Membership functions for fuzzy focal elements

SEBASTIAN PORĘBSKI and EWA STRASZECKA

The paper presents a study on data-driven diagnostic rules, which are easy to interpret by human experts. To this end, the Dempster-Shafer theory extended for fuzzy focal elements is used. Premises of the rules (fuzzy focal elements) are provided by membership functions which shapes are changing according to input symptoms. The main aim of the present study is to evaluate common membership function shapes and to introduce a rule elimination algorithm. Proposed methods are first illustrated with the popular Iris data set. Next experiments with five medical benchmark databases are performed. Results of the experiments show that various membership function shapes provide different inference efficiency but the extracted rule sets are close to each other. Thus indications for determining rules with possible heuristic interpretation can be formulated.

Key words: diagnostic rule extraction, medical diagnosis support, fuzzy focal elements, membership functions, Dempster-Shafer theory.

1. Introduction

The Dempster-Shafer theory (DST) and the fuzzy set theory (FST) are often used together to represent both uncertainty and imprecision of a problem. The key concepts of the DST are focal elements and their basic probability assignment [7], [22], while FST employs a membership function of a fuzzy set. Several authors suggest fuzzyfing the basic probability assignment so the membership value is related to a focal element [35] or linking random set variables to fuzzy sets [9], as well as considering a fuzzy measure as a unifying structure that represents an uncertain variable [34]. These approaches allow for transitions between uncertainty and imprecision. However, sometimes the two kinds of measures have to represent different types of vagueness. In medical diagnosis a fuzzy set may stand for the strength of a symptom (e.g. high blood pressure) while the value of the basic probability determines a disease, given the symptom (a risk of an infarct given a hypertension). Therefore, the present paper suggests a different approach, useful in solving diagnosis support problems.

The basic probability assignment can be considered as a measure of uncertainty of focal elements that are diagnostic symptoms. Yet, the symptoms are usually formulated

with some imprecision. The imprecision measure can be the membership function of the fuzzy set. This approach is introduced in [29, 30] and shares Bezdek's conviction [2] that probability is something different from a membership and each of the measures should be individually interpreted. The method is convenient for applications, particularly when the diagnosis support requires combining data driven and human expert knowledge. Yet, it involves designing membership functions and determining their crucial parameters that are decisive for reasoning [3], [16].

Methods of designing membership functions differ according to a purpose of their use. They can be found from training examples [20], in an identification process [31], by means of neural networks [23] or evolutionary algorithms [6]. Still, membership functions of medical symptoms have to meet specific constraints of this domain. For instance, they should be related to population data [3]. Hence, it would be valuable for the knowledge engineer to know how to find functions which are preferable for suggested method of the diagnosis support.

In the present work, various shapes of membership functions are investigated. Thorough designing of the functions allow for correct reasoning with a limited number of rules. Thus, a rule elimination algorithm is introduced. Experiments are performed for several medical benchmark databases. Results show that membership functions influence a quality of inference. The most advantageous shapes of the functions are indicated. It is demonstrated that both a reduction of a rule set and a more accurate inference is possible with the suggested approach, in comparison to existing methods. As the result limited number of rules with premises that are easy to interpret by humans are determined.

2. The Dempster-Shafer theory in the diagnosis support

Let us assume a diagnostic problem with p symptoms: x_1 ..., x_p. The membership function $\mu_j^{(r_l)}$ of a fuzzy set $A_j^{(r_l)}$ is assigned to each of the p variables ($j = 1, ..., p$) that characterizes a linguistic value of the symptom in the r-th rule corresponding the D_l diagnosis. Thus, the r_l-th diagnostic rule is of the form:

$$\text{IF } X_j \text{ is } A_j^{(r_l)} \wedge ... \wedge X_k \text{ is } A_k^{(r_l)} \text{ THEN } D_l, \quad j, k \in \{1, ..., p\}, \quad j \neq k. \tag{1}$$

The premise of the rule (1) includes a subset of all possible combinations of the symptom set. The complete set of premises for D_l is denoted by S_l. Then the maximum number of rules is $|S_l| = 2^p - 1$. Each premise can be considered as the $s^{(r_l)}$ focal element. Hence, $S_l = \{s^{(r_l)}\}$, $r_l = 1, ..., n^{(r_l)}$, and the greatest possible value of $n^{(r_l)}$ is $|S_l|$. For such focal elements the basic probability assignment (BPA) can be defined as [7]:

$$m_l(f) = 0, \quad \sum_{s^{(r_l)} \in S_l} m_l(s^{(r_l)}) = 1. \tag{2}$$

In this way uncertainty of the r-th rule is given by $m_l(s^{(r_l)})$. The imprecision of the premise is separately represented by the membership functions of the fuzzy sets $A_j^{(r_l)}$ in

rule (1). The $\mu_j^{(r_l)}$ membership function describes one condition in (1). The imprecision of the whole premise can be determined according to possibility theory as the matching level [30]:

$$\eta_i^{(r_l)} = \min_{1 \leqslant j \leqslant K^{(r_l)}} \mu_j^{(r_l)}(x_j), \tag{3}$$

where $K^{(r_l)}$ is the number of conditions in the r_l-th rule.

The membership functions as well as the BPA can be determined for training data. Since determination of the former is one of the main topics of this paper, let us leave it to next sections and let as describe the latter first.

The theoretical background of the BPA calculation is given in [29]. Here only a brief summary of the algorithm is mentioned with an introduction of notations. Let us assume the data set $X_l = \{x_1, ..., x_i, ..., x_{n_l}\}$ of p-dimensional vectors, which are cases assigned to the l-th diagnosis. The BPA can be calculated in the following way:

1. Choose an acceptable threshold of uncertainty η_{BPA} for BPA calculation.

2. Count how many objects x_i ($i = 1, ..., n_l$) is conformed with the r_l-th fuzzy focal element $s^{(r_l)}$ at least to the determined η_{BPA} threshold, i.e.:

$$\eta_i^{(r_l)} > \eta_{BPA}, \quad i = 1, ..., n_l \tag{4}$$

for each x_i data case the matching level to r_l-th rule calculated as:

$$\eta_i^{(r_l)} = \min_{s_k \in s^{(r_l)}} \mu_k^{(l)}(x_{ki}), \tag{5}$$

where $\mu_k^{(r_l)}$ is the membership function in $s^{(r_l)}$ fuzzy focal element which correspond to the k-th variable and l-th diagnosis and x_{ki} is the value of the k-th variable given by the k-th element of x_i vector.

3. The obtained number of satisfactory matching objects $\gamma_i^{(r_l)}$ divided by the total sum of data objects conformed with all fuzzy focal elements makes the BPA value:

$$m_l(s^{(r_l)}) = \frac{\sum\limits_{i=1}^{n_l} \gamma_i^{(r_l)}}{\sum\limits_{r_l=1}^{n^{(l)}} \left(\sum\limits_{i=1}^{n_l} \gamma_i^{(r_l)} \right)} \tag{6}$$

where

$$\gamma_i^{(r_l)} = \begin{cases} 1, & \eta_i^{(r_l)} > \eta_{BPA} \\ 0, & \eta_i^{(r_l)} \leqslant \eta_{BPA}. \end{cases} \tag{7}$$

The η_{BPA} threshold value can be changed ($\eta_{BPA} \in (0, 1]$). The higher is the η_{BPA} the smaller is number of data cases conformed with focal elements. If so, the BPA value allows for different rule evaluation.

The belief (*Bel*) and plausibility (*Pl*) measures described in the Dempster-Shafer theory [22] are also useful for evaluation of a diagnosis [30]. Particularly the *Bel* value can be used to choose the best supported diagnosis among $l = 1, ..., C$ diagnostic hypotheses for the investigated $\boldsymbol{x}$ data case. This measure is calculated as:

$$Bel_l(\boldsymbol{x}) = \sum_{\substack{\eta^{(r_l)} > \eta_T \\ 1 \leqslant r_l \leqslant n^{(l)}}} m_l(s^{(r_l)}), \tag{8}$$

The belief measure (8) provides that patient symptoms matching the fuzzy focal elements are counted for the final diagnosis. The conformation of symptoms with the focal element is evaluated by the η_T threshold ($\eta_T \in (0, 1]$) individually assumed for the diagnosis.

When $Bel_1(\boldsymbol{x})$, ..., $Bel_C(\boldsymbol{x})$ values for all diagnoses ($l = 1, ..., C$) are known then the final conclusion is stated by their comparison. The diagnostic hypothesis with the greatest belief value wins. If the greatest value occurs for more than one diagnosis, the decision cannot be made and it is interpreted as the error of inference.

The number of focal elements (rules) can be lowered when for some r_l $\sum_{i=1}^{n_l} \gamma_i^{(r_l)} = 0$. This may happen if membership functions and thresholds ensures the most effective of inference and in consequence redundant conditions are neglected.

3. The design of membership functions

Fuzzy focal elements are defined by membership functions. The membership functions should provide proper interpretation of symptoms. Statistical investigation helps to determine characteristics of processing data. Let us define $\boldsymbol{x}_{jl} = [x_{j1}, ..., x_{ji}, ..., x_{jn_l}]^T$ as the vector of the *j*-th feature values corresponding to the *l*-th class. The following statistical parameters are important for a fuzzy membership function shape:

- mean value of the *j*-th feature of the *l*-th class:

$$\bar{x}_{jl} = \frac{1}{n_l} \sum_{i=1}^{n_l} x_{ji}, \tag{9}$$

- standard deviation:

$$\sigma_{jl} = \sqrt{\frac{1}{n_l - 1} \sum_{i=1}^{n_l} x_{ji} - \bar{x}_{jl}}, \tag{10}$$

- lower quartile Q_{1_j} as maximal x_{ji} value for which:

$$\left| X_j^* \right| \leqslant \frac{1}{4} n_l. \tag{11}$$

where $X_j^* = \left\{ x_{ji} \leqslant Q_{1_j} \right\}$.

- upper quartile Q_{3_j} as minimal x_{ji} value for which:

$$\left| X_j^{**} \right| \geqslant \frac{1}{4} n_l. \tag{12}$$

where $X_j^{**} = \left\{ x_{ji} \geqslant Q_{3_j} \right\}$.

Moreover, distributions of data for different diagnoses should be investigated. The two neighbouring normal distributions defined by the probability density functions $P_1(x)$ and $P_2(x)$ always have at least one point of intersection x_{cross}. It could be calculated when the mean values ($\bar{x}_1$ and $\bar{x}_2$) and the standard deviations (σ_1 and σ_2) are known. The x_{cross} value is the solution of the equation:

$$P_1(x) = P_2(x), \tag{13}$$

where

$$P_1(x) = \frac{1}{\sigma_1 \sqrt{2\Pi}} \cdot \exp \left(-\frac{(x - \bar{x}_1)^2}{2\sigma_1^2} \right)$$

and

$$P_2(x) = \frac{1}{\sigma_2 \sqrt{2\Pi}} \cdot \exp \left(-\frac{(x - \bar{x}_2)^2}{2\sigma_2^2} \right).$$

If $\bar{x}_1 \neq \bar{x}_2$ two values x_{01} and x_{02} are the solution of the equation (13). The x_{cross} point is the one of the roots which fulfil the condition:

$$x_{cross} = \{x\} \, | x \in [\bar{x}_1, \bar{x}_2]. \tag{14}$$

However, it may happen that (14) does not occur, for instance if the distributions have similar means and different variances. In this case, equivalent x_{cross} is calculated as $\frac{\bar{x}_1 + \bar{x}_2}{2}$ and used to membership function design. Fig. 1 shows the characteristic values calculated for the data with normal distribution.

Among a number of different membership function shapes triangular, trapezoidal, Gaussian combination and generalized bell are considered in this paper. Since the shapes should be illustrated by examples, the benchmark Iris data set [12] is chosen to depict the functions. This set obviously is not a diagnostic base, but it is a popular benchmark, well-known and easy to handle, so it makes an illustrative example of the presented method. The base include 150 objects, which are a numerical data about three species of the Iris flowers named *Iris Setosa*, *Iris Versicolour* and *Iris Virginica*. Every data case has five parameters: four of them provide metrical dimensions of sepal and petal. Hence,
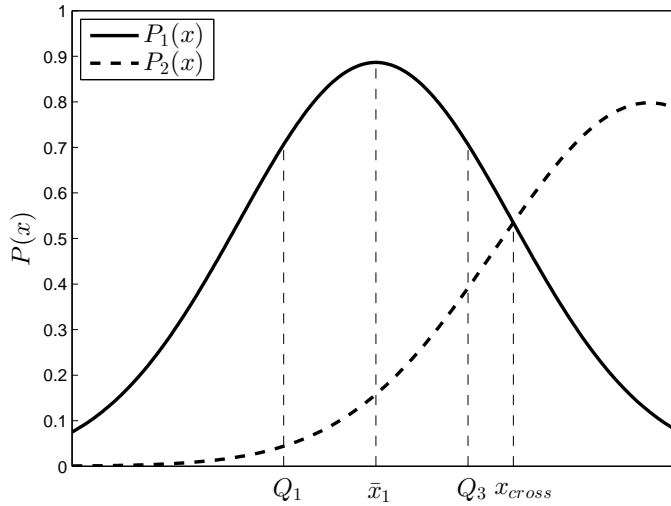
Figure 1: Mean $\bar{x}_1$, lower quartile Q_1, upper quartile Q_3 for normal distribution density function $P_1(x)$ and intersection x_{cross} of the two neighbouring distributions $P_1(x)$ and $P_2(x)$

the number of features in the set equals four ($p = 4$). The fifth parameter is the class label. The values contained in the Iris data set satisfy the central limit theorem and could be considered as normally distributed. Hence, it is possible to use mentioned data set characteristic values to construct the fuzzy membership functions.

3.1. Triangular and trapezoidal membership functions

A triangular membership function $\mu_T(x,[a,b,c])$ is defined as follows:

$$\mu_T(x,[a,b,c]) = \begin{cases} 0, & x \leqslant a, \\ \frac{x-a}{b-a}, & a < x \leqslant b, \\ \frac{c-x}{c-b}, & b < x \leqslant c, \\ 0, & x > c, \end{cases} \tag{15}$$

where a, b and c are determined as in Fig. 2.

The Iris data set includes cases which are classified into three classes ($C = 3$). Hence it is desired to design three membership functions for every variable. The intersection points are calculated between the neighbouring classes when the $\bar{x}_{jl}$ and σ_{jl} pairs ($l = 1, ..., C$) are sorted in ascending order of the $\bar{x}_{jl}$. The intersection point $x_{cross}^{(l)}$ is calculated between l-th and $(l+1)$-th class ($1 \leqslant l \leqslant C-1$). If we keep the later definition of $\boldsymbol{x}_{jl} = [x_{j1},...,x_{ji},...,x_{jn_l}]^T$ as the vector of the j-th feature values corresponding to the l-th class, the a, b and c values are equal:
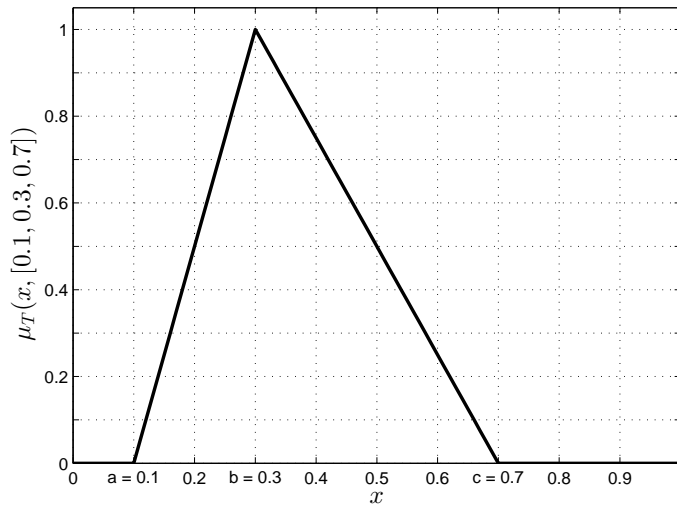
Figure 2: The example of a triangular membership function $\mu_T(x,[a,b,c])$ for $a = 0.1$, $b = 0.3$ and $c = 0.7$

- $a = \begin{cases} 0.01 \cdot \min\limits_{1 \leqslant i \leqslant n_l}(x_{ji}) & \text{for } l = 1, \\ \bar{x}_{jl} - 2(\bar{x}_{jl} - x_{cross}^{(l-1)}) & \text{for } 1 < l \leqslant C, \end{cases}$

- $b = \bar{x}_{jl}$,

- $c = \begin{cases} \bar{x}_{jl} + 2(x_{cross}^{(l)} - \bar{x}_{jl}) & \text{for } 1 \leqslant l < C, \\ 100 \cdot \max\limits_{1 \leqslant i \leqslant n_l}(x_{ji}) & \text{for } l = C. \end{cases}$

The values a for the first ($l=1$) class and c for the last ($l=C$) class provide a full membership for the $\min\limits_{1 \leqslant i \leqslant n_l}(x_{ji})$ and $\max\limits_{1 \leqslant i \leqslant n_l}(x_{ji})$ values, respectively. The 0.01 and 100 coefficients are introduced to extend the support of the membership function in case test data have smaller (or greater) values than training data. In the other cases the a and b values provide a half membership function for data located in the intersection point x_{cross} between two neighbouring classes. Designed triangular membership functions for the Iris data set with the intersection points are shown in Fig. 3.

A trapezoidal membership function can be interpreted as a generalization of a triangular function. The function has four characteristic parameters and is defined in the
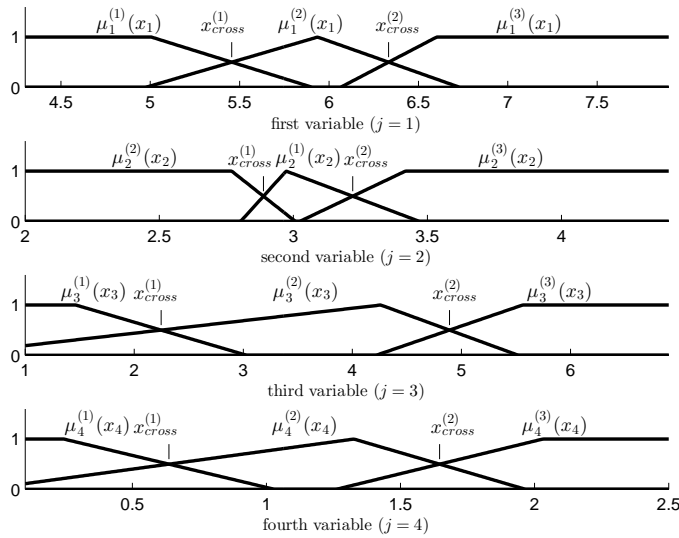
Figure 3: The triangular membership functions for $p = 4$ features and $C = 3$ classes of the Iris data set representing the fuzzy sets $A_j^{(r_l)}$

following way:

$$\mu_{TR}(x, [a,b,c,d]) = \begin{cases} 0, & x \leqslant a, \\ \frac{x-a}{b-a}, & a < x \leqslant b, \\ 1, & b < x \leqslant c, \\ \frac{d-x}{d-c}, & c < x \leqslant d, \\ 0, & x > d, \end{cases} \tag{16}$$

where the a, b, c and d are located as it is shown in Fig. 4.

Trapezoidal membership functions are created similarly to the triangular yet with more constraints. For instance:

- the class with the lowest value of the $\bar{x}_{jl}$ ($l = 1$):

  { $a = 0.01 \cdot \min_{1 \leqslant i \leqslant n_l} (x_{ji})$,

  { $b = \min_{1 \leqslant i \leqslant n_l} (x_{ji})$,

  { $c = Q_{3jl} + \alpha \left| x_{cross}^{(l)} - Q_{3jl} \right|$,

  { $d = c + 2(x_{cross}^{(l)} - c)$.

- The classes with the intermediate values of the $\bar{x}_{jl}$ ($1 < l < C$):

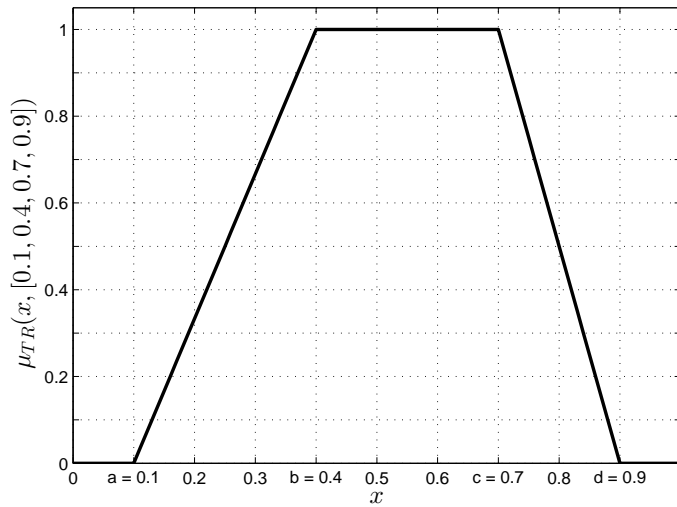 { $b = Q_{1jl} - \alpha \left| Q_{1jl} - x_{cross}^{(l-1)} \right|$,

Figure 4: The example of a trapezoidal membership function $\mu_{TR}(x, [a, b, c, d])$ for $a = 0.1$, $b = 0.4$, $c = 0.7$ and $d = 0.9$

$$\{ \ a = b - 2(b - x_{cross}^{(l-1)}),$$

$$\{ \ c = Q_{3_{jl}} + \alpha \left| x_{cross}^{(l)} - Q_{1_{jl}} \right|,$$

$$\{ \ d = c + 2(x_{cross}^{(l)} - c).$$

- The class with the highest value of the $\bar{x}_{jl}$ ($l = C$):

$$\{ \ b = Q_{1_{jl}} - \alpha \left| Q_{1_{jl}} - x_{cross}^{(l-1)} \right|,$$

$$\{ \ a = b - 2(b - x_{cross}^{(l-1)}),$$

$$\{ \ c = \max_{1 \leqslant i \leqslant n_l} (x_{ji}),$$

$$\{ \ d = 100 \cdot \max_{1 \leqslant i \leqslant n_l} (x_{ji}).$$

where the α parameter ($\alpha \in (-1, 1)$) provides different gradient of the trapezoid slopes while maintaining the condition of a half membership value for intersection points $x_{cross}^{(l)}$. The α value can be tuned according to training data. However, if lower or upper quartile extends beyond suitable $x_{cross}^{(l)}$ point, a correction of points location by making functionŠs slopes almost vertical is necessary [30]:

- if $b \leqslant x_{cross}^{(l-1)}$ ($1 < l \leqslant C$):

$$a = 0.99 \cdot x_{cross}^{(l-1)},$$

$$b = 1.01 \cdot x_{cross}^{(l-1)}.$$

- if $c \geqslant x_{cross}^{(l)} \ (1 \leqslant l < C)$:

$$c = 0.99 \cdot x_{cross}^{(l)},$$
$$d = 1.01 \cdot x_{cross}^{(l)}.$$

The a, b parameters for $l = 1$ and the c, d parameters for $l = C$ provide full membership value for the smallest and the greatest possible values of data j-th feature, respectively. In other cases the b and c parameters are located between the proper couple of the quartiles and intersection points. Trapezoidal membership functions determined for the Iris data set with intersection points are shown in Fig. 5.
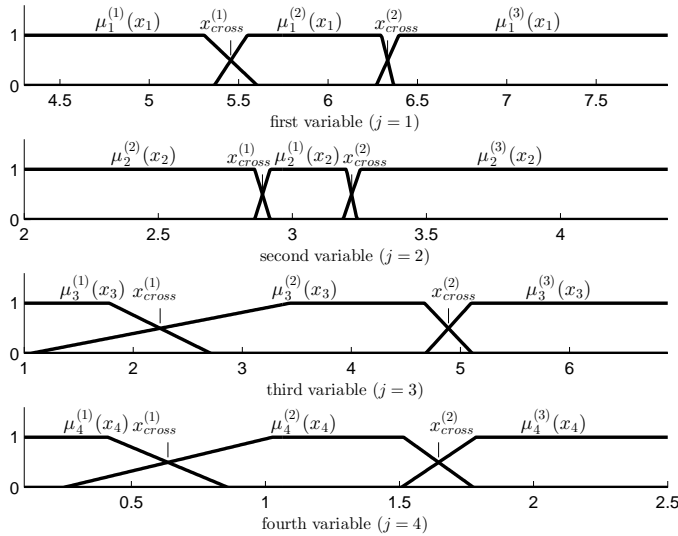

Figure 5: The trapezoidal membership functions for $p = 4$ features and $C = 3$ classes of the Iris data set representing the fuzzy sets $A_j^{(r_l)}$

3.2. Generalized bell membership function

The generalized bell is another membership function shape and it is defined as:

$$\mu_B(x, [a, b, c]) = \frac{1}{1 + \left| \frac{x-c}{a} \right|^{2b}}, \tag{17}$$

where c defines middle value of the membership function (mean value of the data), a affects function's width and b influences function's slope. The example of a generalized bell membership function is shown in Fig. 6.

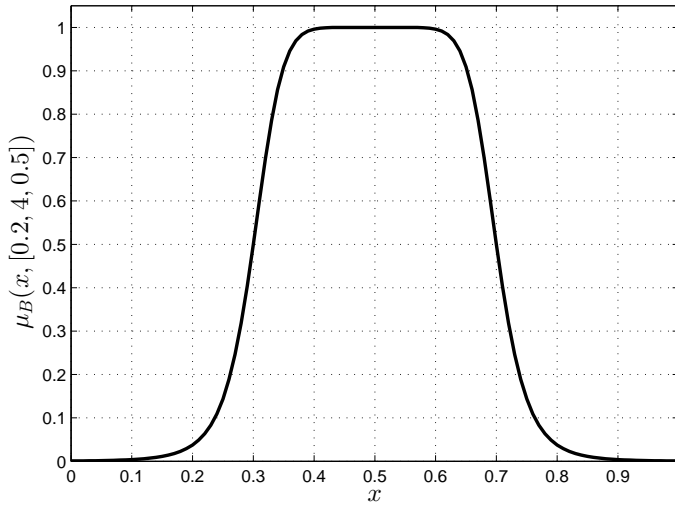The a and c values are calculated in the following way :

Figure 6: The example of a generalized-bell-shaped membership function $\mu_B(x,[a,b,c])$ for $a = 0.2$, $b = 4$ and $c = 0.5$

- $a = \begin{cases} \frac{1}{2} \cdot (\min\limits_{1 \leqslant i \leqslant n_l} (x_{ji}) + x^{(l)}_{cross}) & \text{for } l = 1, \\ \frac{1}{2} \cdot (x^{(l-1)}_{cross} + x^{(l)}_{cross}) & \text{for } 1 < l < C, \\ \frac{1}{2} \cdot (x^{(l-1)}_{cross} + \max\limits_{1 \leqslant i \leqslant n_l} (x_{ji})) & \text{for } l = C, \end{cases}$

- $c = \begin{cases} \frac{1}{2} \cdot (x^{(1)}_{cross} - \min\limits_{1 \leqslant i \leqslant n_l} (x_{ji})) & \text{for } l = 1, \\ \frac{1}{2} \cdot (x^{(l)}_{cross} - x^{(l-1)}_{cross}) & \text{for } 1 < l < C, \\ \frac{1}{2} \cdot (\max\limits_{1 \leqslant i \leqslant n_l} (x_{ji}) - x^{(l-1)}_{cross}) & \text{for } l = C. \end{cases}$

The b parameter is assumed to 4 because it provides proper slopes and significantly wide interval of big values of the function. The generalized bell membership functions designed for the Iris data set with their intersection points are shown in Fig. 7.

3.3. Gaussian combination membership function

This shape is defined as combination of the two Gaussian membership functions defined as:

$$\mu_G(x,[\sigma,c]) = e^{\frac{-(x-c)^2}{2\sigma^2}}, \tag{18}$$

where c and σ are determined mean and standard deviation values.

The Gaussian combination membership function can be used instead of the classical Gaussian, because it provides possibility to design independently two slopes of the
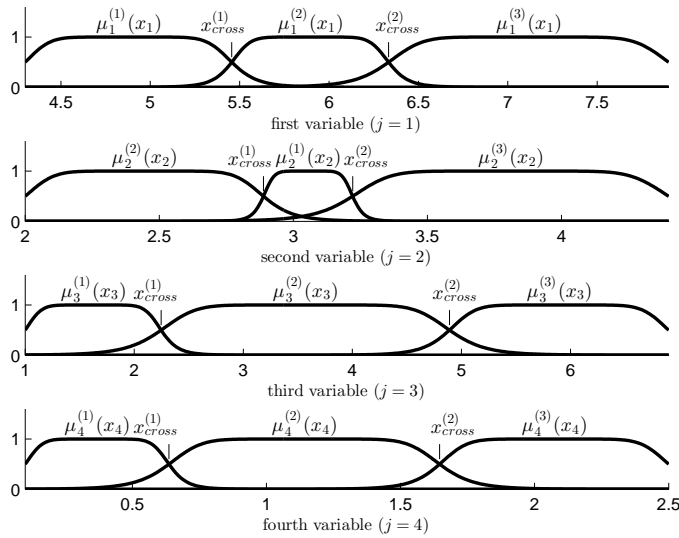
Figure 7: The generalized bell membership functions for $p = 4$ features and $C = 3$ classes of the Iris data set representing the fuzzy sets $A^{(jl)}$

function. The aim of such a design is to maintain maximum membership value for the mean value and half membership value for data located near suitable $x_{cross}^{(l)}$ points. The Gaussian combination membership function can be defined as:

$$\mu_{\Gamma}(x, [\sigma_1, c_1, \sigma_2, c_2]) = \begin{cases} \mu_G(x, [\sigma_1, c_1]), & x < c_1, \\ 1, & c_1 \leqslant x \leqslant c_2, \\ \mu_G(x, [\sigma_2, c_2]), & x > c_2, \end{cases} \tag{19}$$

The example of the Gaussian combination membership function is shown in Fig. 8. In our approach c_1 and c_2 coefficients are determined commonly ($c_1 = c_2$). This choice ensures smooth change of functions as their derivatives are equal in the point of mean value of the l-th class. Moreover, fuzzy sets defined by the membership functions for the first class ($l = 1$) and the last class ($l = C$) are left- and right-open, respectively. Hence, our Gaussian combination membership functions are defined three-fold:

• for the class with the lowest value of the $\bar{x}_{jl}$ ($l = 1$):

$$\mu_{\Gamma}(x, [\sigma_1, c_1, \sigma_2, c_2]) = \begin{cases} 1 & x \leqslant \bar{x}_{jl}, \\ \mu_G(x, [\sigma_2, c_2]), & x > \bar{x}_{jl}, \end{cases} \tag{20}$$
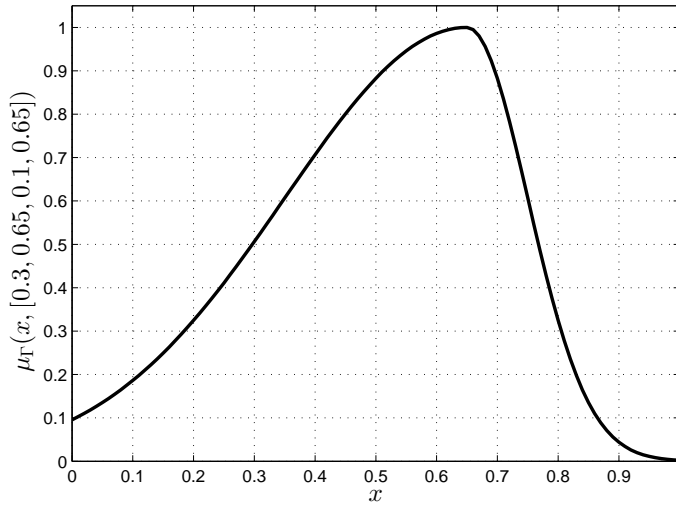
Figure 8: The example of a Gaussian combination membership function $\mu_\Gamma(x, [\sigma_1, c_1, \sigma_2, c_2])$ for $\sigma_1 = 0.3$, $c_1 = 0.65$, $\sigma_2 = 0.1$ and $c_2 = 0.7$

- for the classes with the intermediate values of the $\bar{x}_{jl}$ $(1 < l < C)$:

$$\mu_\Gamma(x, [\sigma_1, c_1, \sigma_2, c_2]) = \begin{cases} \mu_G(x, [\sigma_1, c_1]), & x < \bar{x}_{jl}, \\ 1 & x = \bar{x}_{jl}, \\ \mu_G(x, [\sigma_2, c_2]), & x > \bar{x}_{jl}, \end{cases} \tag{21}$$

- for the class with the highest value of the $\bar{x}_{jl}$ $(l = C)$:

$$\mu_\Gamma(x, [\sigma_1, c_1, \sigma_2, c_2]) = \begin{cases} \mu_G(x, [\sigma_1, c_1]), & x < \bar{x}_{jl}, \\ 1 & x \geqslant \bar{x}_{jl}, \end{cases} \tag{22}$$

where coefficients σ_1, c_1 and σ_2, c_2 are determined as:

- $\sigma_1 = \sqrt{\frac{(x_{cross}^{(l-1)} - \bar{x}_{jl})^2}{2 \ln 0.5}}$,

- $c_1 = \bar{x}_{jl}$,

- $\sigma_2 = \sqrt{\frac{(x_{cross}^{(l)} - \bar{x}_{jl})^2}{2 \ln 0.5}}$,

- $c_2 = \bar{x}_{jl}$.

to provide $\mu_\Gamma(\bar{x}_{jl}) = 1$ and $\mu_\Gamma(x_{cross}^{(l)}) = 0.5$ as it has been mentioned. The Gaussian combination membership functions designed for the Iris data set with their intersection points are shown in Fig. 9.
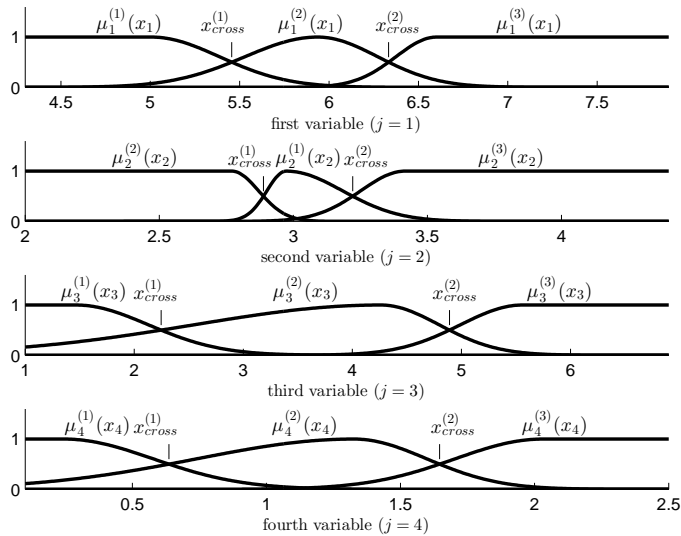
Figure 9: The Gaussian combination membership functions for $p = 4$ features and $C = 3$ classes of the Iris data set representing the fuzzy sets $A^{(jl)}$

4. The rule elimination algorithm

The algorithm described in [28] makes it possible to determine thresholds η_{BPA} and η_T to ensure the best accuracy of a diagnosis. The algorithm with some changes can be also used to proceed iterative elimination of rules from a full rule set. In every step of the algorithm the fuzzy focal element $s^{(r_l)}$ which has the lowest basic probability value (so it fits to the fewest l-th class objects) is deleted. The procedure is finished if the error of an inference grows up or only one rule for some class remains. The improved algorithm of rule elimination used now in experiments could be described as follows:

1. For training data $x_i = \{x_1, ..., x_j, ..., x_p, D_l\}$, $i = 1, ..., N$, $l = 1, ..., C$ design membership functions $\mu_l(x_j)$, $j = 1, ..., p$.

2. Create the focal element sets S_l as maximal sets (then $n^{(l)} = 2^p - 1$) or according to an expert opinion.

3. Calculate m_l according to the algorithm given in section 2

4. Choose the η_T threshold for the inference.

5. Calculate $Bel_l(x_i)$, $l = 1, ..., C$, $i = 1, ..., N$ according to (8) and determine the final diagnosis by the comparison of the belief values.

6. If the final diagnosis cannot be determined or it is different from the training class D_l then note an error of inference.

7. Find $m_{l*} = \min(m_l(s^{(r_l)}))$.

8. Compare $n^{(l)}$, $l = 1, ..., C$. If number of rules for diagnoses are approximately equal then eliminate all rules for which $m_l(s^{(r_l)}) = m_{l*}$. Otherwise eliminate rules of the diagnosis with maximal $n^{(l)}$.

9. Repeat steps 3÷8. Continue until the error is increased in two succeeding eliminations and choose solution with the minimal error.

In the 8. step of the algorithm the elimination is limited when the number of rules for diagnoses significantly differ. In practice the acceptable ratio of rules is assumed as greater than $\frac{2}{3}$. This limitation should secure the algorithm against a convergence to a local minimum solution with few rules of several diagnoses and one diagnosis with multiple rules. The example in the next section illustrates this algorithm in details. It also presents methods of its verification.

5. Example

The mentioned method is illustrated with the Iris data set. Four features ($p = 4$) create $|S_l| = 15$ different fuzzy focal elements for the l-th class. The S_l focal element set includes the following premises for each diagnosis:

$s^{(1)}$: X_1 is $A_1^{(1)}$
$s^{(2)}$: X_2 is $A_2^{(2)}$
$s^{(3)}$: X_3 is $A_3^{(3)}$
$s^{(4)}$: X_4 is $A_4^{(4)}$
$s^{(5)}$: X_1 is $A_1^{(5)} \wedge X_2$ is $A_2^{(5)}$
$s^{(6)}$: X_1 is $A_1^{(6)} \wedge X_3$ is $A_3^{(6)}$
$\vdots$
$s^{(15)}$: X_1 is $A_1^{(15)} \wedge X_2$ is $A_2^{(15)} \wedge X_3$ is $A_3^{(15)} \wedge X_4$ is $A_4^{(15)}$

In the present solution it is assumed that only one membership function is made for the chosen variable and diagnoses, so e.g. $A_1^{(1)}$, $A_1^{(5)}$, ..., $A_1^{(15)}$ are the same linguistic variables defined by the $\mu_1^{(l)}(x_1)$ membership functions. The initial (full) set of focal elements is the same for each diagnosis, but $A_k^{(r_l)}$ and $\mu_k^{(l)}$ which is assigned to it ($l = 1, ..., C$) change according to the diagnosis. The index of diagnosis for $A_k^{(r_l)}$ is omitted for the simplification of notations.

The rule elimination algorithm is executed with the entire data set for the η_{BPA} and η_T thresholds in range 0.05 to 0.95 with step 0.05. In such a way thresholds that ensure the best performance are found. Four described membership function shapes were

checked in the premises of the rules. Let us assume the inference efficiency as:

$$E = (1 - \frac{\varepsilon}{N}) \cdot 100\%, \tag{23}$$

where ε is the number of misclassified cases and N is the number of data cases. The results of the rule elimination algorithm are shown in Tab. 13.

Table 13: The results of the rule elimination algorithm for the Iris data set, different membership functions and η_T, η_{BPA} thresholds

membership function	E [%]	rule set size	η_{BPA}	η_T
triangular (μ_T)	98.00	[1 3 2]	0.5	0.6
trapezoidal (μ_{TR})	98.00	[1 3 2]	0.5	0.3
generalized bell (μ_B)	98.00	[1 3 3]	0.5	0.3
Gaussian combination (μ_Γ)	98.00	[1 3 2]	0.5	0.3

The lowest error obtained for the rule elimination algorithm equals three mistaken classifications. It corresponds to 98% efficiency. It occurs for all considered function shapes. The fuzzy focal element set generated for the triangular membership function shape is presented in Tab. 14.

Table 14: The rule set created for the Iris data with triangular membership functions in fuzzy focal elements

fuzzy focal element $s^{(r_l)}$	$m_1(s^{(r_1)})$	$m_2(s^{(r_2)})$	$m_3(s^{(r_3)})$
$r_l = 3 : X_3$ is $A_3^{(3)}$	0	0.3378	0.5
$r_l = 4 : X_4$ is $A_4^{(4)}$	0	0.3311	0.5
$r_l = 10 : X_3$ is $A_3^{(10)} \wedge X_4$ is $A_4^{(10)}$	1	0.3311	0

The number of extracted rules is small for every class. Only one fuzzy focal element is necessary to classify *Iris Setosa* ($l = 1$) properly:

$$s^{(10)} : \ X_3 \text{ is } A_3^{(10)} \wedge X_4 \text{ is } A_4^{(10)}, \ m_1(s^{(10)}) = 1.$$

Other classes cannot be so easily separated, but still fuzzy focal elements are related to the third and fourth data parameters. Required number of rules for the second and third class is three and two, respectively. It is interesting that the extracted rule set is the same or similar for all considered function shapes (see Fig. 10). In this figure points are labelled by variables which membership functions are used in focal elements (rules). It could be suspected that the algorithm always extract the best rules regardless of the
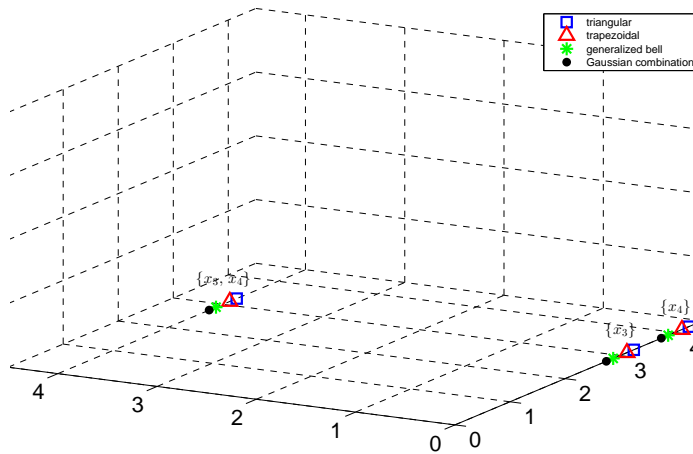
Figure 10: Extracted fuzzy focal elements for the Iris data set ("Iris Versicolor" class) for different membership function shapes. Points represent rules including membership functions of variables which numbers are denoted on axes.

membership function shape. This phenomenon is checked with medical benchmarks in next section.

The cross-validation procedure was introduced to investigate an efficiency of the algorithm for separable training and testing data sets. In the cross-validation done for the Iris data set, 50 training X_U and testing X_T subsets were randomly created with condition that $X_U \cap X_T = \emptyset$. Every training and testing subset contained 75 data cases. Original percentage content of class cases in training and testing subsets is maintained. For every training subset the η_{BPA} and η_T thresholds are searched individually. The η_{BPA} are changed between 0.05 and 0.95 with step 0.05. The η_T is matched to η_{BPA} according to three following conditions suggested in [27]:

$$\eta_T = 0.5 + \left(\frac{\eta_{BPA} - 0.5}{0.5} \right)^3 + 0.5, \tag{24}$$

$$\eta_T = \eta_{BPA}, \tag{25}$$

$$\eta_T = 0.5 + \left(\frac{\eta_{BPA} - 0.5}{0.5} \right)^{\frac{1}{3}} + 0.5. \tag{26}$$

Mean testing efficiency is calculated according to formula:

$$E_T = \frac{1}{K} \sum_{i=1}^{K} E_{ti}, \tag{27}$$

where the E_{ti} is the efficiency (23) obtained for the i-th testing subset and K is the number of the subsets.

Table 15: The results of the cross-validation for the Iris data set for different membership function shapes

membership function	E_T [%]	rule set size (mean)	rule set size (median)
μ_T	97.55	[2 2 2]	[3 3 3]
μ_{TR}	97.33	[3 3 3]	[3 3 3]
μ_B	97.55	[3 3 3]	[3 3 3]
μ_Γ	97.76	[2 2 3]	[3 3 3]

The average efficiencies and sizes of rule sets are shown in Tab. 15. The highest average testing inference efficiency is obtained for the Gaussian combination membership function. It equals 97.76% which means two misclassified cases out of 75 in X_T set. Slightly worse result is obtained for the triangular shapes with one rule less in the S_3 set. Fig. 16a shows average testing efficiency represented by box-and-whiskers diagrams. The best result (97.76% testing efficiency for the Gaussian combination membership function) is better than 90% efficiency mentioned in LS-SVM approach [15], and better than neural network based approach presented and referenced in [36]. Total number (i.e. the number for altogether classes) of Gaussian combination membership function based rules generated for i-th subset of data are shown in Fig. 17a. As we can see the mean number of rules is nine for all data subsets for three classes. The most frequent number of rules is not greater than nine and only six data subsets generated a greater rule set.

6. Experiments

In this section research made for five databases which are well-known medical diagnosis benchmarks is performed. The aim is to find out if an optimal shape of the membership function can be indicated and how efficient can be the algorithm of rule elimination.

6.1. Thyroid Gland

The data set refers to the study of thyroid gland functionality [5]. It contains 215 cases, which are divided to 3 diagnoses:

- normal thyroid functioning – 150 cases,

- hyperthyroidism – 35 cases,

- hypothyroidism – 30 cases.

Every data case is described by six parameters. Five of them provide numerical information about medical test results. The last parameter is the diagnosis label (l = 1, 2,

3). It is possible to create S_l which includes $|S_l| = 31$ fuzzy focal elements for every diagnosis ($l = 1, 2, 3$). Results of the rule elimination algorithm executed for the entire data set are shown in Tab. 16.

Table 16: The results of the rule elimination algorithm for the Thyroid Gland data set, different membership functions and η_T, η_{BPA} thresholds

membership function	E [%]	rule set size	η_{BPA}	η_T
μ_T	98.14	[19 19 19]	0.3	0.4
μ_{TR}	**98.14**	[5 5 4]	0.3	0.15
μ_B	97.67	[8 5 8]	0.5	0.05
μ_Γ	97.21	[6 3 7]	0.25	0.35

The inference efficiency (when the entire data set is considered) equals 98.14% for the triangular and trapezoidal membership functions. Trapezoidal membership functions provide significant reduction of the achieved rule number: five rules for the "normal thyroid activity" and "hyperthyroidism", four for the "hypothyroidism" diagnosis. This result means four wrong diagnoses out of 215 cases. Fuzzy focal elements extracted for trapezoidal membership function are subsets of much greater sets of 19 focal elements when the triangular membership functions are used. The remaining results for the generalized bell and Gaussian combination shapes are also the subsets of the same fuzzy focal element sets. Every detailed elimination provides quite similar focal element sets independently of chosen membership function shape. This observation can be noticed in Fig. 11 where focal element set (i.e. S_1) extracted for four different function shapes is presented.

6.2. Wisconsin Breast Cancer (Original)

This data set contains 699 instances [33], which correspond to two possible diagnoses: the benign or the malignant type of a breast cancer. Every instance of the data set has 10 attributes. Nine of them are nominal values which give information about character of the patient cells examined in the laboratory test. Every of the mentioned attributes gets integer value from 1 to 10. The tenth attribute is the diagnosis label (2 for the benign, 4 for the malignant). Thus, the full rule set size is 511. The results of the rule elimination algorithm are shown in Tab. 17.

The best inference efficiency equals 97.57% and is achieved for the Gaussian combination membership functions. It is related to total 9 rules. The algorithm eliminates 99.4% from the set of 511 rules. The extracted fuzzy focal element sets are also the subsets of bigger sets extracted for methods with worse results. It is shown also in the pseudo-3D diagram in Fig. 12. There are three fuzzy focal elements related to the 'malignant' diagnosis:

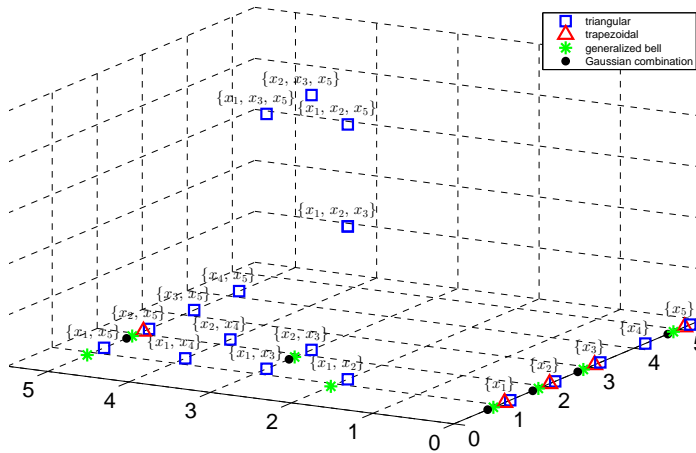$$s^{(2)} : \text{IF } X_2 \text{ is } A_2^{(2)},$$

Figure 11: Extracted fuzzy focal elements for the Thyroid Gland data set ("normal thyroid functioning" diagnosis) for different membership function shapes. Points represent rules including membership functions of variables which numbers are denoted on axes.

Figure 12: Extracted fuzzy focal elements for the Breast Cancer data set ("benign" diagnosis) for different membership function shapes. Points represent rules including membership functions of variables which numbers are denoted on axes.
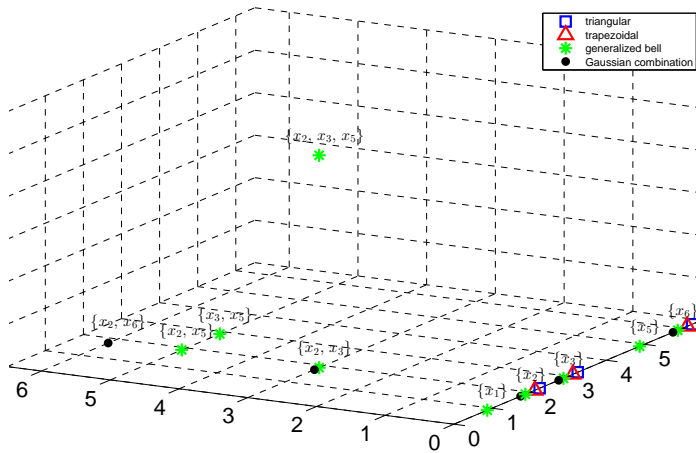
$$s^{(3)} : \text{ IF } X_3 \text{ is } A_3^{(3)},$$
$$s^{(6)} : \text{ IF } X_6 \text{ is } A_6^{(6)},$$

Table 17: The results of the rule elimination algorithm for the Wisconsin Breast Cancer data set, different membership functions and η_T, η_{BPA} thresholds

membership function	E [%]	rule set size	η_{BPA}	η_T
(μ_T)	97.00	[6 6]	0.5	0.3
(μ_{TR})	96.42	[3 3]	0.5	0.05
(μ_B)	97.00	[7 8]	0.15	0.05
(μ_Γ)	**97.57**	[4 5]	0.95	0.25

which occur for every considered membership function shape. Additionally, two complex fuzzy focal elements:

$$s^{(18)} : \text{IF } X_2 \text{ is } A_2^{(18)} \text{ and } X_3 \text{ is } A_3^{(18)},$$

and

$$s^{(21)} : \text{IF } X_2 \text{ is } A_2^{(21)} \text{ and } X_6 \text{ is } A_6^{(21)}$$

are needed to provide the best efficiency when the Gaussian combination shapes are used.

6.3. Cleveland Heart Disease

The Cleveland Heart Disease [8] is another medical data set. Every instance has 76 attributes, but only 14 of them are used in published experiments [4, 10, 11]. The entire data set contains 303 instances. Every instance has 13 numerical attributes of various origin which are considered in heart disease diagnosis. The 14-th attribute is the diagnosis label. Its value equal to 0 stands for the absence of heart disease and 1, 2, 3, 4 for the presence of heart disease with different severity. Hence, classification could be reduced to two diagnoses problem: diagnosis of absence and presence of a heart disease. If 13 attributes are considered, it is possible to create 8191 rules for every diagnosis. The results of the rule elimination algorithm are shown in Tab. 18.

Table 18: The results of the rule elimination algorithm for the Cleveland Heart Disease data set, different membership functions and η_T, η_{BPA} thresholds

membership function	E [%]	rule set size	η_{BPA}	η_T
μ_T	**82.84**	[6 6]	0.2	0.9
μ_{TR}	79.87	[2 3]	0.05	0.7
μ_B	81.19	[6 6]	0.3	0.2
μ_Γ	82.51	[6 5]	0.65	0.35

The best efficiency obtained for the entire set of Heart Disease data is 82.84% for the triangular membership functions. The required number of rules equals six for the

"absence of heart disease" and "presence of heart disease" diagnosis. This is the biggest number of rules when different shapes are considered. Extracted focal element sets are similar for considered shapes (see Fig. 13). Some rules are built as superposition of attributes occurred in elements extracted for other shapes.
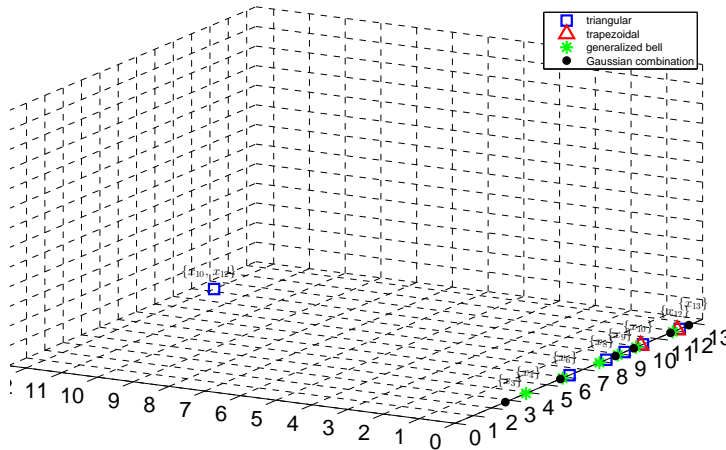

Figure 13: Extracted fuzzy focal elements for the Heart Disease data set ("heart disease absent" diagnosis) for different membership function shapes. Points represent rules including membership functions of variables which numbers are denoted on axes.

6.4. Appendicitis

The data set [32] contains 106 instances. The whole set is corresponded to the appendix diagnosis. Proportion of the diagnoses in data set is as follows:

- acute appendicitis – 88 cases,

- other medical problem – 18 cases.

Every data case is a vector of eight parameters. Seven of them are numerical result of some medical tests. The last parameter is the diagnosis label ($l = 1, 2$). It is possible to create 127 rule premises for single diagnosis. The results of rule elimination algorithm for the entire data set are shown in Tab. 19.

The best efficiency is obtained for three focal elements regardless of the membership function shapes. It is demonstrated in Fig. 14 where fuzzy focal element sets created for the "acute appendicitis" diagnosis are compared. The same efficiency of 88.68% is obtained for the triangular and the Gaussian combination shapes. Since extracted rules are the same and the optimal thresholds are almost equal it can be stated that the shapes of the triangular and the Gaussian combination function have quite similar character. Since for this database the fuzzy focal element set generated for all shapes is the same

Table 19: The results of the rule elimination algorithm for the Appendicitis data set, different membership functions and η_T, η_{BPA} thresholds

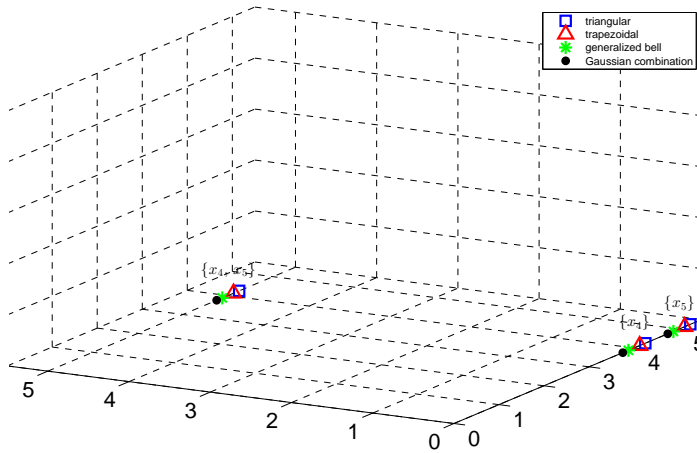| membership function	E [%]	rule set size	η_{BPA}	η_T
μ_T	**88.68**	[3 3]	0.5	0.15
μ_{TR}	84.91	[3 3]	0.45	0.35
μ_B	87.74	[3 3]	0.2	0.05
μ_Γ	**88.68**	[3 3]	0.5	0.05

Figure 14: Extracted fuzzy focal elements for the Appendicitis data set ("not acute appendicitis" diagnosis) for different membership function shapes. Points represent rules including membership functions of variables which numbers are denoted on axes. Points represent rules including membership functions of variables which numbers are denoted on axes.

and the set sizes are small, let us demonstrate all obtained fuzzy focal elements with Gaussian combination membership function for the both diagnoses:

- $S_1 = \{s^{(2)}, s^{(3)}, s^{(6)}\}$:

$$s^{(2)} : \text{IF } X_2 \text{ is } A_2^{(2)}, \text{ THEN } D_1, m_1(s^{(2)}) = 0.34,$$

$$s^{(3)} : \text{IF } X_3 \text{ is } A_3^{(3)}, \text{ THEN } D_1, m_1(s^{(3)}) = 0.32,$$

$$s^{(6)} : \text{IF } X_6 \text{ is } A_6^{(6)}, \text{ THEN } D_1, m_1(s^{(6)}) = 0.34.$$

- $S_2 = \{s^{(4)}, s^{(5)}, s^{(23)}\}$:

$$s^{(4)} : \text{IF } X_4 \text{ is } A_4^{(4)}, \text{THEN } D_1, m_2(s^{(4)}) = 0.33,$$

$$s^{(5)} : \text{IF } X_5 \text{ is } A_5^{(5)}, \text{THEN } D_1, m_2(s^{(5)}) = 0.34,$$

$$s^{(23)} : \text{IF } X_4 \text{ is } A_4^{(23)} \wedge X_5 \text{ is } A_5^{(23)}, \text{THEN } D_1, m_2(s^{(23)}) = 0.33.$$

Above X_k is $A_k^{(r_l)}$ conditions are described by suitable $\mu_k^{(l)}$ membership functions. Basic probability values for all extracted focal elements is approximately 0.33. Hence, conformation of r_l-th rule with the single data case increases the l-th diagnosis belief by 0.33 for it.

6.5. Pima Diabetes

The data set [24] refers to the study of Pima Indians diabetes diagnosis and contains 768 instances divided into two diagnoses: 268 instances are labelled as "tested positive for diabetes" and 500 otherwise. Every instance is nine-dimensional vector (eight attributes and the diagnosis label). All of the attributes are numerical values. It is possible to create full rule set of 255 focal elements for every diagnosis. The results of the rule elimination algorithm for the entire data set are shown in Tab. 20 and in Fig. 15.
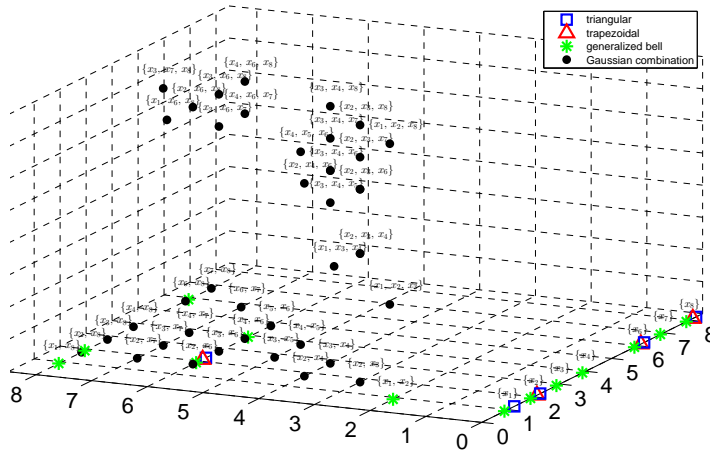

Figure 15: Extracted fuzzy focal elements for the Pima Diabetes data set ("diabetes" diagnosis) for different membership function shapes. Points represent rules including membership functions of variables which numbers are denoted on axes.

As we can see the greatest efficiency is obtained for the generalized bell function. Nevertheless, it is related to the great number of rules: 13 for both diagnoses. The result obtained for the triangular shape is related to the least rule number, with only 1.5%

Table 20: The results of the rule elimination algorithm for the Pima Diabetes data set, different membership functions and η_T, η_{BPA} thresholds

membership function	E [%]	rule set size	η_{BPA}	η_T
μ_T	76.04	[5 5]	0.55	0.2
μ_{TR}	74.74	[4 4]	0.2	0.6
μ_B	**77.34**	[13 13]	0.6	0.1
μ_Γ	74.87	[57 56]	0.35	0.05

efficiency lowering. Moreover it can be noticed again that the rule sets extracted for the less effective shapes are subsets of that created when the best shape (i.e generalized bell) is used.

7. Cross-validation

The cross-validation is executed in the same way as it has been done for the Iris data set (section 5). When the rule elimination algorithm is finished with the training data set, testing procedure with different data set is done. The statistical results of cross-validation with 50 subsets are shown in Tab. 21. Moreover, obtained testing efficiencies are presented by box-and-whiskers diagrams in Fig. 16.

The greatest efficiencies reached in testing for the cross-validation is thrice related to the same membership functions which were indicated for the whole data sets (Tab. 16÷20). A conformity is observed for the Heart Disease, Pima Diabetes and Breast Cancer data, though in the last case the maximum efficiency is achieved at the cost of great rule number. Size of a rule set is often large when the elimination algorithm fails to extract a suitable small focal element set i.e. at some step a rule is wrongly eliminated. Nevertheless, some regularities to the results obtained for the entire data set can be noticed. The average rule set numbers are often increased by the training subsets, for which the algorithm was not able to extract suitably small focal element set (see Fig. 17). This phenomenon can be noticed especially for the Breast Cancer and Appendicitis data set. Total rule number which are extracted for every training subset for mentioned data sets are shown in Fig. 17c and 17e, respectively. The median indicates the rule number neglecting of a few outlying results. This may suggest that the solutions of the best efficiency sometimes lack generalization and that the number of rules should be equally important criterion for the ultimate choice of the focal element set, i.e. membership functions and rules.

(a) The Iris data set

(b) The Thyroid Gland data set

(c) The Wisconsin Breast Cancer data set

(d) The Cleveland Heart Disease data set

(e) The Appendicitis data set
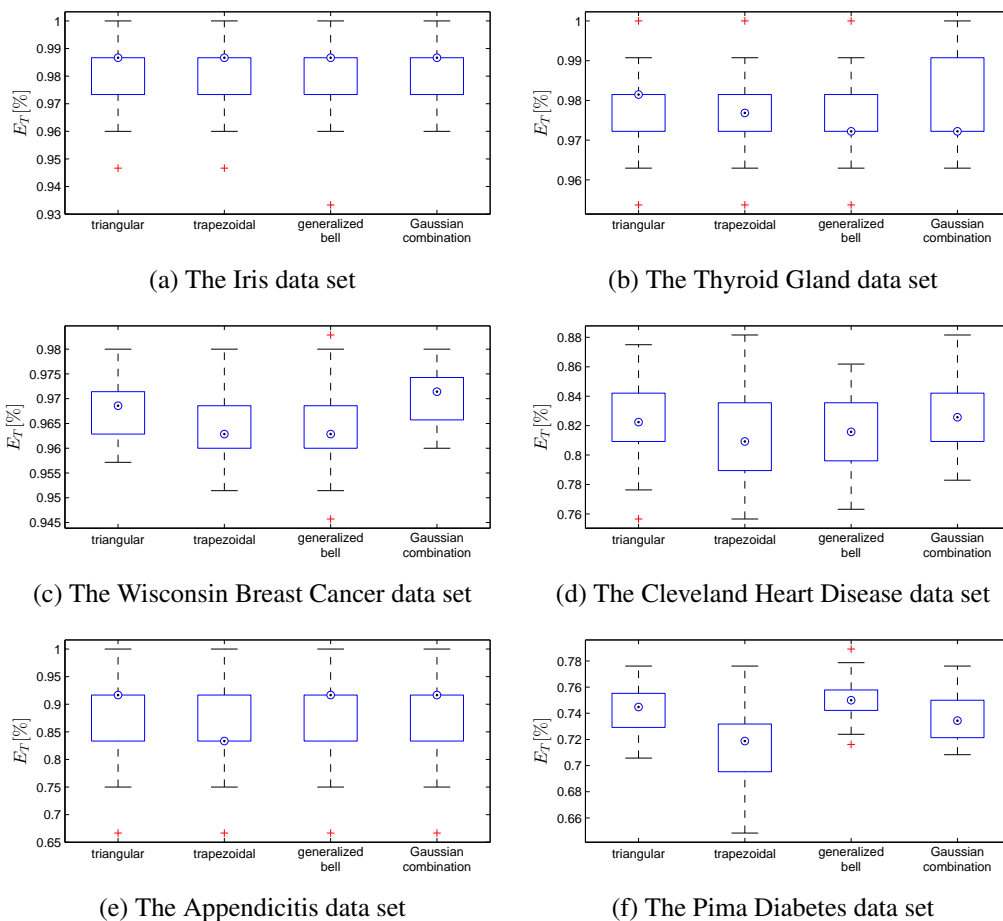

(f) The Pima Diabetes data set

Figure 16: Testing efficiency of the inference in cross-validation procedure represented by box-and-whiskers diagrams. The "box height" is determined by interquartile range, the maximum "whiskers length" is specified as one and half interquartile range. Outlying results are denoted with '+'. Dot inside circle represents median in each "box".

8.   Discussion

Let us summarize results from the point of view of membership function shapes that could be advised in application trials. In Tab. 22 and Tab. 23 the membership function which provides the best efficiency for the chosen data is denoted as „I", the second best as „II" etc.

Some regularities in the tables are noticeable. The generalized bell and trapezoidal shapes are noted with "III" and "IV" most frequently. The triangular membership functions shapes are by far the highest evaluated since they are noted with "I" for two bench-

Table 21: The results of the cross-validation for all data sets for different membership function shapes

data set	membership function	$E_T[\%]$	rule set size (mean)	rule set size (median)
Thyroid Gland	μ_T	97.91	[11 11 10]	[11 11 7]
	μ_{TR}	97.93	[13 12 12]	[12 11 10]
	μ_B	97.91	[15 14 14]	[15 12 11]
	μ_Γ	**97.94**	[11 10 10]	[11 9 8]
Breast Cancer	μ_T	96.78	[17 18]	[7 7]
	μ_{TR}	96.47	[14 14]	[7 7]
	μ_B	96.45	[28 28]	[7 7]
	μ_Γ	**97.00**	[36 36]	[14 14]
Heart Disease	μ_T	82.50	[8 8]	[6 6]
	μ_{TR}	81.09	[6 6]	[5 5]
	μ_B	81.38	[14 14]	[6 6]
	μ_Γ	**82.67**	[10 10]	[6 6]
Appendicitis	μ_T	88.33	[14 12]	[11 5]
	μ_{TR}	85.83	[10 10]	[10 10]
	μ_B	**90.17**	[21 19]	[17 17]
	μ_Γ	89.67	[18 18]	[17 17]
Pima Indians	μ_T	74.23	[4 4]	[3 3]
	μ_{TR}	71.34	[3 3]	[3 3]
	μ_B	**75.04**	[4 4]	[3 3]
	μ_Γ	73.70	[6 6]	[4 4]

mark data sets (Heart Disease and Appendicitis) and with 'II' for the others. The Gaussian combination shape is chosen as the best for two databases, too. Hence it seems that the two distinguished shapes should be preferred for further research in diagnostic rule extraction issues.

The cross-validation procedure shows us that the membership function shapes are thrice evaluated by the same notes as for the entire data sets. Moreover, the trapezoidal and generalized bell shapes again receive most often the "III" and "IV" note, respectively. The best noted shape seems to be the Gaussian combination this time. It obtains the first note for two data sets (Thyroid Gland and Breast Cancer) and the second note
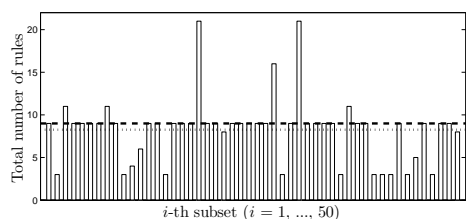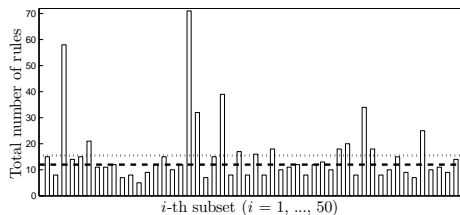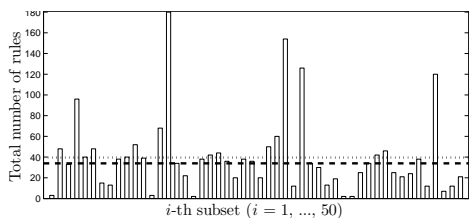
(a) The Iris data set
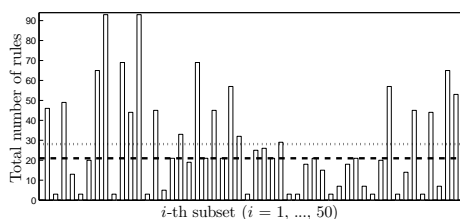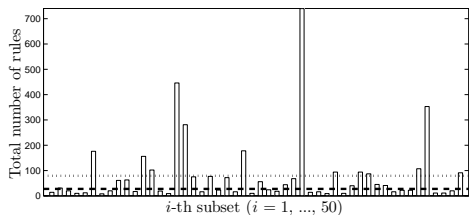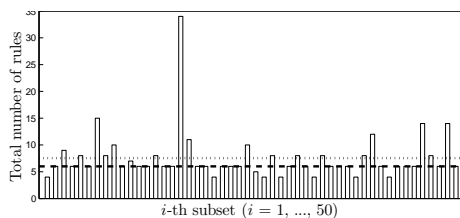

(b) The Thyroid Gland data set

(c) The Wisconsin Breast Cancer data set

(d) The Cleveland Heart Disease data set

(e) The Appendicitis data set

(f) The Pima Diabetes data set

Figure 17: Total number of the fuzzy focal elements in every data subset of the cross-validation. Horizontal dotted line and dashed line represent mean and median number of the total rule number for all subsets, respectively.

Table 22: Comparison of the best methods chosen for the benchmark data sets (when the entire data set is processed)

	Methods			
	triangular	trapezoidal	generalized bell	Gaussian combination
Thyroid Gland	II	**I**	III	IV
Breast Cancer	II	IV	III	**I**
Heart Disease	**I**	IV	III	II
Appendicitis	**I**	III	IV	**I**
Pima Diabetes	II	IV	**I**	III

Table 23: Comparison of the best methods chosen for the benchmark data sets (cross-validation procedure)

	Methods			
	triangular	trapezoidal	generalized bell	Gaussian combination
Thyroid Gland	III	II	III	**I**
Breast Cancer	II	IV	III	**I**
Heart Disease	**I**	IV	III	II
Appendicitis	III	IV	**I**	II
Pima Diabetes	II	IV	**I**	III

for the other two data sets (Heart Disease and Appendicitis). The results of the triangular shape are slightly worse if compared with results presented in Tab. 22. Anyway, they are still not worse that results obtained for the trapezoidal and generalized bell shapes. Taking into account similarity of Gaussian and triangular shapes and possibility of converting the one shape into another [26] we can confirm the usefulness of such functions.

The second remarkable issue is so-called "trade-off" between the efficiency of the diagnosis and the complexity of the fuzzy focal element (rule) set. The most significant example of this phenomenon is observed for the Cleveland Heart Disease data set, where the algorithm eliminates rules from 8191 rule set. Few percentage decrease of mistaken diagnoses is achieved at the cost of the increase at several dozen (or hundreds) rules.

The obtained results seem to be valuable while comparing to the references. When the Iris data set is considered, efficiency results could be compared to number of approaches. Efficiency obtained for the entire data set (98.00%) is higher than approaches presented in [1] or [14] where it was 95 and 94.67%, respectively. The approach based on SSV method proposed in [10] provides two rules per class. Approaches presented in [17] are based on fuzzy clustering and fuzzy based classification method. They provide 98% efficiency with about three rules per class. The results obtained for the Thyroid Gland data set (98.14 and 97.94% for the entire data set and cross-validation, respectively) are better than results presented in [13, 18] and competitive to approach presented in [21] where so-called interval-valued fuzzy reasoning model is introduced. Wisconsin Breast Cancer data set is not such popular, but our results can be compared to approaches (92-99%) presented and referenced in [10] for different rule set sizes. Although the efficiency results for the Heart Disease and Appendicitis data sets are not enough satisfactory, they are better or competitive to approaches presented and referenced in [19, 25].

9.   Conclusions

The paper focuses on extraction of diagnostic rules from databases using the Dempster-Shafer theory and fuzzy set theory. It is shown that such rules can be determined and are easy for heuristic interpretation. Different membership functions that are used in testing turned out to be optimal for various data sets, still the triangular and Gaussian combination functions appear to be preferable for initial investigations. These functions provide creation of few rule sets for five medical benchmark databases. The elimination algorithm is proposed to extract the smallest set of necessary diagnostic rules out of their complete set. It is noticeable that regardless of the functions, the resulting rule sets are the same or similar. In majority of calculations greater sets include as subsets smaller sets, obtained for different membership functions. In the performed experiment, in several cases an increase of the diagnosis accuracy is achieved at the cost of a significant enlargement of the rule set size. In these cases a little decrease of efficiency allow for great decrease of rule number which may evidence that small rule sets are more general and spare rules are specific for data. The latter are not always desirable. Hence it could be indicated to extract the optimal rule sets firstly and then try to match function shapes and tune them to obtain the best inference efficiency. Generally, for the mentioned benchmarks, the obtained accuracy is better or comparable to other methods of rule extraction. The form of rules, the used membership functions that can be described by exemplary linguistic variables "low", "normal" and "high", as well as the limited number of rules, create optimal conditions of cooperation between a knowledge engineer and an expert. Further research of the Dempster-Shafer theory based inference is worthy of interest in view of these results. It could indicate a method to create the set of rules starting from the initial simplest rules instead of rule elimination from the maximal set.

References

[1] V. ANTOINE, N. LABROCHE and V. VIET-VU: Evidential seed-based semi-supervised clustering. In *15th Int. Symp. on Soft Computing and Intelligent Systems (SCIS), Joint 7th Int. Conf. on and Advanced Intelligent Systems (ISIS), 2014*, (2014), 706-711.

[2] J. C. BEZDEK: Fuzzy models – what are they, and why? [editorial]. *IEEE Trans. on Fuzzy Systems*, **1**(1), (1993), 1-6.

[3] K. BOEGL, K.-P. ADLASSNIG, Y. HAYASHI, T.E. ROTHENFLUH and H. LEITICH: Knowledge acquisition in the fuzzy knowledge representation framework of a medical consultation system. *Artificial Intelligence in Medicine*, **30**(1), (2004), 1-26.

[4] A.H. CHEN, S.Y. HUANG, P.S. HONG, C.H. CHENG and E.J. LIN: Hdps: Heart disease prediction system. In *Computing in Cardiology, 2011*, (2011), 557-560.

[5] D. COOMANS: Thyroid Gland Data. https://archive.ics.uci.edu/ml/machine-learning-databases/thyroid-disease/new.thyroid.data, 1992. [Online; accessed 18-February-2016].

[6] K. CPAŁKA, O. REBROVA, R. NOWICKI and L. RUTKOWSKI: On design of flexible neuro-fuzzy systems for nonlinear modelling. *Int. J. of General Systems*, **42**(6), (2013), 706-720.

[7] A. P. DEMPSTER: A generalization of bayesian inference. *J. of the Royal Statistical Society. Series B (Methodological)*, **30**(2), (1968), 205-247.

[8] R. DETRANO: Heart Disease Database (Cleveland). https://archive.ics.uci.edu/ml/machine-learning-databases/heart-disease/processed.cleveland.data, 1990. [Online; accessed 18-February-2016].

[9] D. DUBOIS and H. PRADE: Mathematical aspects of fuzzy sets random sets and fuzzy interval analysis. *Fuzzy Sets and Systems*, **42**(1), (1991), 87-101.

[10] W. DUCH, R. ADAMCZAK and K. GRABCZEWSKI: A new methodology of extraction, optimization and application of crisp and fuzzy logical rules. *IEEE Trans. on Neural Networks*, **12**(2), (2001), 277-306.

[11] M. FATHURACHMAN, U. KALSUM, N. SAFITRI and C.P. UTOMO: Heart disease diagnosis using extreme learning based neural networks. In *Int. Conf. of Advanced Informatics: Concept, Theory and Application (ICAICTA), 2014*, (2014), 23-27.

[12] R. A. FISHER: Iris Plants Database. https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data, 1988. [Online; accessed 10-February-2016].

[13] QINGHUA HU, D. ZHANG, SHUANG AN, D. ZHANG and DAREN YU: On robust fuzzy rough set models. *IEEE Trans. on Fuzzy Systems*, **20**(4), (2012), 636-651.

[14] S. KANTARCI, A. VAHAPLAR, A. O. KINAY and E. NASIBOV: Influence of t-norm and t-conorm operators in fuzzy id3 algorithm. In *IEEE Int. Conf. on Fuzzy Systems (FUZZ-IEEE), 2015*, (2015), 1-6.

[15] L. FANG and Z. ZHIGUANG: A new data classification method based on chaotic particle swarm optimization and least square-support vector machine. *Chemometrics and Intelligent Laboratory Systems*, **147** (2015), 147-156.

[16] A.L. MEDAGLIA, S.-C. FANG, H.L.W. NUTTLE and J.R. WILSON: An efficient and flexible mechanism for constructing membership functions. *European J. of Operational Research*, **139**(1), (2002), 84-95.

[17] C. MENCAR, C. CASTIELLO, R. CANNONE and A.M. FANELLI: Design of fuzzy rule-based classifiers with semantic cointension. *Information Sciences*, **181**(20), (2011), 4361-4377. Special Issue on Interpretable Fuzzy Systems.

[18] K. NESHATIAN, MENGJIE ZHANG and P. ANDREAE: A filter approach to multiple feature construction for symbolic learning classifiers using genetic programming. *IEEE Trans. on Evolutionary Computation*, **16**(5), (2012), 645-661.

[19] KEMAL POLAT and SALIH GÜNES: A new feature selection method on classification of medical datasets: Kernel f-score feature selection. *Expert Systems with Applications*, **36**(7), (2009), 10367-10373.

[20] I. ROJAS, H. POMARES, E. ROS and A. PIETRO: Automatic construction of fuzzy rules and membership functions from training examples. In *Proc. EUFIT'98*, (1998), 618-622.

[21] J.A. SANZ, A. FERNANDEZ, H. BUSTINCE and F. HERRERA: Ivturs: A linguistic fuzzy rule-based classification system based on a new interval-valued fuzzy reasoning method with tuning and rule selection. *IEEE Trans. on Fuzzy Systems*, **21**(3), (2013), 399-411.

[22] G. SHAFER: *A mathematical theory of evidence*. Princeton University Press, New Jersey, 1976.

[23] J. SHEN, W. SHEN, H.J. SUN and J.Y. YANG: Fuzzy neural nets with non-symmetric π membership functions and applications in signal processing and image analysis. *Signal Processing*, **80**(6), (2000), 965-983.

[24] V. SIGILLITO: Pima Indians Diabetes Database. https://archive.ics.uci.edu/ml/machine-learning-databases/pima-indians-diabetes/pima-indians-diabetes.data, 1990. [Online; accessed 18-February-2016].

[25] M SIKORA: *Transactions on Rough Sets XI*, volume 5946, chapter Decision rule-based data models using TRS and NetTRS – methods and algorithms, pages 130-160. Springer Berlin Heidelberg, Berlin, Heidelberg, 2010.

[26] J. ŁĘSKI: *Systemy neuronowo-rozmyte*. WNT, Warsaw, 2008. (in Polish).

[27] E. STRASZECKA: Combining uncertainty and imprecision in models of medical diagnosis. *Information Sciences*, **176**(20), (2006), 3026-3059.

[28] E. STRASZECKA: Combining knowledge from different sources. *Expert Systems*, **27**(1), (2010), 40-52.

[29] E. STRASZECKA and J. STRASZECKA: Medical diagnosis support using uncertainty and imprecision measures. In *IEEE Int. Conf. on Systems, Man and Cybernetics, 2004*, **6** (2004), 5835-5840.

[30] E. STRASZECKA and J. STRASZECKA: *Computer Recognition Systems: Proceedings of the 4th Int. Conf. on Computer Recognition Systems CORES '05*, chapter Interpretation of Medical Symptoms Using Fuzzy Focal Elements, pages 287–293. Springer Berlin Heidelberg, Berlin, Heidelberg, 2005.

[31] T. TAKAGI and M. SUGENO: Fuzzy identification of systems and its applications to modeling and control. *IEEE Trans. on Systems, Man, and Cybernetics*, SMC-**15**(1), (1985), 116-132.

[32] S.M. WEISS and I. KAPOULEAS: An empirical comparison of pattern recognition, neural nets, and machine learning classification methods. In *Proc. of the Eleventh Int. Joint Conf. on Artificial Intelligence*, (1989), 781-787.

[33] W. H. WOLBERG: Wisconsin Breast Cancer Database (Original). https://archive.ics.uci.edu/ml/machine-learning-databases/breast-cancer-wisconsin/breast-cancer-wisconsin.data, 1992. [Online; accessed 18-February-2016].

[34] R. R. YAGER: Uncertainty representation using fuzzy measures. *IEEE Trans. on Systems, Man, and Cybernetics, Part B (Cybernetics)*, **32**(1), (2002), 13-20.

[35] J. YEN: Generalizing the dempster-schafer theory to fuzzy sets. *IEEE Trans. on Systems, Man, and Cybernetics*, **20**(3), (1990), 559-570.

[36] YUNONG ZHANG, YONGHUA YIN, DONGSHENG GUO, XIAOTIAN YU and LIN XIAO: Cross-validation based weights and structure determination of chebyshev-polynomial neural networks for pattern classification. *Pattern Recognition*, **47**(10), (2014), 3414-3428.