diagnosis, because some data about the patient, such as age and other objective data from the medical history, are always available to the physician.

For the purpose of our research we have adopted some assumptions. Firstly, in order to avoid bias, we assume that the dataset on which the evaluation is carried out has an equal distribution of positive and negative cases. To achieve this state, at each iteration we randomly select an equal number of patients within each class. This also allows us to reduce the iteration time and to increase the randomness of the procedure. Secondly, the evaluation can only apply to diagnostic scales based on mathematical models that are simple enough to be efficiently computed after the extension to interval values. Another assumption is that we consider levels of missing data in the range of $0 - 50\%$, because higher levels are unacceptable in practice. The quality of medical diagnosis is most often described by five basic statistics: accuracy (ACC), sensitivity (SEN), specificity (SPEC), positive predictive value (PPV) and negative predictive value (NPV). Therefore, the methodology is aimed at the simple calculation of these parameters.

The evaluation process can be described as follows. For each level of missing data, assessment of the efficacy of aggregation operators is repeated N times. The input data for each repetition is a complete dataset which conforms to the adopted assumptions. In each such iteration two subsets of instances (positive and negative) are drawn independently. Both of them are subjected to a process of obscuration, which involves random erasure of attribute values from the dataset so that the total level of missing data corresponds to the target level of obscuration (OL). The data is next converted to interval form according to equation (1) and passed to uncertaintified diagnostic scales.

The results of all uncertaintified diagnostic scales are represented as IVFSs. The list of those IVFSs forms an input to the aggregation operators. Each operator synthesises the input diagnoses in accordance with its principle of operation. The result of the aggregation is an IVFS representing the synthesised diagnosis. In order to make the final diagnosis it is required to perform a binarisation process. The resulting diagnoses are compared with reference values in order to calculate the necessary statistics. The final statistics for a given level of missing data are calculated by averaging the results of all iterations. A diagram of the evaluation process appears in Fig. 2.

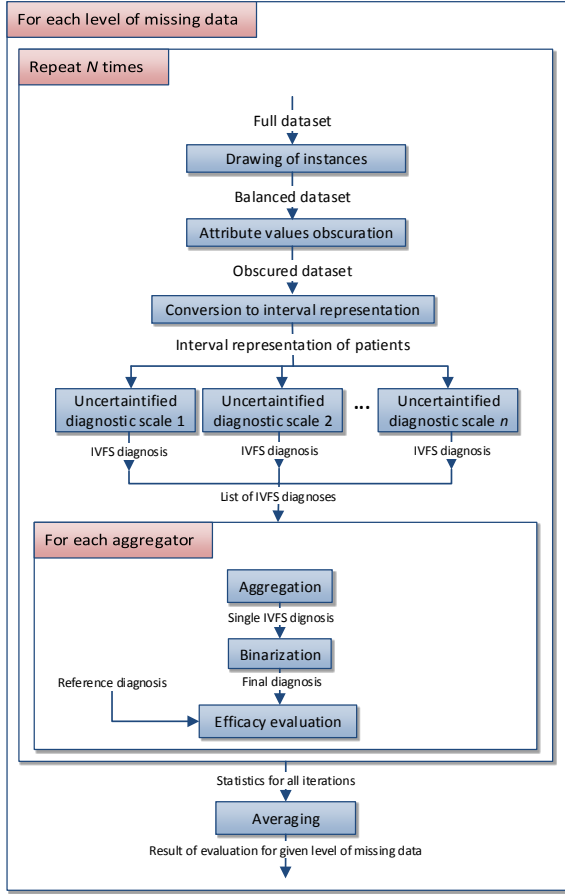


Figure 2: Visualisation of the evaluation process. Data flow is represented as arrows and boxes represent operations on data.

4.2. Evaluation and results analysis

Our study group consisted of 268 women diagnosed and treated for ovarian tumor in the Division of Gynaecological Surgery, Poznań University of Medical Sciences, between 2005 and 2012. Among them, 62% were diagnosed with a benign tumor and 38% with a malignant tumor. At each iteration we selected 50 patients for the positive and negative groups. All patients had no missing values in the attributes required by the diagnostic scales. The dataset is described in detail in [10].

In the evaluation we used six different uncertaintified diagnostic scales $\hat{m}_1, \dots, \hat{m}_6$. These scales were obtained from the basic scales listed in Table 1 by applying the uncertaintification process described by (2).

For the experiment we chose the simplest methods of aggregation, which base on weighted average, sum and intersection of sets, and majority vote. Such methods are most often used in the problem of group decision making [19]. However, the authors are aware that these methods do not cover recent research in that field.

The construction of a certain method of aggregation consists in making choices of aggregation strategy and binarisation strategy. Moreover, methods

ID	aggregation	strategy of weight calc.	binarization
A	Interval avg.	width	margin
B	Interval avg.	entropy	no margin
C	Interval avg.	constant	no margin
D	Lower bound avg.	width	margin
E	Upper bound avg.	width	margin
F	Center avg.	width	margin
G	Center avg.	entropy	margin
H	Intersection	-	no margin
I	Sum	-	no margin
J	Majority vote	-	margin

Table 2: Selected aggregation methods.

based on weighted average require definition of the weight of intervals. We evaluated 177 combinations of the strategies, which will be described below. Many of the resulting methods are similar and have almost equal outcomes. For the sake of the presentation of evaluation results, we chose 10 methods of aggregation which are representative of their groups. These 10 methods are presented in Table 2.

4.2.1. Aggregation strategy

Diagnoses for each patient are independent, it is possible to perform aggregation per patient $\hat{\mathbf{p}}_i$:

$$\hat{m}(\hat{\mathbf{p}}_i) = \text{Agg}(\hat{m}_1(\hat{\mathbf{p}}_i), \dots, \hat{m}_6(\hat{\mathbf{p}}_i)).$$

The first group of aggregation operators (A–C) is based on arithmetic mean with the use of interval arithmetic:

$$\text{Agg}(\hat{x}_1, \hat{x}_2, \dots, \hat{x}_n) = \frac{\sum_{i=1}^n \omega(\hat{x}_i) \times \hat{x}_i}{\sum_{i=1}^n \omega(\hat{x}_i)}.$$

The second group of operators (D–G) is based on a weighted mean calculated with reference to a representative of the interval. Selected strategies in choosing representatives (rep) are the minimum, maximum and centre of an interval:

$$\text{Agg}(\hat{x}_1, \hat{x}_2, \dots, \hat{x}_n) = \frac{\sum_{i=1}^n \omega(\hat{x}_i) \cdot \text{rep}(\hat{x}_i)}{\sum_{i=1}^n \omega(\hat{x}_i)}.$$

The next two operators (H–I) are based on set-theoretical sum and intersection:

$$\text{Agg}(\hat{x}_1, \hat{x}_2, \dots, \hat{x}_n) = \bigcup_{i=1}^n \hat{x}_i$$

and

$$\text{Agg}(\hat{x}_1, \hat{x}_2, \dots, \hat{x}_n) = \bigcap_{i=1}^n \hat{x}_i.$$

The last method (J) differs from the previous ones in that firstly it binarises input intervals, and after that it chooses the diagnosis which appeared more frequently. In the case of a tie, a diagnosis is not made.

4.2.2. Binarisation strategy

In our research we chose the simplest variant of interval binarisation:

$$\tau_{\epsilon}([a, b]) = \begin{cases} 0 & \text{if } b < 0.5 + \epsilon \\ 1 & \text{if } a \geq 0.5 - \epsilon \\ \text{NA} & \text{otherwise.} \end{cases} \quad (3)$$

In this approach, an instance is classified as positive when the whole interval is greater than 0.5 with respect to the margin ϵ . The negative case is defined similarly. In a case where the first or second condition is not met, it is not possible to make a decision. For example, when the margin is set to $\epsilon = 0.025$, a diagnosis of $[0.1, 0.3]$ will be classified as benign, but for the interval $[0.1, 0.6]$ it is not possible to make a decision.

In our evaluation we arbitrarily chose two values for ϵ : 0 (no margin) and 0.025.

4.2.3. Weight calculation strategy

In our evaluation we selected three strategies for choosing weights:

- constant value: $\omega([a, b]) = 1$,
- interval length: $\omega([a, b]) = b - a$,
- normalised interval distance from 0.5 (entropy):

$$\omega([a, b]) = \begin{cases} 0 & \text{if } a \leq 0.5 \leq b \\ 2(a - 0.5) & \text{if } a \geq 0.5 \\ 2(0.5 - b) & \text{otherwise.} \end{cases}$$

4.2.4. Evaluation

In each classification problem it is important to set a baseline. In the case of ovarian tumor diagnosis, this is achieved by means of a classifier which makes a decision using only one attribute *menopausal status* [10]. This classifier achieves 69% accuracy. The aim of this operation is to set a lower bound – all methods of aggregation should be better than a baseline classifier and single diagnostic scales. An aggregation operator which has worse efficacy than the lower bound is useless in practical applications.

Statistical evaluation, as well as implementation the of proposed methodology, were performed using R software, version 3.1.1 [24]. We set levels of missing data to vary from 0% to 50% with a 5% step size. For each level we performed 1000 repetitions of random data obscuration with other calculations. With this number of repetitions the averaged results are stable, so that it is possible to reliably analyse them.

The most significant results are presented in Fig. 3. Diagrams (a) and (b) show how the aggregators and single diagnostic scales perform with an increasing level of missing data. Diagram (a) presents diagnostic accuracy (ACC) and diagram (b) shows the percentage of patients for whom the decision

could be made. The upper and lower bounds of the shaded regions in diagrams (a,b,d) correspond to the largest and the smallest values achieved by the diagnostic scales.

Diagrams (a) and (b) show that preserving high diagnosability frequently prevents models from achieving high accuracy, and *vice versa*. In order to understand the relation between these two parameters, let us present individual aggregation operators on a plane. Diagram (c) shows a relationship between diagnosability and accuracy for each level of missing data. All points of a hypothetical ideal aggregation operator would be localised in the upper right corner of the diagram (100% accuracy and 100% diagnosability). For this reason, a measure of the quality of an aggregation operator for a given level of missing data is defined as the Euclidean distance of such a point in the plane from the point (100%, 100%):

$$q(\text{Agg}_i, OL) = \sqrt{(ACC_{Agg_i}^{OL} - 1)^2 + (DIAG_{Agg_i}^{OL} - 1)^2}.$$

Diagram (d) presents a quality measure q with respect to levels of missing data. Smaller values of q correspond to better diagnostic quality of a model. From the diagram it can be seen that in practice it is impossible to achieve a measure arbitrarily close to 0.

4.2.5. Results summary

The developed methodology led us to the conclusion that in the case of our medical diagnosis problem, aggregation is useful as a tool to solve the problem of missing data. Even the simplest methods presented in this paper achieved an efficacy which exceeded that of the individual diagnostic scales, both in terms of accuracy and the number of diagnosed patients, despite the missing data. There are three interesting cases:

1. the result of aggregation is the achievement of very high and stable accuracy (over 95%, regardless of the level of missing data) at the cost of a small number of patients for whom it was possible to make a diagnosis (below 50%, less than individual diagnostic scales) – see e.g. the *I* operator;
2. the result of aggregation is the achievement of very high diagnosability (over 90%) regardless of the level of missing data, at the cost of decreasing accuracy with increasing level of missing data (accuracy might be even lower than is achieved by individual diagnostic scales) – see e.g. the *E* and *F* operators;
3. the result of aggregation is the achievement of persistent high accuracy comparable to that achieved by individual diagnostic scales, with simultaneous high level of diagnosability (significantly higher than achieved by individual diagnostic scales) – see e.g. *F* i *G*.

In the problem of ovarian tumor diagnosis, it appears that the most promising results were obtained

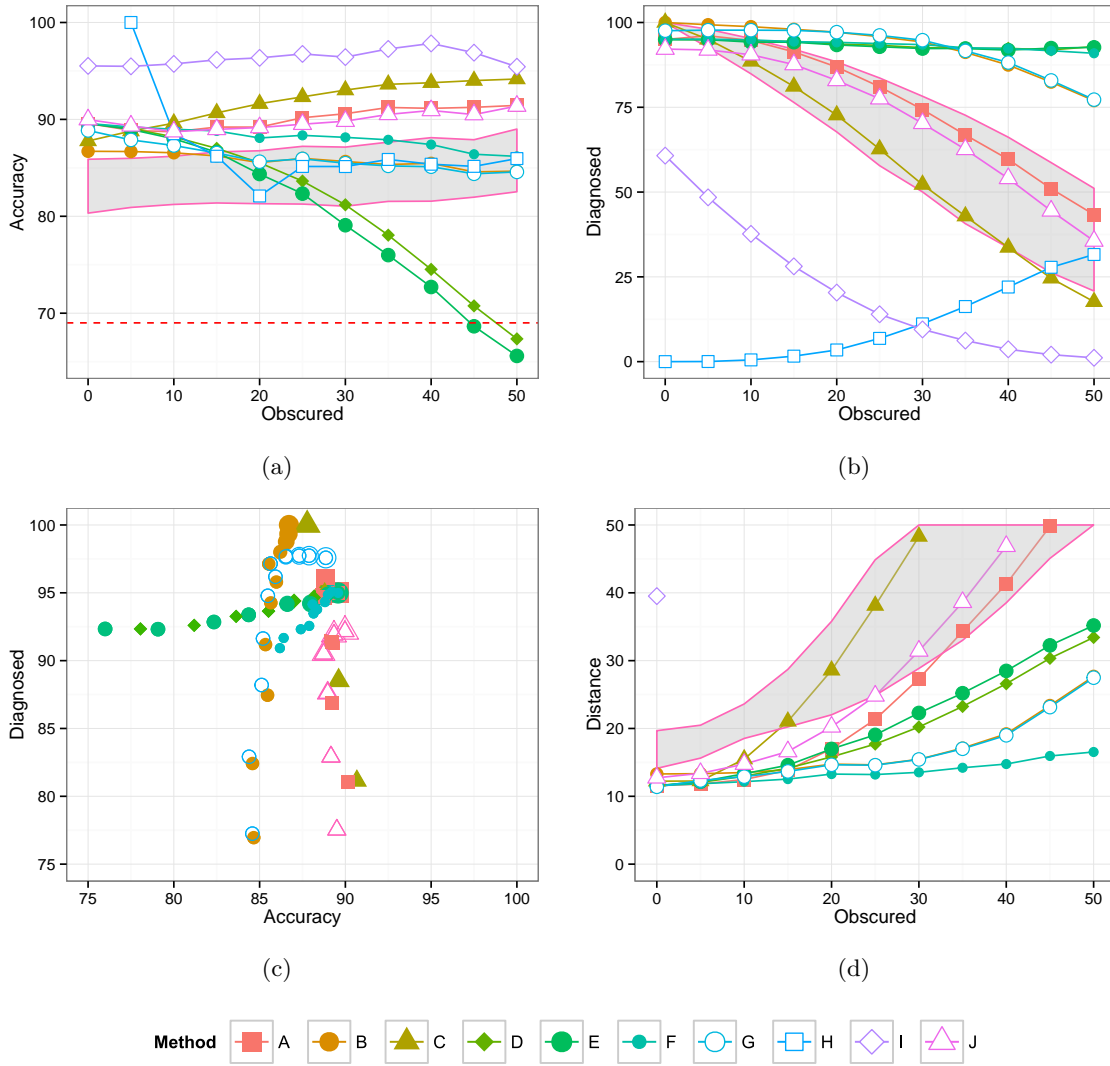


Figure 3: Results of evaluation of selected aggregation methods. The shaded region indicates bounds of single diagnostic methods. The dashed horizontal line in diagram (a) indicates the accuracy of the baseline classifier.

by aggregation operator F , as shown in diagram (d). This operator is capable of maintaining high accuracy and diagnosability. Furthermore, its distance from the optimal effectiveness point is the smallest and increases very slowly with an increase in the level of missing data. Its low sensitivity to lack of data makes it a promising candidate in the search for robust operators.

An interesting result is obtained in the case of the operator H , which is based on set intersection. In the case of complete data it is not able to make any decision. This is because in such a situation the intervals degenerate to single points, and the intersection of such intervals is usually the empty set.

The authors are aware that since the evaluation was performed on the whole dataset with arbitrarily chosen binarisation margins, general conclusions on the performance of the presented aggregation operators should not be drawn. To make the results more reliable, performance should be validated on a separate dataset with optimised aggregation pa-

rameters.

5. Conclusions and further research

The results presented here are promising and show that the competent selection and use of aggregation methods can significantly improve the quality of decisions taken by a diagnostic system. The problem is particularly significant when the knowledge is based on incomplete information. Proper selection of the method of aggregation is essential for reducing the negative impact of data incompleteness on the quality of decisions. Because the design of an aggregation method depends on the particular problem, extensive evaluation is needed on each occasion. This can be done using our proposed method.

This paper proposes a method for the uncertainification of classifiers which may have broader application in the study of the influence of incomplete data on the quality of decisions. This issue will be examined and described in more detail in the future.

In future work, we plan to study a broader range of aggregation methods and determine guidelines on their applicability to various problems. Generally, performance measurement methods depend on the domain of application. Thus we will also focus on other statistics besides accuracy, such as sensitivity and specificity.

It is important to note that there are still many open problems related to the proposed approach. Aggregation methods should be evaluated on many common datasets in order to examine their characteristics and properties. The results obtained by the proposed approach should be compared with those obtained with the use of imputation methods. Finally, it should be checked whether the proposed approach easily scales to the multiclass classification problem.

The results presented in this paper are promising and provide a starting point for a broader analysis of the problem of effective construction of aggregating methods on sources with incomplete data.

References

- [1] A. Pankowska. The Problem of Uncertain and Incomplete Information in Group Decision Making Process. In K. Atanassov et al., editors, *Development in Fuzzy Sets, Intuitionistic Fuzzy Sets, Generalized Nets and Related Topics. Foundations. Vol I.*, pages 265–274. IBS PAN – SRI PAS, Warsaw, 2008.
- [2] R. Scherer. *Multiple Fuzzy Classification Systems*. Springer, 2012.
- [3] K. Dyczkowski, A. Wójtowicz, et al. An intelligent system for computer-aided ovarian tumor diagnosis. In D. Filev et al., editors, *Intelligent Systems’2014*, volume 323 of *Advances in Intelligent Systems and Computing*, pages 335–343. Springer International Publishing, 2015.
- [4] World Health Organization. Mortality database. http://www.who.int/healthinfo/mortality_data/. Online. Accessed on 17/11/2014.
- [5] United Nations. World Population Prospects, the 2012 Revision. <http://esa.un.org/unpd/wpp/>. Online. Accessed on 17/11/2014.
- [6] J. L. Alcázar, L. T. Mercé, et al. A new scoring system to differentiate benign from malignant adnexal masses. *Obstetrical & Gynecological Survey*, 58(7):462–463, 2003.
- [7] D. Timmerman, A. C. Testa, et al. Simple ultrasound-based rules for the diagnosis of ovarian cancer. *Ultrasound in Obstetrics & Gynecology*, 31(6):681–690, 2008.
- [8] D. Timmerman, A. C. Testa, et al. Logistic regression model to distinguish between the benign and malignant adnexal mass before surgery: a multicenter study by the International Ovarian Tumor Analysis Group. *Journal of Clinical Oncology*, 23(34):8794–8801, 2005.
- [9] C. Van Holsbeke, B. Van Calster, et al. External validation of mathematical models to distinguish between benign and malignant adnexal tumors: a multicenter study by the International Ovarian Tumor Analysis Group. *Clinical Cancer Research*, 13(15):4440–4447, 2007.
- [10] R. Moszyński, P. Żywica, et al. Menopausal status strongly influences the utility of predictive models in differential diagnosis of ovarian tumors: An external validation of selected diagnostic tools. *Ginekologia Polska*, 85(12):892–899, 2014.
- [11] D. Szpurek, R. Moszyński, et al. An ultrasonographic morphological index for prediction of ovarian tumor malignancy. *European Journal of Gynaecological Oncology*, 26(1):51–54, 2005.
- [12] D. Timmerman, T. H. Bourne, et al. A comparison of methods for preoperative discrimination between malignant and benign adnexal masses: the development of a new logistic regression model. *American Journal of Obstetrics and Gynecology*, 181(1):57–65, 1999.
- [13] I. Jacobs, D. Oram, et al. A risk of malignancy index incorporating CA 125, ultrasound and menopausal status for the accurate preoperative diagnosis of ovarian cancer. *BJOG: An International Journal of Obstetrics & Gynaecology*, 97(10):922–929, 1990.
- [14] D. Timmerman, L. Valentin, et al. Terms, definitions and measurements to describe the sonographic features of adnexal tumors: a consensus opinion from the International Ovarian Tumor Analysis (IOTA) group. *Ultrasound in Obstetrics and Gynecology*, 16(5):500–505, 2000.
- [15] A. Wójtowicz, P. Żywica, et al. Dealing with Uncertainty in Ovarian Tumor Diagnosis. In K. Atanassov et al., editors, *Modern Approaches in Fuzzy Sets, Intuitionistic Fuzzy Sets, Generalized Nets and Related Topics*, pages 151–158. IBS PAN – SRI PAS, Warsaw, 2014.
- [16] A. Stachowiak, P. Żywica, et al. An interval-valued fuzzy classifier based on an uncertainty-aware similarity measure. In P. Angelov et al., editors, *Intelligent Systems’2014*, volume 322 of *Advances in Intelligent Systems and Computing*, pages 741–751. Springer International Publishing, 2015.
- [17] T. De Waal, J. Pannekoek, and S. Scholtus. *Handbook of statistical data editing and imputation*, volume 563. John Wiley & Sons, 2011.
- [18] R. E. Moore. *Interval analysis*. Prentice-Hall, 1966.
- [19] G. Beliakov, A. Pradera, et al. *Aggregation functions: A guide for practitioners*. Springer, 2008.
- [20] G. Deschrijver and E.E. Kerre. Aggregation operators in interval-valued fuzzy and atanassov’s intuitionistic fuzzy set theory. In *Fuzzy Sets and Their Extensions: Representation, Aggregation and Models*, pages 183–203. Springer, 2008.
- [21] R.R. Yager. Owa aggregation of intuitionistic fuzzy sets. *International Journal of General Systems*, 38(6):617–641, 2009.
- [22] G. Deschrijver and E.E. Kerre. Implicators based on binary aggregation operators in interval-valued fuzzy set theory. *Fuzzy Sets and Systems*, 153(2):229–248, 2005.
- [23] G. Beliakov, H. Bustince, et al. Aggregation for Atanassov’s intuitionistic and interval valued fuzzy sets: The median operator. *Fuzzy Systems, IEEE Transactions on*, 20(3):487–498, 2012.
- [24] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2014.