

Learning fuzzy measures for aggregation in fuzzy rule-based models

Emran Saleh¹, Aida Valls¹, Antonio Moreno¹,
Pedro Romero-Aroca², Vicenç Torra³, and Humberto Bustince⁴

¹ Departament d'Enginyeria Informàtica i Matemàtiques, Universitat Rovira i Virgili
`emran.saleh@urv.cat`, `aida.valls@urv.cat`, `antonio.moreno@urv.cat`

² Ophthalmic Service, University Hospital Sant Joan de Reus, Institut d'Investigació
Sanitària Pere Virgili (IISPV), Universitat Rovira i Virgili
`pedro.romero@urv.cat`

³ School of informatics, University of Skövde, Skövde, Sweden
`vtorra@his.se`

⁴ Departamento de Automática y Computación, Universidad Pública de Navarra,
Institute of Smart Cities, Universidad Pública de Navarra
`bustince@unavarra.es`

Abstract. Fuzzy measures are used to express background knowledge of the information sources. In fuzzy rule-based models, the rule confidence gives an important information about the final classes and their relevance. This work proposes to use fuzzy measures and integrals to combine rules confidences when making a decision. A Sugeno λ -measure and a distorted probability have been used in this process. A clinical decision support system (CDSS) has been built by applying this approach to a medical dataset. Then we use our system to estimate the risk of developing diabetic retinopathy. We show performance results comparing our system with others in the literature.

Keywords: Fuzzy Measures, Aggregation functions, Choquet Integral, Sugeno Integral, Fuzzy rule-based Systems, Diabetic Retinopathy.

1 Introduction

Aggregation operators are mathematical functions to merge a set of numerical arguments into a single one that summarizes them. They are widely used in many knowledge fields, such as sensor data fusion and decision making [16].

Among the vast number of aggregation operators found in the literature, fuzzy integrals are one of the most general ones. Due to their parametrisation, fuzzy integrals as Choquet and Sugeno show a great flexibility in aggregating the inputs. Choquet integral generalizes both the weighted mean and the OWA operator [2] and Sugeno integral generalizes weighted maximum, weighted minimum and the median operators [4],[16].

Fuzzy integrals rely on a fuzzy measure (or capacity), which is a set function that indicates the importance of the information sources (i.e., of each of possible

groups of input sources). Defining a proper fuzzy measure for each problem is a crucial point in order to make a suitable aggregation of the inputs and obtain the correct corresponding outputs.

In this paper we will focus on solving a classification problem in the medical field. We have been working on the definition and construction of a clinical decision support system for improving the diagnosis of diabetic retinopathy (DR). This disease is one of the major complications of diabetes and one of the most important causes of loss vision in young diabetic people all over the world. The effects of this disease can be controlled if it is detected at an early stage. With the collaboration of experts from different medical centers in Catalonia we have collected a dataset of information of 3,000 diabetic patients. This data has been used to train and test a binary DR classification model using Fuzzy Random Forests (FRF) [12].

One of the characteristics of FRF is that a large number of classification rules are generated using different samples of the data. In our case, we have 100 trees with about 100 rules each one. When a new patient has to be classified, his data is introduced into the system and all rules are fired at different levels of satisfaction. Merging the outcome of all these rules is usually done with the Winner strategy, which consists on taking as answer the output of the rule with maximum activation [11]. However, the information provided by the rest of rules is lost.

In this paper we propose to use other aggregation methods in order to merge the contribution of the different rules that are activated by a certain patient's data. In particular, we study the use of fuzzy integrals and a new way of constructing the fuzzy measure is proposed, based on the confidence score of each of the contributing rules.

The rest of the paper is organized as follows. Section 2 presents the main concepts used in this work. In section 3, we introduce the induction algorithms of fuzzy decision trees (FDT) and fuzzy random forest (FRF) models, the proposed fuzzy measures and aggregation process. In Section 4, we describe the dataset and discuss the experimental results. Finally, section 5 shows the conclusion and future work.

2 Preliminaries

In this section, we define the basic concepts that are used in this work. We follow [16] for the definitions.

Definition 1 *A function $agg : [0, 1]^n \rightarrow [0, 1]$ is an aggregation function if and only if it fulfills the following properties:*

- $agg(x, \dots, x) = x$ (*Identity*)
- $agg(0, \dots, 0) = 0$ and $agg(1, \dots, 1) = 1$ (*Boundary conditions*)
- If $(x_1, \dots, x_n) \leq (y_1, \dots, y_n)$ then $agg(x_1, \dots, x_n) \leq agg(y_1, \dots, y_n)$ (*Non decreasing*)

Note that some authors use identity only in 0 and 1 as eg. [1] and [6].

Definition 2 A function $T : [0, 1]^2 \rightarrow [0, 1]$ is a *t-norm function* if and only if it fulfills the following properties:

- $T(x, y) = T(y, x)$ (Commutativity)
- $T(x, y) \leq T(u, v)$ if $(x \leq u)$ and $(y \leq v)$ (Increasing monotonicity)
- $T(x, T(y, z)) = T(T(x, y), z)$ (Associativity)
- $T(x, 1) = x$ (Neutral element)

Examples of T-norms include minimum and product.

Definition 3 A fuzzy measure (also known as non-additive measure) m on a set X with cardinality η is a set function $m : 2^X \rightarrow [0, 1]$ fulfilling the following properties:

- $m(\emptyset) = 0, m(X) = 1$, (Boundary condition)
- $A \subseteq B$ implies $m(A) \leq m(B)$, for all $A, B \subset X$ (Monotonicity)

Fuzzy measures are a way to represent background knowledge about the importance of the sources of some values. In that way, they are used to weight the arguments in some aggregation operators like the Choquet and Sugeno integrals. The fuzzy measure can be defined manually or it can be obtained from some domain data.

In [6], it is proposed a fuzzy measure obtained as the power mean of the cardinality of the set of values aggregated. This fuzzy measure $m_{PM} : 2^X \rightarrow [0, 1]$ is defined as follows:

$$m_{PM}(A) = \left(\frac{|A|}{\eta} \right)^q \quad \text{with } q > 0 \quad (1)$$

For classification problems, the value of q can be optimized for each of the classes considered [1].

In this work, we use Choquet and Sugeno integrals in order to aggregate the input data. The discrete Choquet integral is defined as:

Definition 4 Let X be a reference set with cardinality η and let m be a fuzzy measure on X ; then, the Choquet integral of a function $f : X \rightarrow \mathbb{R}^+$ with respect to the fuzzy measure m is defined by

$$Choquet(f) = \sum_{i=1}^{\eta} [f(x_{s(i)}) - f(x_{s(i-1)})] \cdot m(A_{s(i)}), \quad (2)$$

where $f(x_{s(i)})$ indicates that the indices have been permuted so that $0 \leq f(x_{s(1)}) \leq \dots \leq f(x_{s(\eta)}) \leq 1$, and where $f(x_{s(0)}) = 0$ and $A_{s(i)} = \{x_{s(i)}, \dots, x_{s(\eta)}\}$.

For the sake of simplicity, we will use $Choquet(x_1, \dots, x_\eta)$.

In [6] the Choquet-like Copula-based fuzzy integral (CC-integral) is defined. It uses a copula as main operator \bullet instead of the product \cdot as usual in the Choquet integral. I. e., when $\bullet = \cdot$ the CC-integral is Choquet integral. The properties of this extended fuzzy integral have been studied in [5] and [7].

Definition 5 Let X be a reference set with cardinality η and let m be a fuzzy measure on X ; then, the CC-integral of a function $f : X \rightarrow \mathbb{R}^+$ with respect to the fuzzy measure m is defined by

$$CC\text{-integral}(f) = \sum_{i=1}^{\eta} [f(x_{s(i)}) \bullet m(A_{s(i)}) - f(x_{s(i-1)}) \bullet m(A_{s(i)})], \quad (3)$$

where $f(x_{s(i)})$ indicates that the indices have been permuted so that $0 \leq f(x_{s(1)}) \leq \dots \leq f(x_{s(\eta)}) \leq 1$, and where $f(x_{s(0)}) = 0$ and $A_{s(i)} = \{x_{s(i)}, \dots, x_{s(\eta)}\}$.

For the sake of simplicity, we will use $CC\text{-integral}(x_1, \dots, x_\eta)$.

Definition 6 Let m be a fuzzy measure on X with cardinality η ; then, the Sugeno integral of a function $f : X \rightarrow [0, 1]$ with respect to m is defined by

$$Sugeno(f) = \max_{i=1, \eta} \min(f(x_{s(i)}), m(A_{s(i)})), \quad (4)$$

where $f(x_{s(i)})$ indicates that the indices have been permuted so that $0 \leq f(x_{s(1)}) \leq \dots \leq f(x_{s(\eta)}) \leq 1$ and $A_{s(i)} = \{x_{s(i)}, \dots, x_{s(\eta)}\}$.

For the sake of simplicity, we will use $Sugeno(x_1, \dots, x_\eta)$.

3 Methodology

In this section, we explain how to build a fuzzy decision tree (FDT) and fuzzy random forest (FRF) and then we propose how to use fuzzy integrals to merge the conclusions of the rules when the FDT or FRF is used to classify a new instance.

3.1 Fuzzy rule-based models construction

There are many techniques to build fuzzy rule-based models. In this section, we describe the main steps of constructing them based on Yuan and Shaw [17]. That is, an induction method is used to build a fuzzy decision tree (FDT), and a bag of FDTs is used to build a fuzzy random forest (FRF). The following is the notation which is used in the induction procedure.

Let us consider a set of labeled examples $U = \{u_1, u_2, \dots, u_m\}$. Each u_i is an example described by attributes $A = \{a_1, \dots, a_n\}$.

Each attribute $a \in A$ takes values on a linguistic fuzzy partition [3] $T_a = \{t_1, \dots, t_s\}$ with membership functions μ_{t_i} . The membership values on the universe can be understood as a possibility distribution.

The U -uncertainty (or non-specificity measure) of a possibility distribution π on any set with cardinality d is defined in [17] as:

$$g(\pi) = \sum_{i=1}^d (\pi_i^* - \pi_{i+1}^*) \ln i \quad (5)$$

where $\pi^* = \{\pi_1^*, \pi_2^*, \dots, \pi_d^*\}$ is a permutation of $\pi = \{\pi(1), \pi(2), \dots, \pi(d)\}$ such that $\pi_i^* \geq \pi_{i+1}^*$, for $i = 1, \dots, d$, and $\pi_{d+1}^* = 0$.

Fuzzy Decision Tree Induction The induction algorithm proposed in [17] is an extension of the classic ID3 method for crisp data. It incorporates two parameters to manage the uncertainty:

- The *significance level* (α) is used to ignore insignificant evidences. If the fuzzy evidence membership value is lower than α then turns it to 0.
- The *truth level threshold* (β) controls the growth of the tree. Very high β may lead to overfitting and very low β may lead to low classification accuracy.

The main steps of the fuzzy decision tree induction process are the following ones:

1. Choose the attribute with the **smallest ambiguity** (see the expression below) for the root node.
2. For each value of the attribute, create a branch if it has examples with support higher than α .
3. For each branch, calculate the truth level of classification to each class.
4. If **the truth level of classification** (see the expression below) is higher than β then end the branch with the class label which has the highest truth level of classification.
5. If no then check if an additional attribute will reduce the classification ambiguity.
6. If so, choose the attribute with **smallest classification ambiguity with the accumulated evidence** (see the expression below) for the new node, and repeat from step 2 to 6 until no more tree growth is possible.
7. If no, end the branch with the label of the class that has the highest truth level of classification.

The ambiguity of an attribute is calculated as an average of the uncertainty of this attribute for an example using the following equation:

$$Ambiguity(a) = \frac{1}{m} \sum_{j=1}^m g(\pi_j),$$

where

$$\pi_j = \{\mu'_{t_1}(u_j), \dots, \mu'_{t_s}(u_j)\} \quad (6)$$

and $\mu'_{t_i}(u_j)$ is the normalized possibility distribution of $\mu_{t_i}(u_j)$:

$$\mu'_{t_i}(u_j) = \mu_{t_i}(u_j) / \max_{1 \leq k \leq s} \{\mu_{t_k}(u_j)\} \quad (7)$$

The truth level of classification defines the possibility of classifying an object u_i into a class $C_k \in C$ where $C = \{C_1, \dots, C_p\}$ given the fuzzy evidence E .

$$Truth(C_k|E) = S(E, C_k) / \max_{1 \leq j \leq p} \{S(E, C_j)\} \quad (8)$$

where S is the subsethood of the fuzzy set X on the fuzzy set Y

$$S(X, Y) = \frac{M(X \cap Y)}{M(X)} = \frac{\sum_{i=1}^m \min(\mu_X(u_i), \mu_Y(u_i))}{\sum_{i=1}^m \mu_X(u_i)} \quad (9)$$

and $M(X)$ is the cardinality or sigma count of the fuzzy set X . The truth level of classification can be understood as the possibility distribution on the set U . $\pi(C|E)$ is the normalisation of the truth level. It has been defined above to be used in the calculation of *Classification ambiguity*.

Classification ambiguity: Suppose we have a fuzzy partition $P = \{E_1, \dots, E_k\}$ on a fuzzy evidence F , the following equation is used to calculate the classification ambiguity of a fuzzy partition on a fuzzy evidence denoted by $G(P|F)$.

$$G(P|F) = \sum_{i=1}^k W(E_i|F)g(\pi(C|E_i \cap F)) \quad (10)$$

where $W(E_i|F)$ is the weight. The weight is calculated using the following equation: $W(E_i|F) = M(E_i \cap F) / \sum_{i=1}^k M(E_i \cap F)$.

Fuzzy Random Forests construction :

The main steps to build a fuzzy random forest are as follows:

1. Randomly, select a subset of the training examples (bootstrap) for training. It has to have a balanced distribution of each class. It is recommended that the size of each dataset (bootstrap) has to be 2/3 of the total training dataset size. The repetition of examples is acceptable. Use each bootstrap to construct a fuzzy decision tree (see 3.1).
2. While constructing the FDT, a random subset of the remaining attributes with size γ will be used when deciding for next tree node.
3. Repeat steps 1 and 2 until the number of the FDTs n is reached.

3.2 Fuzzy measure based on the rule confidence

Fuzzy measures are used to give background knowledge in relation to the elements which are going to be aggregated. In our context, we aggregate data from a set of rules and we have a degree of support for each rule (*rule confidence*). These degrees define a possibility distribution of the data. These values give an important information about the system. Taking them into account while we are giving the decision of the model is valuable. In this paper we propose the use of two fuzzy measures that will be built from these *rule confidence* values. The first measure is a distorted probability. The second measure is a Sugeno λ -measure.

Let us define the notation used in the following equations: R = the total set of all rules, RC_i = *Rule Confidence* of the i th rule and n = the total number of rules.

Distorted probability based fuzzy measures :

The proposed distorted probability is defined using the following equation:

$$m_{DP}(A) = \left(\frac{\sum_{RC_j \in A} RC_j}{\sum_{RC_i \in R} RC_i} \right)^q, \text{ with } q > 0 \quad (11)$$

where the value q needs to be optimised. Different methods can be used to optimize q like evolutionary algorithms [6],[1]. We use here a gradient descent and wide search. Note that this fuzzy measure is a distorted probability because $m = f \odot P$ with

$$P_j = \frac{RC_j}{\sum_{RC_i \in R} RC_i}, \quad \text{and } f(x) = x^q \quad (12)$$

Sugeno λ -measures based fuzzy measures :

Another way of using domain knowledge to construct a fuzzy measure is by means of the definition of a Sugeno λ -measure as proposed in [16].

Definition 7 Let $v : X \rightarrow [0, 1]$ and $\lambda > -1$ be such that

$$\begin{aligned} - (1/\lambda)(\prod_{x_i \in X} [1 + \lambda v(x_i)] - 1) &= 1 & \text{if } \lambda \neq 0 \\ - \sum_{x_i \in X} v(x_i) &= 1 & \text{if } \lambda = 0 \end{aligned}$$

then, the fuzzy measure defined by

$$m_{SL}(A) = \begin{cases} v(x_i) & \text{if } A = \{x_i\} \\ (1/\lambda)(\prod_{x_i \in A} [1 + \lambda v(x_i)] - 1) & \text{if } |A| \neq 1 \text{ and } \lambda \neq 0 \\ \sum_{x_i \in A} v(x_i) & \text{if } |A| \neq 1 \text{ and } \lambda = 0 \end{cases} \quad (13)$$

is a Sugeno λ -measure. In our proposal, the weights $v(x_i) = RC_i$ are the *rule confidence* values. Therefore, first the *rule confidence* values are used to build the Sugeno λ -measures finding an appropriate λ and then this fuzzy measure is used in the aggregation process.

3.3 Classification using the fuzzy rules

A binary classification is done using the Mamdani inference procedure. Class 0 represents that patients do not suffer from DR and class 1 that they suffer. All rules are applied and the rule membership degree to the conclusion class (*RMCC*) values of the same class are aggregated to obtain the final decision. The proposed procedure is the following:

1. Use a t-norm function to calculate the satisfaction degree of each rule $\mu_R(u)$.

2. Use the product between the satisfaction degree of each rule $\mu_R(u)$ and the degree of support of the rule (*rule confidence*) to obtain the membership degree to the conclusion class (*RMCC*).
3. Calculate a fuzzy measure (distorted probability $m_{DP}(A)$ or Sugeno λ -measure $m_{SL}(A)$) using the degree of support of the rule (rule confidence).
4. Aggregate the final value of each class using a fuzzy integral (CC-integral (equation 3) or Sugeno integral (equation 4)). In the aggregation process, the obtained *RMCCs* from the same class are weighted using fuzzy measures as explained above.
5. Compare the aggregation values, the final decision is the class label which has the maximum aggregation value.

4 Experimental results

In this section, we describe the data used to train and validate the proposed models. The results achieved by these models are discussed in 4.2.

4.1 The diabetic retinopathy problem and dataset

An early diagnosis of DR is crucial to improve the quality of life of these patients. At the moment, the detection of DR is done by screening of the eye fundus with a non-midriatic camera. This technique requires a lot of resources from the medical centers both in terms of cost, specialized personnel and time [9]. Due to the large amount of diabetic people it is not possible to perform this test early as recommended by the medical guidelines. Therefore, tests are done every two or three years. For some patients, the detection arrives too late.

The clinical decision support system that we are developing may significantly decrease these costs because it will be used by the family physicians during the regular visits that diabetic people have. The incidence of DR is scarce, which means that most of the people do not need an eye fundus screening. Therefore, the resources could be used to the patients that really need it, facilitating the detection of DR in its first signs.

Sant Joan de Reus University Hospital (SJRUH) in Catalonia (Spain) has been systematically collecting the data of the diabetic patients of many years. These data include demographic, metabolic and analytical information which is stored in the Electronic Health Records (EHR) of the people who has diabetes. The dataset used in the work has the information of 3346 diabetic patients and it is labeled regarding to diabetic retinopathy presence. This dataset has been split into a training dataset with 2243 diabetic patients (1605 not suffering from RD and 638 who suffer from DR) and a testing dataset with 1103 examples (863 not suffering from RD and 240 who suffer from DR). The datasets are imbalanced because the patients with DR are less than healthy people. This imbalance distribution hampers the performance of some machine learning techniques. To solve this problem with FDT model, random over sampling technique has been done to the minor class until both classes have the same number of examples. FRF

model internally does random under sampling technique which is a technique to deal with imbalanced datasets.

A statistical analysis on the data was done by the ophthalmologists in SJRUH [8]. Out of that study, nine attributes were identified as the important ones to detect the risk of RD development. Most of the attributes are numerical but there are some categorical ones too. With the collaboration of the experts, the numerical attributes have been fuzzified into linguistic variables according to the medical knowledge.

4.2 Tests, results and discussion

In this section, we study the results achieved by using the proposed aggregators with FRF and FDT models on the testing dataset. A comparison between the different aggregation proposals and the traditional methods is done as well. The aim is to improve the performance of the models and achieve a good performance that is acceptable in the medical treatments.

To evaluate the performance of the models on such kind of problems, we use specificity and sensitivity (*recall*). They are usually used in the medical field. To make it easier to the reader to follow the performance results, the harmonic mean (HM) of specificity and sensitivity is calculated as well (Equation 14).

$$\begin{aligned} \text{Sensitivity} &= \frac{TP}{TP + FN}, & \text{Specificity} &= \frac{TN}{TN + FP}, \\ \text{HM} &= 2 * \frac{\text{Sensitivity} * \text{Specificity}}{\text{Sensitivity} + \text{Specificity}} \end{aligned} \quad (14)$$

The parameters of FDT and FRF were studied in previous works [14], [11] and the best values have been used in this paper.

In FRF two ways of aggregating the outcome of the trees have been tested. On the one hand, the direct aggregation of all the rules of all the trees into a unique result (one-step). On the other hand, the aggregation first of the rules of each tree and in a second step the aggregation of the outcome of each tree (two-steps). To avoid effect of randomness, all one-step and two-steps FRF models are experimented with the same parameters 10 times then we take as result the ones of the model with the median HM performance. The median is more robust to outliers than the arithmetic mean.

The different methods tested are shown in Table 1. The basic winner rule (WR) for making decisions is well-known in rule-based models, it uses max t-conorm to aggregate the outputs of the rules. To verify the quality of the aggregation methods proposed in this work in comparison with the state of art, first the aggregation method based on Choquet-like Copula-based integral is used with the power mean as fuzzy measure (equation 1) as proposed in [6]. The rest of methods correspond to the different versions of Choquet and Sugeno integrals using the Rule Confidence (RC) for the fuzzy measure construction. Table 1 indicates the t-norm operator used in the fuzzy integral and the t-norm used in to calculate the degree of activation of each rule.

Short name	Aggregation method name	Aggregator	
		T-norm	T-norm
WR	max	-	min
CCPM	CC-integral with power mean	min	min
ICMM	<i>RC</i> based CC-integral, distorted probability	min	min
ICMMS	<i>RC</i> based CC-integral, Sugeno λ -measures	min	min
ICPM	<i>RC</i> based CC-integral, distorted probability	product	min
ICPMS	<i>RC</i> based CC-integral, Sugeno λ -measures	product	min
ISM	<i>RC</i> based Sugeno, distorted probability	min	min
ISMS	<i>RC</i> based Sugeno, Sugeno λ -measures	min	min

Table 1: Notation of the aggregation methods used in this work

q	HM	Sensit.	Specif.	Accuracy	TP	FN	FP	TN	Method
2	76.51	76.67	76.36	76.34	184	56	204	659	WR
1	77.15	76.67	77.64	77.43	184	56	193	670	CCPM
2	76.85	75.42	78.33	77.70	181	59	187	676	ICMM
2	76.89	76.67	76.31	76.38	187	53	208	655	ICMMS
3	77.65	80.00	75.43	76.43	192	48	212	651	ICPM
2	71.63	86.67	63.41	68.48	210	48	311	552	ICPMS
2	77.98	78.33	77.64	77.79	188	52	193	670	ISM
2	76.37	76.67	76.07	76.20	185	55	210	653	ISMS

Table 2: Classification results of Fuzzy Decision Tree with $\beta = 0.70$ and $\alpha = 0.30$

With each fuzzy measure, several q values were tested to find the optimal value. Notice that if $q = 1$, then the aggregation (Choquet or Sugeno integral) corresponds to the weighted mean. When q value increases, the performance of the models decreases. Low q values always showed better results, being the best ones $q = 2$, $q = 3$ for most of FRF and FDT models. The models with the best performance are highlighted in Tables 2, 3 and 4.

Observing the basic method WR (winning rule), in FRFs this aggregation method shows high specificity (around 81%) but it shows low sensitivity as well (between 71% and 73%) and HM value is around 76%. The FDT model with WR achieved specificity=76.36%, sensitivity=76.67% and HM=76.51%. By checking the models' performance in Tables 2, 3 and 4, the first conclusion is that the models based on Choquet and Sugeno integrals offer better results than WR. Method ICPMS is an exception (low HM) because it achieves a very good sensitivity but then specificity is too low to be acceptable for medical diagnosis.

Two different methods of calculating fuzzy measures have been proposed in this work. The first one is distorted probability and the second one is a Sugeno λ -measure. The results of FDTs models are presented in table 2. ISM has the best performance with a very good HM, near 78%, and has the highest sensitivity value (78.33%). We can see that models based on distorted probability improve the ones based on λ -Sugeno measures.

q	HM	Sensit.	Specif.	Accuracy	TP	FN	FP	TN	Method
2	76.05	71.67	81.00	78.97	172	68	164	699	WR
2	78.31	78.75	77.87	78.06	189	51	191	672	CCPM
3	79.77	80.42	79.14	79.42	193	47	180	683	ICMM
2	76.89	77.92	75.90	76.34	187	53	208	655	ICMMS
3	78.61	76.67	80.65	79.78	184	56	167	696	ICPM
2	71.63	81.40	63.96	67.98	210	48	311	552	ICPMS
2	79.74	78.75	80.76	80.33	189	51	166	697	ISM
2	76.37	77.08	75.67	75.97	185	55	210	653	ISMS

Table 3: Classification results of two-steps Fuzzy Random Forest with $\alpha = 0.40$, $\beta = 0.80$

q	HM	Sensit.	Specif.	Accuracy	TP	FN	FP	TN	Method
2	77.11	72.92	81.81	79.87	175	65	157	706	WR
2	76.96	75.00	79.03	78.15	180	60	181	682	CCPM
2	79.31	78.33	80.30	79.87	188	52	170	693	ICMM
2	76.59	76.25	76.94	76.79	183	57	199	664	ICMMS
2	78.10	77.08	79.14	78.69	185	55	180	683	ICPM
3	73.27	83.75	65.12	69.17	201	39	301	562	ICPMS
3	79.73	79.17	80.30	80.05	190	50	170	693	ISM
2	78.00	79.58	76.48	77.15	191	49	203	660	ISMS

Table 4: Classification results of one-step Fuzzy Random Forest with $\alpha = 0.40$, $\beta = 0.80$

In Table 3 when using FRF with two-steps aggregation, CCPM method obtained HM=77.38%, sensitivity=76.67% and specificity=78.10%. Whereas, all the distorted probability measures based on the *rule confidence* values (ICMM, ICPM and ISM) obtain HM greater than 79%. ICMM achieved the highest performance (HM around 80%). In table 4, the results of one-step FRF are presented. ICMM, ICPM achieved HM higher than 78% while and ISM has HM around 80%. Methods based on the λ -Sugeno fuzzy measure obtain quite good performance in sensitivity(see ICPMS) but the specificity decreases too much. These results clearly show that the aggregator that uses a fuzzy measure with distorted probability based on the *rule confidence* values outperforms the one using the number of the rules in the fuzzy measure (CCPM). We see a difference in the best method when using one or two steps in the aggregation of the rules of the set of trees. However, both ICMM and ISM achieve quite similar HM with values of sensitivity and specificity close to 80%.

By looking to the results presented in Tables 2,3 and 4, FRF models usually offer better performance than FDTs with every aggregation method. In FDT model with (ICPM) aggregation method, the model obtains specificity = 80%, which is higher than FRFs models results. The same model shows sensitivity=75.43% which is lower than the sensitivity obtained by FRF models. In general, FRF models show more balance in sensitivity and specificity values.

5 Conclusion and future work

The use of fuzzy measures in aggregation operators shows good performance. *Rule confidence* values showed that it can play an important role in the aggregation process.

In this work, a Sugeno λ -measure and a distorted probability are used with Choquet and Sugeno integrals. These new aggregation approaches are used within fuzzy random forests (FRF) and Fuzzy decision trees (FDT). The models with these new aggregation approaches outperforms the same models with max t-conorm aggregation operator.

In comparison with the models that use the same Choquet and Sugeno integrals with fuzzy measures based only on the number of rules, the new approach obtains better performance results as well. Experiments also showed that one-step and two-steps FRFs have better performance results than FDTs. Moreover, two-steps FRF is recommended because it offers better results than one-step FRF, with 80% of sensitivity and 79% of specificity on the testing dataset. With these results, we conclude that the new aggregation operators based on the proposed fuzzy measures improve the performance of our previous works [14], [12], [13] and [10].

This work is oriented to build a clinical decision support system (CDSS). The CDSS will be used in the medical centers by family physicians who are not expert ophthalmologist. The goal is to help the physicians to estimate the risk of developing DR with the new patients. The proposed methods can be easily integrated into the CDSS in order to merge the rule's predictions made with the data of each patient.

Future work includes studying how to improve the performance of current fuzzy measures. Sugeno λ -measures and distorted probabilities do not permit to structure the information sources. As the data to be aggregated is highly dimensional, other families of measures as the hierarchically decomposable ones can be useful [15]. We plan to work in this line. The current approach is going to be validated using other datasets in order to see if the same conclusions about the performance different proposals depend on the problem data or not.

6 Acknowledgements

This work is supported by the URV grant 2017PFR-URV-B2-60, and by the Spanish research projects no: PI12/01535 and PI15/01150 for (Instituto de Salud Carlos III and FEDER funds). Mr. Saleh has a Pre-doctoral grant (FI 2017) provided by the Catalan government and an Erasmus+ travel grant by URV. Prof. Bustince acknowledges the support of Spanish project TIN2016-77356-P.

References

1. Barrenechea, E., Bustince, H., Fernandez, J., Paternain, D., Sanz, J.A.: Using the choquet integral in the fuzzy reasoning method of fuzzy rule-based classification systems. *Axioms* 2(2), 208–223 (2013)

2. Beliakov, G., Pradera, A., Calvo, T.: Aggregation functions: A guide for practitioners, vol. 221. Springer (2007)
3. Bodjanova, Slavka: Fuzzy Sets and Fuzzy Partitions, pp. 55–60. Springer Berlin Heidelberg, Berlin, Heidelberg (1993)
4. Grabisch, M., Labreuche, C.: A decade of application of the choquet and sugeno integrals in multi-criteria decision aid. *Annals of Operations Research* 175(1), 247–286 (2010)
5. Lucca, G., Sanz, J.A., Dimuro, G.P., Bedregal, B., Bustince, H., Mesiar, R.: Cf-integrals: A new family of pre-aggregation functions with application to fuzzy rule-based classification systems. *Information Sciences* 435, 94 – 110 (2018)
6. Lucca, G., Sanz, J.A., Dimuro, G.P., Bedregal, B., Asiain, M.J., Elkano, M., Bustince, H.: Cc-integrals: Choquet-like copula-based aggregation functions and its application in fuzzy rule-based classification systems. *Knowledge-Based Systems* 119, 32–43 (2017)
7. Mesiar, R., Stupanov, A.: A note on cc-integral. *Fuzzy Sets and Systems* (2018)
8. Romero-Aroca, P., de la Riva-Fernandez, S., Valls-Mateu, A., Sagarra-Alamo, R., Moreno-Ribas, A., Soler, N.: Changes observed in diabetic retinopathy: eight-year follow-up of a spanish population. *British Journal of Ophthalmology* 100(10), 1366–1371 (2016)
9. Romero-Aroca, P., de la Riva-Fernandez, S., Valls-Mateu, A., Sagarra-Alamo, R., Moreno-Ribas, A., Soler, N., Puig, D.: Cost of diabetic retinopathy and macular oedema in a population, an eight year follow up. *BMC ophthalmology* 16(1), 136 (2016)
10. Romero-Aroca, P., Valls, A., Moreno, A., Sagarra-Alamo, R., Basora-Gallisa, J., Saleh, E., Baget-Bernaldiz, M., Puig, D.: A clinical decision support system for diabetic retinopathy screening: Creating a clinical support application. *Telemedicine and e-Health*. In press (2018)
11. Saleh, E., Baszczyski, J., Moreno, A., Valls, A., Romero-Aroca, P., de la Riva-Fernandez, S., Sowiski, R.: Learning ensemble classifiers for diabetic retinopathy assessment. *Artificial Intelligence in Medicine* 85, 50 – 63 (2018)
12. Saleh, E., Moreno, A., Valls, A., Romero-Aroca, P., de la Riva-Fernandez, S.: A fuzzy random forest approach for the detection of diabetic retinopathy on electronic health record data. In: *Artificial Intelligence Research and Development*. vol. 288, p. 169. IOS Press (2016)
13. Saleh, E., Valls, A., Moreno, A., Romero-Aroca, P.: Integration of different fuzzy rule-induction methods to improve the classification of patients with diabetic retinopathy. In: *Recent Advances in Artificial Intelligence Research and Development*. pp. 6–15 (2017)
14. Saleh, E., Valls, A., Moreno, A., Romero-Aroca, P., de la Riva-Fernandez, S., Sagarra-Alamo, R.: Diabetic retinopathy risk estimation using fuzzy rules on electronic health record data. In: *Modeling Decisions for Artificial Intelligence*. pp. 263–274. Springer (2016)
15. Torra, V.: On hierarchically S-decomposable fuzzy measures. *International journal of intelligent systems* 14(9), 923–934 (1999)
16. Torra, V., Narukawa, Y.: *Modeling decisions: information fusion and aggregation operators*. Springer Science & Business Media (2007)
17. Yuan, Y., Shaw, M.J.: Induction of fuzzy decision trees. *Fuzzy Sets and Systems* 69(2), 125–139 (1995)