Development of a Fuzzy-Driven System for Ovarian Tumor Diagnosis

Patryk Żywica[a,*], Krzysztof Dyczkowski[a], Andrzej Wójtowicz[a], Anna Stachowiak[a], Sebastian Szubert[b], Rafał Moszyński[b]

[a]*Department of Imprecise Information Processing Methods*
Faculty of Mathematics and Computer Science, Adam Mickiewicz University
Umultowska 87, 61-614 Poznań, Poland
[b]*Division of Gynecological Surgery*
Poznan University of Medical Sciences, Poland
Polna 33, 60-535 Poznań, Poland

Abstract

In this paper we present *OvaExpert*, an intelligent system for ovarian tumor diagnosis. We give an overview of its features and main design assumptions. As a theoretical framework the system uses fuzzy set theory and other soft computing techniques. This makes it possible to handle uncertainty and incompleteness of the data, which is a unique feature of the developed system. The main advantage of *OvaExpert* is its modular architecture which allows seamless extension of system capabilities. Three diagnostic modules are described, along with examples. The first module is based on aggregation of existing prognostic models for ovarian tumor. The second presents the novel concept of an Interval-Valued Fuzzy Classifier which is able to operate under data incompleteness and uncertainty. The third approach draws from cardinality theory of fuzzy sets and IVFSs and leads to a bipolar result that supports or rejects certain diagnoses.

Keywords: supporting medical diagnosis, ovarian tumor, soft computing, imprecise and incomplete data, fuzzy methods

1. Introduction

One of the most challenging current problems in gynaecology is the correct differentiation of adnexal masses. Early identification of malignant ovarian tumors versus benign neoplasms and functional lesions is crucial, because it determines the necessity of surgery, the pre-operative work-up and adequate timing in the operation room [1]. It also has great importance for determining who should perform the surgery – a gynaecological oncologist or a general gynaecologist. The problem of correct and early diagnosis of this kind of tumor is

*Corresponding author
Email address: bikol@amu.edu.pl (Patryk Żywica)

This is an Accepted Manuscript of an article published by Elsevier in Biocybernetics and Biomedical Engineering on 2016-10-02, available online: https://dx.doi.org/10.1016/j.bbe.2016.08.003

a difficult task especially for inexperienced gynaecologists [1]. Moreover, small medical centers lack specialised equipment for advanced medical examinations. Such deficiencies lead to problems with the collection of complete data by physicians during examinations and with interpretation of the results. This, in turn, hinders the making of a final decision.

Gynaecologists around the world have developed many prognostic models, ultrasonographic morphological scales, and other risk of malignancy calculators that are used for differential diagnosis of ovarian tumors. The most common diagnostic models are based on scoring systems [2, 3] and logistic regressions [4, 5]. Over 10 years ago, the International Ovarian Tumor Analysis (IOTA) Group began a project to improve the ability to differentiate between benign and malignant ovarian tumors. Several years of comprehensive and broad studies resulted in a number of predictive models. Among these, the most important are the LR1 and LR2 models, based on logistic regression [6], and the most recent IOTA model – ADNEX [7]. However, the plurality of diagnostic models confirms their imperfections. Both the sensitivity and specificity of those models rarely exceed 90% in external evaluation by independent research centers [8, 9]. Another limitation is that they cannot be applied when some of the patient data is missing, which is a common problem resulting from, for example, the technical limitations of the health care unit or high costs of medical examinations.

OvaExpert is an intelligent medical diagnosis support system for ovarian tumor, intended as a solution to the problem of effective diagnosis in the presence of low-quality (uncertain and incomplete) data. It is being developed by scientists from two Polish universities: Adam Mickiewicz University in Poznan and Poznan University of Medical Sciences. The main aim of *OvaExpert* is to equip a physician with a convenient tool that makes it possible:

- to gather and manage a patient's data in a standardized format;
- to reduce the impact of low data quality on the final diagnosis;
- to present the result in a way that gives maximum information to the doctor.

Currently *OvaExpert* is in the testing phase, and its demo version is available at the project website `http://ovaexpert.pl/en` to provide insight into all of the functions of the system described below. The system is easy to use and intuitive, yet it utilizes recent methods mainly from the areas of machine learning, soft computing and fuzzy set theory [10].

In the following sections we describe the system in detail, focusing on its modular architecture. We present the main features and components of *OvaExpert*, namely the diagnostic modules, and some of their theoretical background.

2. Design of the System

The following section describes the features of *OvaExpert* along with its place and role in the diagnostic process. The third part of this section contains technical and technological details.

2.1. Features

OvaExpert is intended to integrate present knowledge about ovarian tumors (models, scoring systems, reasoning schemes, etc.) into a single computer-based system. It is a unique tool for many reasons. To the best of our knowledge this is the first time that incompleteness of data has been taken into account and incorporated into a system for ovarian tumor diagnosis in a comprehensive manner [11]:

- at the stage of collecting data about a patient;
- at the stage of data processing;
- and finally, at the stage of presenting the diagnosis.

We briefly introduce the principles of the system and its main elements. The system covers four main areas: medical data acquisition, expert knowledge gathering, decision support in the process of selecting the optimal diagnostic path, and decision support in making a final diagnosis (see Figure 1).

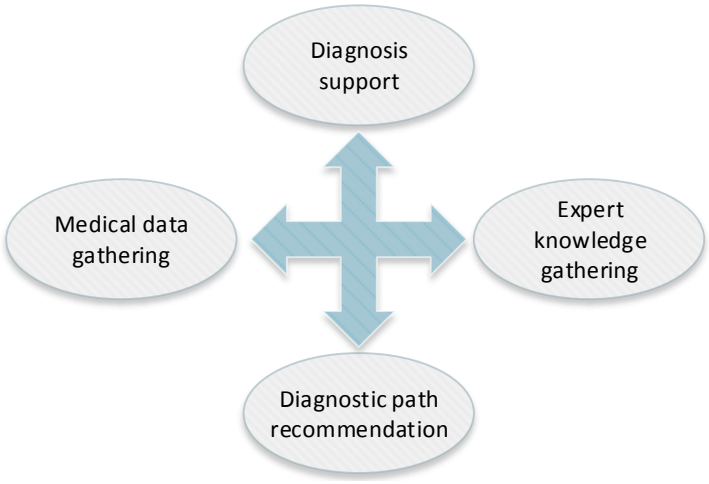

Figure 1: The main components of the *OvaExpert* system.

The system was created for physicians who need a comprehensive tool that is able to offer more than support for a final diagnosis. *OvaExpert* assists the gynaecologist during the whole diagnostic process, from gathering medical data to exchanging data and experience among physicians. The following subsections describe the features of the system which enable the technical realisation of these objectives.

2.1.1. Uncertainty Modelling

Uncertainty has attracted increasing attention as an important problem in health care practice and in medical publications. As noted in [12] there are

multiple meanings and varieties of uncertainty in medicine, each of them having unique effects for diagnosis. Types of uncertainty can be distinguished according to its nature – whether it is objective (arising from the complex or probabilistic nature of a phenomenon), subjective (personal opinion or interpretation), or a result of low quality of information (incompleteness).

Working under information uncertainty is an everyday experience in medical practice, and it is impossible to eliminate it completely. However, many tools that support gynaecologists neglect this problem and shift the responsibility for providing good-quality data to the doctor. A different approach is proposed in the *OvaExpert* system.

OvaExpert introduces a completely new approach to the uncertainty related to incompleteness and lack of data [10, 13]. The aim of the system is to store and process uncertain data, so as to extract as much information as possible and to present the resultant diagnosis still retaining information about the level of uncertainty. For example, in addition to precise examination results, uncertain values can also be handled. This can occur if the physician is not sure of the result or if results are ambiguous. For example, if the tumor thickness in ultrasound examination is between 30 and 50 mm, such values can be stored and handled in the system. The presentation of uncertain results also depends on the type of a given attribute. In the case of integers or decimals, the result is an interval, whereas for boolean attributes one may indicate that both values are possible. For attributes with a list of possible values (enums) one may indicate several possible examination outcomes.

2.1.2. Medical Data Gathering

One of the main objectives of the *OvaExpert* system is to provide a simple, convenient and efficient way of collecting patient data and reaching a final diagnosis. Currently, due to the absence of a common data format, there are limitations in cooperation between physicians from different centers. Some data may also be lost when converting one data format to another. To date, data has been collected by individual doctors using traditional methods, such as spreadsheets or notebooks, without paying sufficient attention to its quality and format. The system, by providing standardised data schemas developed on the basis of the recommendations of the IOTA group [14], enables the collection of data in a common, centralised database. This has made it possible to begin building a knowledge base of different medical cases. This database also enables quality assessment of the diagnostic decisions taken by the system, performed by specialists from different medical centers, and the collection of data for further scientific research.

2.1.3. Simple and Intuitive Interface

The design of the interface was the subject of thorough consultation with gynaecologists to meet the need for ease of use in all conditions, including on mobile devices, especially smartphones. An example screenshot of the *OvaExpert* interface appears in Figure 2. At any time, the attending physician can be provided with the history and a visualisation of the patient's diagnostic process.

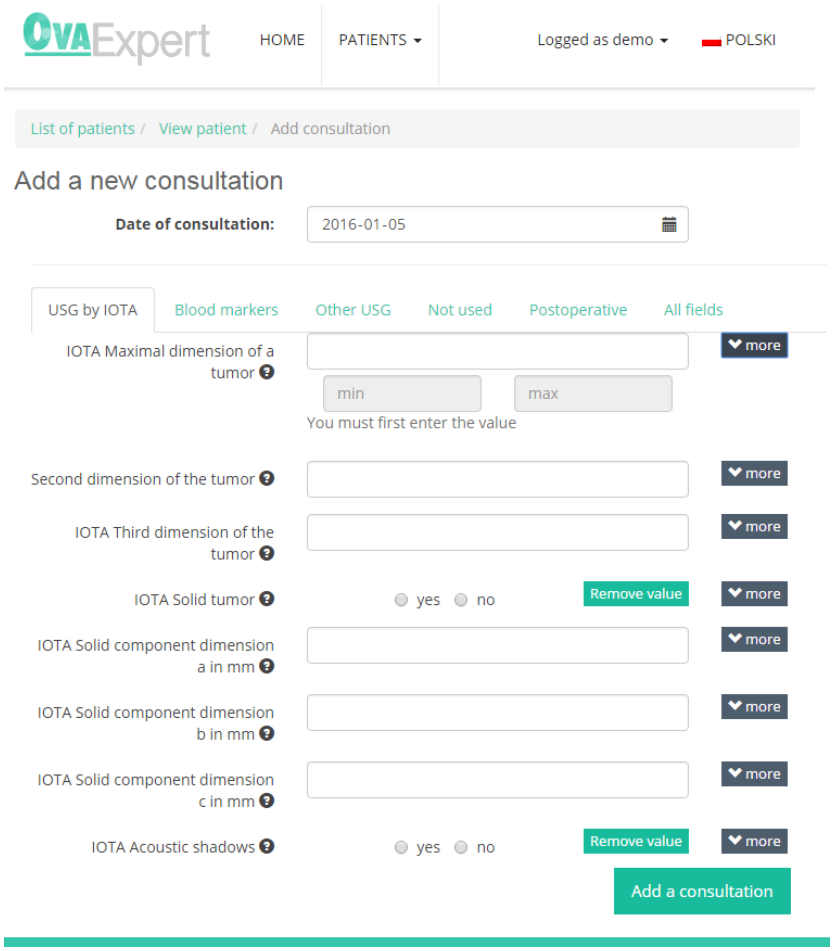

Figure 2: An example screenshot of the *OvaExpert* interface.

During the whole process the gynaecologist is accompanied by a system that supports him or her by identifying further examinations which may increase the likelihood of an accurate diagnosis. Such a solution is of great assistance to inexperienced gynaecologists, and moreover helps to avoid unnecessary examinations and the associated costs.

2.1.4. Variety of Diagnostic Models

OvaExpert is designed to provide classical diagnostic models as well as new ones. First of all, well-known prognostic models were implemented, including the models of the IOTA group, namely the SM scale [3], Alcazar scale (further denoted as Alc.) [2], IOTA LR1 model [4], IOTA LR2 model [4], Timmerman model (Tim.) [15] and Risk of Malignancy Index (RMI) [5]. Since many gynecologists are familiar with those methods and trust their results, *OvaExpert*

makes it possible to use them at any time (including their uncertaintified versions) and to compare their results. Moreover, modern diagnostic methods were proposed and some of them were implemented, including those based on fuzzy aggregation and interval-valued fuzzy classification. The modular architecture of the system makes it easy to extend it with new methods in the future.

2.1.5. Bipolar Diagnosis

OvaExpert presents the result of the diagnostic process in a bipolar manner [11], giving simultaneously the possibility of diagnosis towards malignant and towards benign, together with a degree of impossibility of determining the nature of malignancy. Such a presentation informs the physician about the reliability and completeness of the diagnosis. A classical approach to the medical diagnostic process involves identifying the most adequate diagnosis. However, it is also possible to follow criteria that exclude certain diagnoses. It is apparent that in case of doubts regarding the diagnosis, such a bipolar – positive and negative – perspective is valuable and conveys more information to the doctor.

OvaExpert uses an approach based on Atanassov's intuitionistic fuzzy sets (IF-sets) [16, 17] to model bipolarity in the diagnostic process. This concept is innovative in medicine, its use in diagnosis having only been indicated as a possibility [18, 19]. It is consistent with the basic premise of the *OvaExpert* system, that it must be able to accept and to cope with uncertainty. The patient's condition is described on the one hand by a degree to which the tumor is regarded as malignant, and on the other by a degree to which it is regarded as benign. These two degrees need not sum to 100%, and the system may suggest further examinations to increase the reliability and completeness of the diagnosis.

IF-sets representation which is mathematically equivalent to Interval-Valued Fuzzy Sets (IVFS) allows us to highlight the component of missing information. Thanks to this approach, where the pros and cons do not necessarily add up to 1, the amount of missing data can be visualised. An IF-set $\mathcal{E}$ is a triple:

$$\mathcal{E} = (A^+, A^-, A^?),$$

where A^+ is a fuzzy set of elements that belong to $\mathcal{E}$, and A^- is a fuzzy set of elements that do not belong to $\mathcal{E}$. This theory, in contrast with fuzzy set theory, incorporates uncertainty about the membership of an element, as A^- is not necessarily a negation of A^+, but $A^- \subset (A^+)^c$ where the complement of fuzzy set A is defined as $A^c(x) = 1 - A(x)$ for each x (for a generalised definition of IF-set with a use of a complement generated by an arbitrary strong negation see e.g. [20]). Therefore, the value $A^?(x) = 1 - A^+(x) - A^-(x)$ reflects uncertainty or hesitation about membership of an element x in IF-set $\mathcal{E}$.

2.1.6. Expert Knowledge Gathering

OvaExpert will also enable a user to enter his or her own diagnostic rules derived from personal experience. These rules take the form of IF-THEN clauses with both numerical (e.g. based on the level of a particular blood marker) and

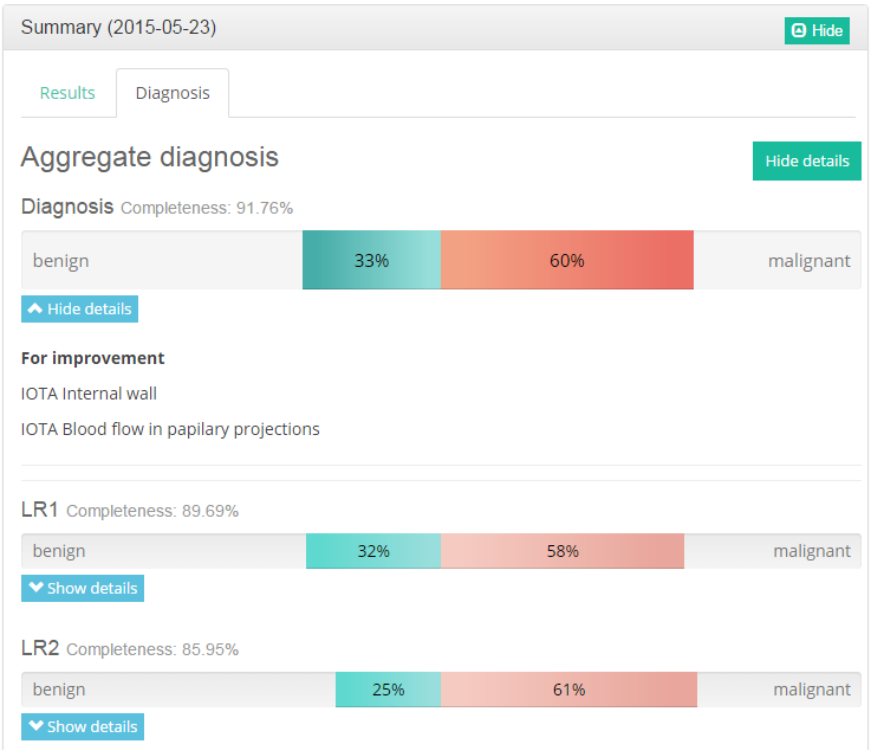

Figure 3: Screenshot of the *OvaExpert* diagnosis recommendation view.

linguistic terms. This means that an individual specialist is able to personalise the behaviour of the system, which may lead to expansion of the knowledge contained in the system. Potentially this can lead to an increase in the effectiveness of diagnoses. This feature is still under development.

2.2. Use of the System in the Diagnostic Process

The diagnostic process begins with a medical history entered by the doctor. It then continues through several stages, where further medical examinations are made. Each of these stages provides additional knowledge for the diagnostic process (e.g. levels of blood markers, ultrasound descriptions, etc.). The process runs iteratively. At each stage *OvaExpert* computes a bipolar recommendation for diagnosis, which consists of a set of possible diagnoses for a particular type of tumor, with an indication of how probable it is that the diagnosis is correct and how probable that it is not (see Figure 3). On the basis of that recommendation the doctor must decide whether to carry out further tests or to make a final diagnosis. The system supports the selection of the optimal diagnostic path. This is done by utilising knowledge from retrospective data (e.g. statistical methods) as well as from fuzzy rules entered into the system by experts. The process of interaction with the physician is illustrated in Figure 4.

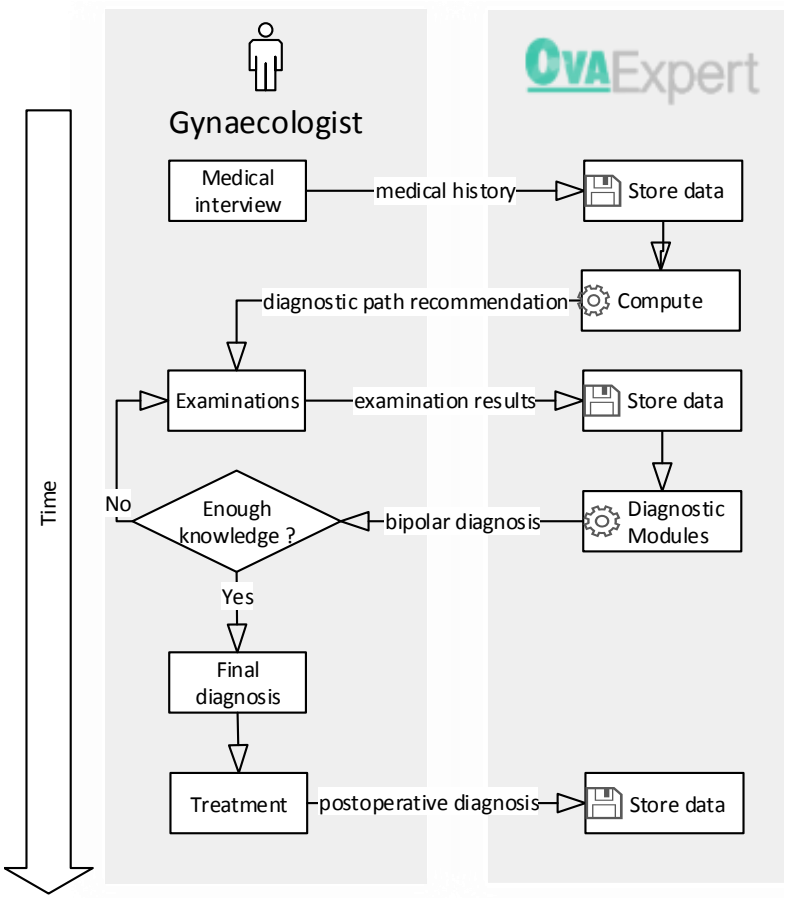

Figure 4: Diagram of interaction between a physician and *OvaExpert*.

2.3. Architecture

A very important aspect in the design of the software is the choice of technology. The main factors that were taken into account when designing the *OvaExpert* system were stability, reliability, and usability on mobile devices. In addition, the chosen technologies need to be up to date and in line with modern standards of software development. We were mainly interested in open source software supported by a large community.

OvaExpert was built with the use of modern software engineering tools and technologies such as Java, Spring, AngularJS and PostgreSQL. The system uses a client–server approach and is available via a web browser based on RESTful web services. In the following, we will give more details on the chosen technologies and their role in the project.

The server is deployed on the Apache Tomcat web container, which makes it possible to run a Java-based HTTP server. As the main framework we selected

Spring, which was used to establish the main components of the application:

- object lifecycle management and dependency injection using the Spring IoC container;
- RESTful web services and enforcement of Model View Controller (MVC) design pattern using Spring MVC;
- system security using Spring Security;
- seamless integration with the database using Spring Object-Relational Mapping (ORM).

As a database management system (DBMS) we chose the free and open source PostgreSQL, which is known for its reliability and supports data integrity.

The client side was created using JavaScript and HTML with the use of the AngularJS framework and Bootstrap CSS. The client and server communicate using RESTful web services and the JSON data format. The client application is available for recent web browsers and is optimised for different screen sizes (desktop, tablet, phone).

In addition to the above, the *OvaExpert* system uses several other libraries to implement specific features. Import from and export to other data formats were implemented with the use of the Apache POI and OpenCSV libraries. Internationalisation of the application was achieved with the *angular-translate* library. The implementation of medical data anonymization uses the *angular-cryptography* library.

An overview of the system architecture is given in Figure 5, which shows the most important parts of the system along with the technologies used to create them. Two components deserve special attention with regard to the further use of the system in medical practice. Medical data anonymization enables the convenient use of sensitive data in the system by both gynaecologists and researchers. The diagnostic modules realise the main objective of the *OvaExpert* project. Here medical data anonymization will be described from a technical point of view. The theoretical basis of the diagnostic modules will be addressed in the next section.

2.3.1. Medical Data Anonymization

Personal data (especially medical records) is sensitive information which is legally protected. In order to conduct research on data collected by the system it must be anonymized. *OvaExpert* has the ability to anonymize medical data automatically in such a way that sensitive data is never sent to the server, but doctors can still access it.

Figure 6 presents the process of medical data storage. The gynecologist enters all medical data into the client application accessed via a web browser. Then automatic anonymization occurs, the data is split into two parts and a unique ID number is generated. Anonymized medical data is stored on the *OvaExpert* server, while sensitive personal information is encrypted and stored locally

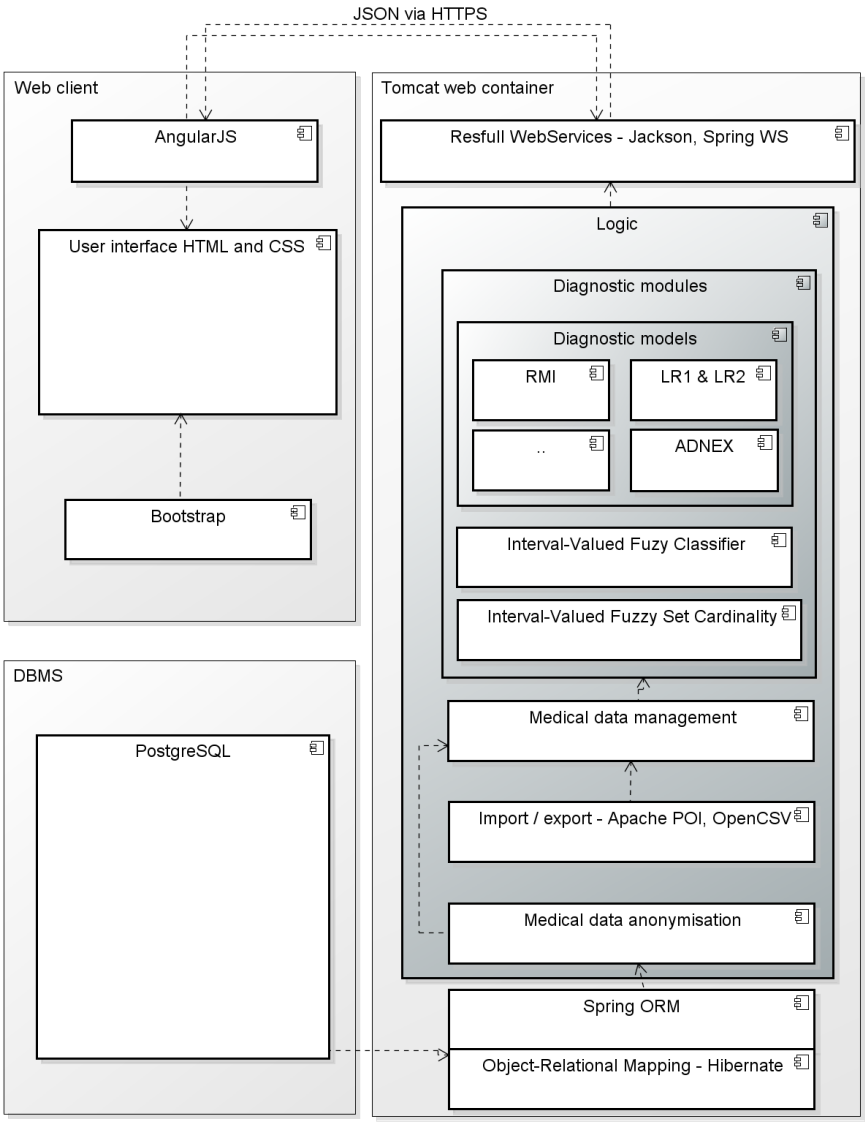

Figure 5: UML diagram showing the most important parts of the *OvaExpert* system.

in the browser's Local Storage. This solution means that sensitive information stays on the doctor's computer and is available only to authorized medical unit employees. Stored sensitive information is as secure as the computer on which it is stored. On the other hand, researchers are able to use anonymized collected medical data to develop new diagnostic models and improve existing ones. When a physician wants to retrieve data from the *OvaExpert* system, sensitive information from the browser's Local Storage is decrypted and merged

with the anonymized medical data from the server with the use of the unique ID.

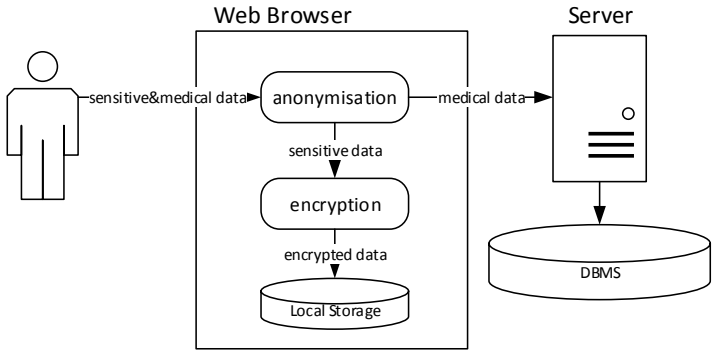

Figure 6: Principle of operation of the proposed method of medical data anonymization.

3. Diagnostic Modules

The main advantage of the system is its modular architecture, which will be discussed in this section. In order to provide a physician with the best recommendation for a final diagnosis, all existing and new methods for supporting the diagnosis of ovarian tumors can be integrated into the system as modules. Currently, three top-level diagnostic modules are implemented: *based on diagnostic models*, *based on an Interval-Valued Fuzzy Classifier* and *based on Interval-Valued Fuzzy Set Cardinality*. Moreover, the system is designed so that it is possible to add new diagnostic modules that provide a diagnosis using techniques completely different from those currently used.

We decided not to develop new models based on classic methods of automated data classification, e.g. SVMs and Bayes classifiers. Such models have been already studied [9] but they are not used nor recognised by the physicians community – this is mostly caused by complexity as well as nontrivial interpretation of the models. For that reason our models base on classifiers which are acknowledged by the community and we utilise a fuzzy-driven methodology due to its natural and approachable way of explanation.

Various decision-making modules are available only for experts who are involved in the process of developing the system. OvaExpert is designed for less experienced physicians and they will be allowed to use only one optimised decision model.

3.1. Module Based on Diagnostic Models Aggregation

The number of different diagnostic models is large (see Introduction, [21]) and it is not generally agreed which one should be used in a particular situ-

ation. Moreover, the original models were not designed to handle incomplete data, while incompleteness is common in medical practice. Thus, the greatest challenge was to support a physician in making an effective final diagnosis under incomplete information.

One of the proposed approaches is to take advantage of the diversity of diagnostic models and to aggregate their results to benefit from synergy effects. Our previous research has shown that fuzzy aggregation is a powerful method to improve the quality of diagnosis as well as to minimise the impact of missing data and uncertainty. The results of a representative instance of this approach can be seen in Figure 7. It is based on Ordered Weighted Average aggregation (OWA, see [22]), marked on the diagram as *OEA*. It achieved an efficacy exceeding that of the individual diagnostic models, despite the missing data. More details concerning diagnostic model aggregation, its evaluation and preliminary results can be found in the original paper [23]. Below we will give a very simplified example of the operation of this diagnostic module.

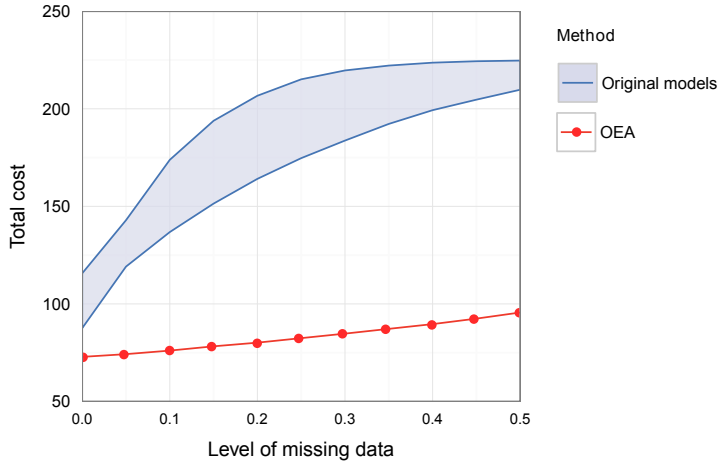

Figure 7: A comparison of total costs between the original diagnostic models and the selected aggregation strategy. The shaded area indicates bounds of the total cost of the original models. All models are evaluated on the same patient data. Section 4 gives more details about evaluation procedure.

For simplicity, in this example we assume that the patient is described by two attributes only, namely patient's age and one cancer antigen test. We define the domains of these attributes as $D_1 = [0, 100]$ and $D_2 = [0, 1500]$. Consider two patients $\mathbf{p}^A = (p_1^A, p_2^A) = (35, 100)$ and $\mathbf{p}^B = (p_1^B, p_2^B) = (60, 1200)$. Let m_1 be a simple example diagnostic model defined by

$$m_1(\mathbf{p}) = 0.0025p_1 + 0.0005p_2\,, \tag{1}$$

which gives results in $[0, 1]$, where values above 0.5 indicate a diagnosis towards malignancy. Now we can easily see that according to diagnostic model m_1

patient A should be diagnosed as benign ($m_1(\mathbf{p}^A) = 0.138$) and patient B as malignant ($m_1(\mathbf{p}^B) = 0.75$).

Now suppose that some patients' values are missing: $\mathbf{p}^A = (35, \text{NA})$ and $\mathbf{p}^B = (\text{NA}, 1200)$. In our approach we define a new interval representation of patients

$$\hat{\mathbf{p}}^A = (\hat{p}_1^A, \hat{p}_2^A) = ([35, 35], [0, 1500]) \tag{2}$$

$$\hat{\mathbf{p}}^B = (\hat{p}_1^B, \hat{p}_2^B) = ([0, 100], [1200, 1200]) \tag{3}$$

and compute diagnoses from *uncertaintified* models using the following formula

$$\hat{m}(\hat{\mathbf{p}}) = \{m(\mathbf{p}) : \mathbf{p} \text{ is such that } \forall_{1 \leq i \leq n}\ p_i \in \hat{p}_i\}\ . \tag{4}$$

This results in

$$\hat{m}_1(\hat{\mathbf{p}}^A) = \left\{m_1(p_1^A, p_2^A) : p_1^A = 35, p_2^A \in [0, 1500]\right\} = [0.088, 0.838] \tag{5}$$

and, analogously, $\hat{m}_1(\hat{\mathbf{p}}^B) = [0.6, 0.85]$. It is easy to see that for the first patient it is hard to make a diagnosis, while for the second, despite the missing data, we can still say with high confidence that she has a malignant tumor.

To illustrate the next step, aggregation, let us assume that there is a new blood marker ($D_3 = [0, 100]$) and it is used in a new diagnostic model

$$m_2(\mathbf{p}) = 0.0025p_1 + 0.0075p_3\,. \tag{6}$$

New marker results were assessed for both patients with the following results: $\hat{\mathbf{p}}^A = (35, NA, 5)$ and $\hat{\mathbf{p}}^B = (NA, 1200, 90)$. The new diagnostic model (after *uncertaintification*) yields $\hat{m}_2(\hat{\mathbf{p}}^A) = [0.125, 0.125]$ and $\hat{m}_2(\hat{\mathbf{p}}^B) = [0.675, 0.925]$.

Having two different pieces of information, we can try to merge them into one that is more reliable. What we know about the first patient is that the diagnostic models yielded $[0.088, 0.838]$ and $[0.125, 0.125]$ as suggested diagnoses. The simplest and naive method of aggregation uses the mean calculated by interval arithmetic. Calculation gives the following results:

$$\hat{Agg}\left(\hat{m}_1(\hat{\mathbf{p}}^A), \hat{m}_2(\hat{\mathbf{p}}^A)\right) = \left[\frac{0.088 + 0.125}{2}, \frac{0.838 + 0.125}{2}\right] = [0.107, 0.482] \tag{7}$$

and

$$\hat{Agg}\left(\hat{m}_1(\hat{\mathbf{p}}^B), \hat{m}_2(\hat{\mathbf{p}}^B)\right) = \left[\frac{0.6 + 0.675}{2}, \frac{0.85 + 0.925}{2}\right] = [0.638, 0.888]\,. \tag{8}$$

In this simple example, thanks to the use of aggregation we have obtained new diagnoses which are less uncertain and make it easier to take a final decision.

3.2. Module Based on an Interval-Valued Fuzzy Classifier

The next diagnostic module implements the novel concept of an Interval-Valued Fuzzy Classifier based on an uncertainty-aware similarity measure [24].

The main idea is to preserve full information – including the uncertainty factor – about the data during the classification process. The classifier is designed to deal with situations in which both the classified objects as well as the classes themselves are imprecise, subjective and/or incomplete. In such cases, the resulting classification will also be imprecise or incomplete.

There are two ways to divide patients into classes. A basic, binary, classification discriminates two kinds of tumor: malignant and benign. A multi-class classification allows more sophisticated discrimination into histopathological types of tumor. For each class, one prototype vector, which represents the entire class, is constructed. We assume that class prototypes, as well as the objects to be classified (patients), are coded as interval-valued fuzzy sets (IVFS, see [25, 17]) and that their attributes are normalised to the interval $[0, 1]$. Then the assignment of the i-th patient to classes labelled by $\{c_1, \cdots, c_m\}$ can be represented as follows:

$$\widetilde{A}_i = \sum_{1 \leq j \leq m} {}^{\hat{sim}(\widetilde{C}_j, \widetilde{P}_i)}/_{c_j} \tag{9}$$

where $\hat{sim}$ is an uncertainty-aware similarity measure, and $\widetilde{C}_j$ and $\widetilde{P}_i$ denote interval-valued fuzzy set representations of a particular class c_j and patient respectively. This approach was discussed in detail in [24].

The crucial issue for this approach is the method of constructing prototypes. Prototypes can be formed from data, for example by using clustering algorithms such as k-means, or can result from the application of expert knowledge. Thus the proposed method provides a valuable opportunity to integrate knowledge acquired from data and from an expert in a single tool. Currently used prototypes were created with the help of gynaecologists based on common medical knowledge.

In the following we illustrate the use of the Interval-Valued Fuzzy Classifier as a diagnostic module in *OvaExpert*. The objective is to assign the best matching histopathological profile of a tumor using the data available before an operation. Both patient and histopathological profiles are coded as IVFSs. For the purpose of the example, we will present only four histopathological types. Two of them were benign – *Endometrioid cyst* and *Mucinous cystadenoma* – and two malignant – *Serous adenocarcinoma* and *Undifferentiated carcinoma*. These types are further referred as HP 1, HP 6, HP 21 and HP 25 respectively. Let us choose five arbitrary patient attributes: age, size of papillary projections (PAP), blood serum levels of CA-125 and HE4 tumor markers, and resistive index (RI). These attributes may be more or less subjective or imprecise. Moreover, some data may not be available at all. A patient's age is known precisely, while blood serum levels of tumor markers are subject to some uncertainties. Resistive index and size of papillary projections are subjective attributes, thus their value is uncertain. Moreover, for technical, medical or financial reasons, values of the last three attributes may not be known. Example histopathological profiles and patient data are presented in Tables 1 and 2. Note that all attributes' values are scaled to the unit interval and that patients' missing values are represented by the $[0, 1]$.

HP type	AGE	PAP	CA125	HE4	RI
HP 1	[0.27, 0.64]	[0.00, 0.27]	[0.00, 0.04]	[0.00, 0.03]	[0.49, 0.78]
HP 6	[0.29, 0.72]	[0.00, 0.14]	[0.00, 0.18]	[0.01, 0.06]	[0.22, 0.83]
HP 21	[0.47, 0.76]	[0.00, 0.52]	[0.30, 1.00]	[0.12, 0.90]	[0.23, 0.56]
HP 25	[0.39, 0.77]	[0.00, 0.58]	[0.15, 0.98]	[0.04, 0.62]	[0.27, 0.45]

Table 1: Profiles of ovarian tumor histopathological type coded as IVFS.

Postoperative diagnosis	AGE	PAP	CA125	HE4	RI
HP 21	[0.62, 0.62]	[0.00, 0.25]	[0.95, 1.00]	[0.95, 1.00]	[0.00, 1.00]

Table 2: Patient profile coded as IVFS.

A classification using the Interval–Valued Fuzzy Classifier can be computed. By definition, the patient's classification is the following:

$$\begin{aligned}\widetilde{A}_1 =& {}^{\hat{sim}(\widetilde{P_1},\widetilde{hp_1})}/_{hp_1} + {}^{\hat{sim}(\widetilde{P_1},\widetilde{hp_6})}/_{hp_6} + \\ & {}^{\hat{sim}(\widetilde{P_1},\widetilde{hp_{21}})}/_{hp_{21}} + {}^{\hat{sim}(\widetilde{P_1},\widetilde{hp_{25}})}/_{hp_{25}}\,. \end{aligned} \tag{10}$$

We use the classical Jaccard index

$$sim(A,B) = \frac{|A \cap B|}{|A \cup B|} = \frac{\sum_i \min(A(x_i), B(x_i))}{\sum_i \max(A(x_i), B(x_i))} \tag{11}$$

to build the uncertainty-aware similarity measure

$$\hat{sim}(\widetilde{A},\widetilde{B}) = \left[\min_{\substack{\underline{A}(x_i)\le A(x_i)\le \overline{A}(x_i) \\ \underline{B}(x_i)\le B(x_i)\le \overline{B}(x_i)}} sim(A,B), \max_{\substack{\underline{A}(x_i)\le A(x_i)\le \overline{A}(x_i) \\ \underline{B}(x_i)\le B(x_i)\le \overline{B}(x_i)}} sim(A,B) \right]. \tag{12}$$

The interval membership of a patient in class HP 1 is calculated as a minimum and a maximum of

$$\frac{\sum_i \min(a_i, b_i)}{\sum_i \max(a_i, b_i)}\,. \tag{13}$$

where a_i, and b_i satisfy:

$$\begin{cases} 0.62 \le a_1 \le 0.62 \\ 0.0 \le a_2 \le 0.25 \\ 0.95 \le a_3 \le 1.0 \\ 0.95 \le a_4 \le 1.0 \\ 0.0 \le a_5 \le 1.0 \end{cases} \quad \text{and} \quad \begin{cases} 0.27 \le b_1 \le 0.64 \\ 0.0 \le b_2 \le 0.27 \\ 0.0 \le b_3 \le 0.04 \\ 0.0 \le b_4 \le 0.3 \\ 0.49 \le b_5 \le 0.78 \end{cases}. \tag{14}$$

The final patient diagnosis represented by IVFS is the following:

$$\widetilde{A} = {}^{[0.07,0.48]}/_{hp_1} + {}^{[0.08,0.52]}/_{hp_6} + {}^{[0.21,0.99]}/_{hp_{21}} + {}^{[0.13,0.90]}/_{hp_{25}}\,. \tag{15}$$

The interval–valued score towards class HP 1 is [0.07, 0.48], which is low. By contrast, the score for class HP 21 is [0.21, 0.99].

3.3. Module Based on Interval-Valued Fuzzy Set Cardinality

In this section we present a new approach to making decisions in an Interval-Valued Fuzzy Set (IVFS) environment based on cardinality. The natural way for humans to make decisions based on many sources (or many experts) is a strategy of counting. By counting, people determine how many sources or experts vote for and how many vote against a given option, and then they choose the decision for which most of them have voted. Since the decision in our case is taken on the basis of multiple source decisions represented as intervals, it seems natural to estimate the maximum and minimum possible confidence for a particular decision. Such an approach suggests the use of the cardinality of IVFSs representing the limits of both intervals: supporting and rejecting the decision.

To be able to describe our method we have to introduce some basic concepts from the theory of cardinality of IVFSs. A comprehensive compendium on this theory can be found in [17]. Another example of an application of fuzzy set cardinalities with weighting functions in decision-making can be found in [26].

A scalar cardinality ($Sigma\,f - Count$) of an IVFS $\widetilde{A}$ can be defined as an interval $sc_f(\widetilde{A}) = [\sigma_f(A_l), \sigma_f(A_u)]$ where

$$\sigma_f(A) = \sum_{x \in supp(A)} f_t(A(x)) \tag{16}$$

and $f : [0,1] \to [0,1]$ is a weighting function (sometimes called cardinality pattern) fulfilling the conditions $f(0) = 0$, $f(1) = 1$ and $f(x) \leq f(y)$ whenever $x \leq y$. The weighting function plays a crucial role in computing a cardinality. It is worth noting that when $f = id$ the cardinality is equal to the sigma-count defined by Zadeh in [27]:

$$\sigma_f(A) = \sum_{x \in supp(A)} A(x) \tag{17}$$

In this article, for simplicity, we will use two basic cases of weighting functions:

- The function $f_{1,t}$, resulting in a cardinality called *counting by thresholding* (see Figure 8)

$$f_{1,t}(x) = \begin{cases} 1, & \text{if } x \geq t, \\ 0 & \text{otherwise} \end{cases} \tag{18}$$

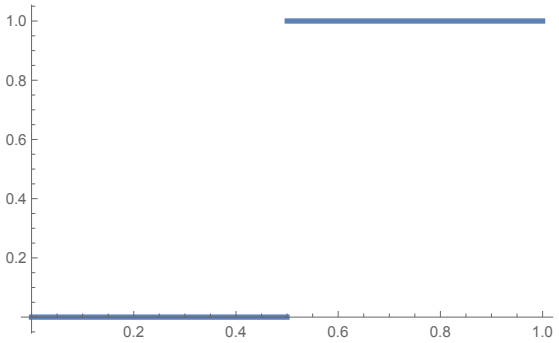

Figure 8: Weighting function $f_{1,t}$for $t = 0.5$

- The function $f_{2,t}$, resulting in a cardinality called *counting by thresholding and joining* (see Figure 9)

$$f_{2,t}(x) = \begin{cases} x, & \text{if } x \geq t, \\ 0 & \text{otherwise} \end{cases} \tag{19}$$

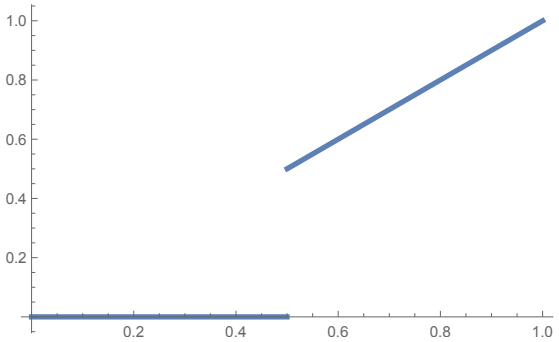

Figure 9: Weighting function $f_{2,t}$ for $t = 0.5$

where $t \in (0, 1]$ is a threshold determining whether we count elements or not.

As in the previously described method, we use existing diagnostic models as the sources for our decision. The data used in the diagnostics modules can contain missing values. So in this approach, a patient diagnosis is represented by an IVFS $\widetilde{A} = (A_l, A_u)$ with $A_l \subset A_u$ in the universe of N applied diagnostic models. Fuzzy sets A_l and A_u represent respectively the lower and upper bounds of the decisions given by the models. These values can be interpreted as follows: a value closer to 0 means a decision towards benign, a value closer to 1 means a decision toward malignant. Figure 10 shows an example of diagnosis using six models for the same patient.

Using a bipolar perspective we define two IVFSs: pro $\widetilde{P}$ and contra $\widetilde{C}$ diagnosis. The set $\widetilde{P} = (A_l, A_u)$ models positive (malignant) diagnosis, and

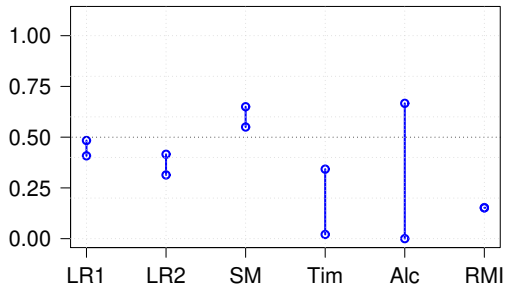

Figure 10: Example of source diagnosis of six diagnostic modules as an IVFS

the set $C = (A_u^{-1}, A_l^{-1})$ models negative (benign) diagnosis. Next we count the cardinality of both sets using Sigma f-Count, which results in two intervals describing the number of models tending towards a malignancy diagnosis $sc_f(\widetilde{P}) = [\sigma_f(A_l), \sigma_f(A_u)]$ and of modules tending towards a benign diagnosis $sc_f(\widetilde{P}) = [\sigma_f((A_u)^{-1}), \sigma_f((A_l)^{-1})]$. It is worth noting that if we use the weighting function $f_{1,t}$ (*counting by thresholding*) this approach will be equivalent to a voting strategy (each model will vote 0 or 1). Having the cardinality intervals $sc_f(\widetilde{P})$ and $sc_f(\widetilde{C})$, we can make a final decision by comparing them in the following way:

Algorithm 1 Decision Algorithm

if $overlap(sc_f(\widetilde{P}), sc_f(\widetilde{C}))$ **or** $distance(sc_f(\widetilde{P}), sc_f(\widetilde{C})) < r$ **then**
  $Decision \leftarrow NA$
else
  if $center(sc_f(\widetilde{P})) > center(sc_f(\widetilde{C}))$ **then**
    $Decision \leftarrow Malignant$
  else
    $Decision \leftarrow Benign$
  end if
end if

If the cardinality intervals are overlapping (function *overlap*) or the distance between the ends of the intervals (function *distance*) is smaller than a given number r, the system cannot make a decision. In other words, if the numbers of models voting pro and contra are similar then it is not possible to make a decision. The parameter r reflects our intuition of how large should be the difference between the numbers of pro and contra votes.

If the intervals are not overlapping, then compare the centers of the cardinal-

		predicted		
		benign	malignant	NA
actual	benign	0	2.5	1
	malignant	5	0	2

Table 3: Cost matrix. The costs were assigned based on expert gynaecologists' opinions.

ity intervals (function *center*) and as a decision select the option with the greater value. This means that we choose the decision with more models supporting it.

The decision algorithm is illustrated in the following example.

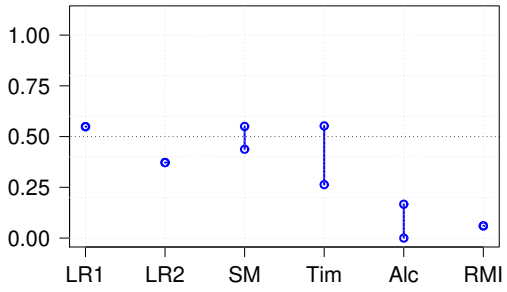

Figure 11: Weighting function $f_{2,t}$ for $t = 0.5$

For the input decisions shown in Figure 11 we execute the algorithm with parameter $r = 1$ (a decision should be taken by a majority of at least one). When the weighting function $f_{2,0.5}$ was used the following sigma f-counts were computed: $sc_f(\widetilde{P}) = [1.7, 2.3]$, $sc_f(\widetilde{C}) = [3.7, 4.3]$. Hence the final decision is malignant. For comparison, if the function $f_{1,0.5}$ were used with the same parameters, the cardinality would be $sc_f(\widetilde{P}) = [1, 3]$, $sc_f(\widetilde{C}) = [3, 5]$ and hence no decision could be taken.

4. Results and Discussion

The study group consisted of 375 patients diagnosed and treated for ovarian tumor in the Division of Gynaecological Surgery at Poznan University of Medical Sciences. Among them, 232 were diagnosed with a benign tumor and 143 with a malignant tumor. The training set consisted of 200 patients, and the test set of 175 patients. All patients in the training set had a complete set of features. All patients in the test set had missing values up to 50%. The original diagnostic models operate directly on the patients' features. The new diagnostic models were evaluated on intervals created by the original diagnostic models as it was described in section 3.1. New diagnostic modules were optimised in the training phase to minimise the total misclassification cost. The cost matrix was based on expert knowledge and is presented in Table 3.

We set the learning and evaluation procedures so that, firstly, in the training phase we simulate possible data incompleteness derived from the patients with

the complete set of features and, secondly, we estimate misclassification error on the patients with real incomplete features. Notably, we extend the training set by a 1000-fold random balanced sampling of 150 patients and random features obscuration to get missing values in range $[0\%, 50\%]$.

Performance measures are presented in Table 4. The original diagnostic models vary in their classification characteristics: some of them tend to be more conservative (LR1, LR2, SM) and some more liberal (RMI, Tim.). This can be observed in the considerable differences between values of sensitivity and specificity. Only one model maintains a balance between those two factors (Alc.). Unfortunately, all six models have low decisiveness (due to the lack of some patient attributes), which results in high total costs for these models.

The new diagnostic modules have high values of both sensitivity and specificity. Two of them tend to be more conservative (OEA and IVFC), and one maintains a balance between sensitivity and specificity (FSC). These algorithms can handle incompleteness of the input data, so levels of decisiveness are very high. As a result, the total costs for the modules are much lower – approximately two times lower than in the original models.

		Total cost	Dec.	Sen.	Spec.	Acc.
Original models	Alc. [2]	189.0	20.6 %	88.2 %	89.5 %	88.9 %
	LR1 [6]	184.0	27.4 %	92.6 %	57.1 %	77.1 %
	LR2 [6]	164.0	33.1 %	94.3 %	65.2 %	82.8 %
	RMI [5]	156.0	56.6 %	75.9 %	87.1 %	83.8 %
	SM [3]	142.0	62.9 %	94.6 %	71.2 %	79.1 %
	Tim. [4]	159.0	47.4 %	66.7 %	97.1 %	91.6 %
New diag. modules	OEA	72.0	96.6 %	90.2 %	86.4 %	87.6 %
	IVFC	72.5	100.0 %	90.4 %	84.6 %	86.3 %
	FSC	78.5	93.7 %	87.2 %	88.9 %	88.4 %

Table 4: Performance measures of the original diagnostic models and new diagnostic modules. Results are obtained on the test set. Abbreviations: Dec. – Decisiveness, Sen. – Sensitivity, Spec. – Specificity, Acc. – Accuracy, OAE – module based on diagnostic models aggregation (section 3.1), IVFC – module based on interval-valued fuzzy classifier (section 3.2), FSC – module based on interval-valued fuzzy set cardinality (section 3.3).

A statistical comparison between the original models and new diagnostic modules is presented in Table 5. The new diagnostic modules classify significantly differently than the original models. Although the modules differ in their performance measures, the classification difference among them did not reach statistical significance.

In the light of the foregoing results, *OvaExpert* provides a promising classification tool. The final diagnosis can be achieved based on binary classification, as well as with the use of a multi-class approach. All of the new methods can handle data incompleteness without imputing missing values, which leads to a high level of decisiveness. The aforementioned division of the dataset is the first step to obtain reliable and stable evaluation results of our approach. In the next phase of development of the system we will evaluate our models on

		New diagnostic modules		
		OEA	IVFC	FSC
Original models	Alc. [2]	< 0.001	< 0.001	< 0.001
	LR1 [6]	< 0.001	< 0.001	< 0.001
	LR2 [6]	< 0.001	< 0.001	< 0.001
	RMI [5]	< 0.001	< 0.001	< 0.001
	SM [3]	< 0.001	< 0.001	< 0.001
	Tim. [4]	< 0.001	< 0.001	< 0.001
New diag. modules	OEA	-	0.579	0.579
	IVFC	0.579	-	0.093
	FSC	0.579	0.093	-

Table 5: McNemar's test with Benjamini-Hochberg correction between the original diagnostic models and new diagnostic modules ($\alpha = 0.05$). Results are obtained on the test set. Abbreviations: OAE – module based on diagnostic models aggregation (section 3.1), IVFC – module based on interval-valued fuzzy classifier (section 3.2), FSC – module based on interval-valued fuzzy set cardinality (section 3.3).

external datasets delivered by cooperating medical centres in Europe. This will also help us in re-assessment of statistical comparison of the classifiers, since our dataset may not be large enough to deliver a robust evidence on classification difference.

5. Conclusions and Further Work

We have presented the development of a system to support decision-making in the diagnosis of ovarian tumors. We have also presented the theoretical basis and the most important diagnostic algorithms used in the system. As reflected in the results, if the input data is incomplete, the system copes much better compared with existing diagnostic models. Hence, it can support the physician more effectively in actual diagnosis. Moreover, we believe that our approach can be effectively adapted to non-medical problems where data quality is a matter of concern [28].

Naturally, we are still faced with the challenge of improving the system. One of the challenges is to collect more data in order to optimise the algorithms more effectively. This can be done by deploying the system at many specialised medical centers. Moreover, it seems to be necessary to carry out studies on the effectiveness of the proposed algorithms on external data. We are currently working on the preparation of such analyses.

References

References

[1] A. du Bois, J. Rochon, J. Pfisterer, W. J. Hoskins, Variations in institutional infrastructure, physician specialization and experience, and out-

come in ovarian cancer: a systematic review, Gynecologic Oncology 112 (2) (2009) 422–436.

[2] J. L. Alcázar, L. T. Mercé, C. Laparte, M. Jurado, G. López-García, A new scoring system to differentiate benign from malignant adnexal masses, Obstetrical & Gynecological Survey 58 (7) (2003) 462–463.

[3] D. Szpurek, R. Moszyński, W. Zietkowiak, M. Spaczyński, S. Sajdak, An ultrasonographic morphological index for prediction of ovarian tumor malignancy, European Journal of Gynaecological Oncology 26 (1) (2005) 51–54.

[4] D. Timmerman, A. C. Testa, T. Bourne, E. Ferrazzi, L. Ameye, M. L. Konstantinovic, B. Van Calster, W. P. Collins, I. Vergote, S. Van Huffel, L. Valentin, 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) (2005) 8794–8801.

[5] I. Jacobs, D. Oram, J. Fairbanks, J. Turner, C. Frost, J. G. Grudzinskas, 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) (1990) 922–929.

[6] B. W. J. Mol, D. Boll, M. De Kanter, A. P. M. Heintz, E. A. Sijmons, S. G. Oei, H. Bal, H. A. M. Brölmann, Distinguishing the Benign and Malignant Adnexal Mass: An External Validation of Prognostic Models, Gynecologic Oncology 80 (2) (2001) 162–167.

[7] B. Van Calster, K. Van Hoorde, L. Valentin, A. C. Testa, D. Fischerova, C. Van Holsbeke, L. Savelli, D. Franchi, E. Epstein, J. Kaijser, V. Van Belle, A. Czekierdowski, S. Guerriero, R. Fruscio, C. Lanzani, F. Scala, T. Bourne, D. Timmerman, Evaluating the risk of ovarian cancer before surgery using the ADNEX model to differentiate between benign, borderline, early and advanced stage invasive, and secondary metastatic tumours: prospective multicentre diagnostic study, BMJ 349 (2014) 5920.

[8] R. Moszyński, P. Żywica, A. Wójtowicz, S. Szubert, S. Sajdak, A. Stachowiak, K. Dyczkowski, M. Wygralak, D. Szpurek, 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) (2014) 892–899.

[9] C. Van Holsbeke, B. Van Calster, L. Valentin, A. C. Testa, E. Ferrazzi, I. Dimou, C. Lu, P. Moerman, S. Van Huffel, I. Vergote, D. Timmerman, 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) (2007) 4440–4447.

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

[11] A. Stachowiak, K. Dyczkowski, A. Wójtowicz, P. Żywica, M. Wygralak, A bipolar view on medical diagnosis in OvaExpert system, in: Flexible Query Answering Systems 2015, 2016, pp. 483–492.

[12] P. Han, W. Klein, N. K. Arora, Varieties of uncertainty in health care: a conceptual taxonomy, Medical Decision Making 31 (6) (2011) 828–838.

[13] A. Wójtowicz, P. Żywica, K. Szarzyński, R. Moszyński, S. Szubert, K. Dyczkowski, A. Stachowiak, D. Szpurek, M. Wygralak, Dealing with Uncertainty in Ovarian Tumor Diagnosis, in: K. T. Atanassov, et al. (Eds.), New Developments in Fuzzy Sets, Intuitionistic Fuzzy Sets, Generalized Nets and Related Topics, IBS PAN – SRI PAS, Warsaw, 2014.

[14] D. Timmerman, L. Valentin, T. H. Bourne, W. P. Collins, H. Verrelst, I. Vergote, 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 & Gynecology 16 (5) (2000) 500–505.

[15] D. Timmerman, T. H. Bourne, A. Tailor, W. P. Collins, H. Verrelst, K. Vandenberghe, I. Vergote, 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) (1999) 57–65.

[16] K. T. Atanassov, Intuitionistic fuzzy sets, Springer, 1999.

[17] M. Wygralak, Intelligent Counting Under Information Imprecision, Vol. 292 of Studies in Fuzziness and Soft Computing, Springer, 2013.

[18] S. K. De, R. Biswas, A. R. Roy, An application of intuitionistic fuzzy sets in medical diagnosis, Fuzzy Sets and Systems 117 (2) (2001) 209–213.

[19] E. Szmidt, J. Kacprzyk, An intuitionistic fuzzy set based approach to intelligent data analysis: an application to medical diagnosis, in: Recent advances in intelligent paradigms and applications, Springer, 2003, pp. 57–70.

[20] Anna Pankowska and Maciej Wygralak, On Hesitation Degrees in IF-Set Theory, in: Leszek Rutkowski and Jorg Siekmann and Ryszard Tadeusiewicz and Lotfi Zadeh (Ed.), Artificial Intelligence and Soft Computing. LNAI, Vol. 3070, Springer, Heidelberg, 2004, pp. 338–343.

[21] M. Stukan, M. Dudziak, K. Ratajczak, J. P. Grabowski, Usefulness of diagnostic indices comprising clinical, sonographic, and biomarker data for discriminating benign from malignant ovarian masses., Journal of Ultrasound in Medicine 34 (2) (2015) 207–217.

[22] R. R. Yager, On ordered weighted averaging aggregation operators in multicriteria decisionmaking, Systems, Man and Cybernetics, IEEE Transactions on 18 (1) (1988) 183–190.

[23] P. Żywica, A. Wójtowicz, A. Stachowiak, K. Dyczkowski, Improving medical decisions under incomplete data using interval–valued fuzzy aggregation, in: Proceedings of 9th European Society for Fuzzy Logic and Technology (EUSFLAT), Gijón, Spain, 2015, pp. 577–584.

[24] A. Stachowiak, P. Żywica, K. Dyczkowski, A. Wójtowicz, An Interval-Valued Fuzzy Classifier Based on an Uncertainty-Aware Similarity Measure, in: Intelligent Systems' 2014, Springer, 2015, pp. 741–751.

[25] L. A. Zadeh, The concept of a linguistic variable and its application to approximate reasoning—i, Information Sciences 8 (3) (1975) 199–249.

[26] M. Diering, K. Dyczkowski, A. Hamrol, New Method for Assessment of Raters Agreement Based on Fuzzy Similarity, in: A. Herrero, J. Sedano, B. Baruque, H. Quintian, E. Corchado (Eds.), 10th International Conference on Soft Computing Models in Industrial and Environmental Applications, Vol. 368 of Advances in Intelligent Systems and Computing, Springer International Publishing, 2015, pp. 415–425.

[27] L. A. Zadeh, A computational approach to fuzzy quantifiers in natural languages, Computers & Mathematics with Applications 9 (1) (1983) 149 – 184.

[28] A. Kujawińska, M. Diering, A. Hamrol, Measurement System Analysis Combined with Shewhart's Approach, Key Engineering Materials 637 (2015) 7–11.