

# Article A Novel Classification Method Using the Takagi–Sugeno Model and a Type-2 Fuzzy Rule Induction Approach

Martin Tabakov \*<sup>D</sup>, Adrian B. Chlopowiec <sup>D</sup> and Adam R. Chlopowiec <sup>D</sup>

Department of Artificial Intelligence, Wroclaw University of Science and Technology, 50-370 Wroclaw, Poland; 254517@student.pwr.edu.pl (A.B.C.); 254518@student.pwr.edu.pl (A.R.C.)

\* Correspondence: martin.tabakow@pwr.edu.pl

**Abstract:** The main purpose of this research was to introduce a classification method, which combines a rule induction procedure with the Takagi–Sugeno inference model. This proposal is a continuation of our previous research, in which a classification process based on interval type-2 fuzzy rule induction was introduced. The research goal was to verify if the Mamdani fuzzy inference used in our previous research could be replaced with the first-order Takagi–Sugeno inference system. In the both cases to induce fuzzy rules, a new concept of a fuzzy information system was defined in order to deal with interval type-2 fuzzy sets. Additionally, the introduced rule induction assumes an optimization procedure concerning the footprint of uncertainty of the considered type-2 fuzzy sets. A key point in the concept proposed is the generalization of the fuzzy information systems' attribute information to handle uncertainty, which occurs in real data. For experimental purposes, the classification method was tested on different classification benchmark data and very promising results were achieved. For the data sets: Breast Cancer Data, Breast Cancer Wisconsin, Data Banknote Authentication, HTRU 2 and Ionosphere, the following F-scores were achieved, respectively: 97.6%, 96%, 100%, 87.8%, and 89.4%. The results proved the possibility of applying the Takagi–Sugeno model in the classification concept. The model parameters were optimized using an evolutionary strategy.

**Keywords:** fuzzy sets; interval type-2 fuzzy sets; Takagi–Sugeno model; fuzzy information systems; classification; fuzzification optimization

# 1. Introduction

Classification is one of the most important and challenging machine learning tasks. As the main goal is to investigate similarities between groups of objects, the classification accuracy strongly depends on the initial data sets used for training. There are a lot of issues to be handled, such as unbalanced or biased training data, but also incomplete or vague information which is often the case for real data. A lot of research was done to address this problem, especially using fuzzy classification and by extending to type-2 fuzzy sets as well. For example, recently in [1], a new fuzzy reasoning method for an interval type-2 fuzzy classification system including cluster-based rules was introduced. Authors propose to incorporate the introduced reasoning procedure with a new Possibilitybased fuzzy measure to handle uncertainty of cluster centers in an interval type-2 fuzzy rule-based classification system. In [2], a robust sparse representation for classification of medical images is proposed based on an introduced adaptive type-2 fuzzy dictionary learning. Remote-sensing image classification techniques using type-2 fuzzy sets were introduced [3–5] where, for example recently, considering the last cited research, a novel robust interval type-2 possibilistic fuzzy clustering model for the classification of complex remote sensing land cover was presented.

There are many fuzzy control applications based on the use of type-2 fuzzy sets. Recently, type-2 fuzzy logic system for ergonomic control of indoor environments was proposed [6]. The system calculates the effective working time and time-dependent change



Citation: Tabakov, M.; Chlopowiec, A.B.; Chlopowiec, A.R. A Novel Classification Method Using the Takagi–Sugeno Model and a Type-2 Fuzzy Rule Induction Approach. *Appl. Sci.* 2023, *13*, 5279. https:// doi.org/10.3390/app13095279

Academic Editor: Diana Kalibatiene

Received: 31 January 2023 Revised: 13 April 2023 Accepted: 18 April 2023 Published: 23 April 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). in carbon dioxide levels and aims to evaluate ergonomic comfort conditions. The authors proved better system performance using type-2 fuzzy sets for ergonomic control in fuzzy environments. Important medical application was proposed in [7]. An interval type-2 fuzzy stochastic modeling and control strategy to consider the uncertainties of the COVID-19 pandemic in order to control the number of infected people was introduced. A common strategy for fuzzy model parameter tuning is the use of evolutionary algorithms. In [8], authors applied a genetic algorithm to tune the parameters of an interval type-2 fuzzy proportional–derivative controller, in order to track the trajectory of a snake robot in the presence of system uncertainties. In [9,10], slime mold and particle swarm optimization algorithms were used for parameter tuning of interval type-2 fuzzy controllers, respectively. In [11], the design of membership functions for interval type-2 fuzzy tracking controllers was optimized using a metaheuristic algorithm.

In general, type-2 fuzzy sets provide better generalizations [12,13] as the main assumption here is the additional fuzzification of the membership values of type-1 fuzzy sets involved. Therefore, type-2 fuzzy sets can handle better with imprecise information [14,15].

Additionally, to deal with information issues, data discovery is used to induce knowledge. Information systems and rough sets, originally introduced by Pawlak [16–18], are often used to represent knowledge. Basic mathematical concepts, using rough sets theory to induce fuzzy rules, have been worked on for a long time. In [19], fuzzy learning methods for the automatic generation of membership functions and fuzzy if-then rules from training examples were introduced. Fuzzy decision rules induction was presented in [20] using the tolerance-based rough sets model. Rough sets were used to reduce the dimensionality of complex datasets as preprocessing for fuzzy rule induction [21].

The above guided us to investigate the possible combination of fuzzy reasoning with Pawlak's information systems. We have already introduced our concept of a fuzzy information system [22]. This research is a continuation of our previous research, by evolving and combining our concept with the Takagi–Sugeno fuzzy model [23]. Fuzzy information systems introduce a relation between attributes and fuzzy sets [24]. Applications of fuzzy information systems can be found in the fields of decision-making [25,26] and rule extraction [26,27]. Recently, new decision-making concepts were introduced in [28] by combining regret theory with three-way decisions in fuzzy incomplete information systems, or in [29], by multi-attribute predictive analysis using fuzzy rough sets.

In this research, a fuzzy information system to induce rules was applied with respect to classification attribute values. Next, a transformation into fuzzy rules was proposed, and finally, the fuzzy sets used were expanded into type-2 fuzzy sets. To provide classification, the Takagi–Sugeno model was applied in order to simplify our previous concept, which was related to the Mamdani fuzzy model [30]. The Takagi–Sugeno model does not require rule consequents defined as fuzzy sets, but introduces additional parameters. For that reason, in this research, an optimization procedure was applied to determine the best fuzzy classifier for the benchmark data considered. The research experiments were provided for binary classification problems defined with well-known benchmark data.

Therefore, our research goal was to provide a classification method, by replacing the Mamdani model used in our previous research with the first-order Takagi–Sugeno fuzzy inference. This simplified the induction of type-2 fuzzy rules in the classification process, as fuzzy sets were not required to be defined as rule consequents. The main research advantage of this work states the proposal of a fuzzy information system used to induce type-2 fuzzy rules, incorporated with the Takagi–Sugeno inference to provide classification.

The rest of the paper is organized as follows: in Section 2, the benchmark data used and some theoretical background are introduced; in Section 3, the introduced classification procedure is explained; in Section 4, the classification results are presented; and finally, Sections 5 and 6 draw corresponding discussion, possible further developments and conclusions.

### 2. Material and Methods

## 2.1. Materials

In this research, classification experiments were provided on well-known benchmark data [31] for binary classification. This is the same benchmark data as those used in our previous study [22], with details given in Table 1.

Table 1. The bench	nmark data use	d.
--------------------	----------------	----

Benchmark	Number of Attributes	Number of Samples	<b>Class Proportions</b>
Blood Transfusion	7	748	76%/24%
Breast Cancer Data	4	569	63%/37%
Breast Cancer Wisconsin	9	683	65%/35%
Data Banknote Authentication	4	1372	56%/44%
Haberman	3	305	73%/27%
Heart	13	303	46%/54%
HTRU 2	8	17,898	9%/91%
Immunotherapy	7	90	79%/21%
Indian Liver Patient	10	583	71%/29%
Ionosphere	34	351	64%/36%
Parkinson	22	187	74%/26%
Pima Indians Diabetes	8	768	65%/35%
Vertebral	6	310	68%/32%

# 2.2. Methods

2.2.1. Type-1 and Type-2 Fuzzy Sets

A Type-1 fuzzy set F consists of a non-empty domain X and a function  $\mu_F : X \to [0, 1]$ , called a membership function [32]. Considering a continuous membership function, a fuzzy set of type-1 as defined in Equation (1):

$$F =_{df} \int_{X} \mu_F(x) / x, x \in X$$
(1)

The above integral denotes the collection of all points  $x \in X$  with associated membership grade  $\mu_F(x) \in [0, 1]$ . The function values define the grade of membership of the elements of X to the fuzzy set F. Membership functions are assumed to describe imprecise or vague information.

The type-1 fuzzy sets concept was expanded by fuzzifying the membership function values themselves and type-2 fuzzy sets were introduced. A type-2 fuzzy set, denoted as  $\tilde{F}$ , is defined in Equation (2) [5]:

$$\widetilde{F} =_{df} \int_{x \in X} \int_{u \in J_x} \mu_{\widetilde{F}}(x, u) / (x, u), J_x \subseteq [0, 1]$$

$$(2)$$

where  $\iint$  denotes the union over all admissible x and u. The most widely used special cases of type-2 fuzzy sets, mostly because of their easy interpretation, are the interval type-2

(IT2) fuzzy sets [5]. Uncertainty about F is handled by the so-called footprint of uncertainty (FOU) of  $\tilde{F}$ , as shown in Equation (3):

$$FOU(F) =_{df} \bigcup_{x \in X} J_x, J_x \subseteq [0, 1]$$
(3)

The area of a FOU is directly related to the uncertainty that is conveyed by an interval type-2 fuzzy set, and what follows, a FOU with more area is more uncertain than the one with less area. The lower membership function (LMF) and upper membership function

(UMF) of F are two type-1 membership functions <u>F</u> and F that bound the FOU, which are used to describe  $J_x$ , see Equation (4):

$$J_{x} =_{df} \left[ \mu_{\underline{F}}(x), \mu_{\overline{F}}(x) \right]$$
(4)

which leads to Equation (5):

$$FOU(\tilde{F}) =_{df} \bigcup_{x \in X} \left[ \mu_{\underline{F}}(x), \mu_{\overline{F}}(x) \right]$$
(5)

Figure 1 illustrates the graphical interpretation of the above definitions.



**Figure 1.** An interval type-2 fuzzy set F determined by membership functions of the type-1 fuzzy sets: F and  $\overline{F}$ .

2.2.2. Takagi-Sugeno Type-2 Fuzzy Inference

The major difference between the type-1 and type-2 Takagi–Sugeno fuzzy models obviously lies in the incorporation of type-2 fuzzy sets in the inference process. This requires corresponding fuzzy rule interpretation and type–reducer mechanism.

A type-1 fuzzy rule in terms of the Takagi–Sugeno model, takes the following form:

$$R :_{df} IF(x_1 \text{ is } X_1) \circ \ldots \circ (x_i \text{ is } X_i) THEN f(x_1, \ldots x_i)$$

where  $X_i$  (i = 1, ..., I) are type-1 fuzzy sets defined over corresponding domains. The operator 'o' is assumed as a  $\otimes$ ,  $\oplus$ :  $[0, 1]^2 \rightarrow [0, 1]$  are the t- and s-norms [33], respectively. These binary operators are applied in fuzzy logic as generalizations of the conjunction and disjunction Boolean logic operators. The Zadeh's t- and s- norms defined by the min and max operators, were applied, respectively. The rule consequents are multivariable

functions, defined as combinations of polynomials. The most popular solution, i.e., firstorder Takagi–Sugeno model, assumes combinations of linear functions. Therefore, any rule consequent takes the form:

Consequent:  $f(x_1, ..., x_i) =_{df} \sum_{k=1}^{i} (a_k x_k + b_k) = \sum_{k=1}^{i} a_k x_k + \sum_{k=1}^{i} b_k$ , where  $a_k$  and  $b_k$  are the parameters of the corresponding linear functions, defined with respect to each  $x_i$ .

It is interesting to note that in the Takagi–Sugeno model, there is no need to define the consequents as fuzzy sets and therefore, there is no need to define any membership function in the consequents as it should be done using, for example, the Mamdani model [30].

The final output of the system is the weighted over all rule outputs, as defined in Equation (6):

Final Output = 
$$_{df} \frac{\sum_{k=1}^{N} w_k f_k}{\sum_{k=1}^{N} w_k}$$
, (6)

where N is the number of rules,  $w_k$  is the rule-firing value derived from the k<sup>th</sup> rule antecedent and  $f_k$  is the value of the kth rule consequent.

Considering type-2 fuzzy sets, for an input vector  $\overline{x} = \{x'_1, x'_2, ..., x'_i\}$ , typical computations of an IT2 fuzzy system consist of the following steps:

- (1) Compute the membership intervals of  $x'_i$  for each  $\tilde{X}^n_i$ ,  $\left\lfloor \mu_{\underline{X}^n_i}(x'_i), \mu_{\overline{X}^n_i}(x'_i) \right\rfloor$ , i = 1, ..., I; n = 1, ..., N (N is the number of rules),
- (2) Compute the firing interval of the  $n^{th}$  rule:

$$F^{n}\left(\bar{x}\right) =_{df} \left[ \mu_{\underline{X}_{1}^{n}}(x_{1}') \circ \dots \circ \mu_{\underline{X}_{i}^{n}}\left(x_{i}'\right), \mu_{\underline{X}_{1}^{-n}}(x_{1}') \circ \dots \circ \mu_{\underline{X}_{i}^{-n}}(x_{i}') \right]$$

(3) Use type reduction to combine  $F^n(\bar{x})$  with corresponding rule consequents.

The most popular one is the center-of-set (COS) type reducer [34] using the Karnik–Mendel algorithm [34,35] or their modifications [36,37].

In order to apply the Karnik–Mendel algorithm, a single type-2 fuzzy set must be derived using the rules outputs. For a Takagi–Sugeno type-2 fuzzy system, the aggregate set is generated using the following procedure:

- 1. Sort the rule outputs from all rules into ascending order,
- 2. For each output, define the UMF value using the maximum firing range of the considered rule,
- 3. For each output, define the LMF value using the minimum firing range of the considered rule.

As an example, consider the data shown in Table 2. Let us assume five rules with output values and firing range limits. For these data, the aggregate type-2 fuzzy set shown in Figure 2 will be generated.

Rule	Rule Output Value (f <sub>k</sub> )	Minimum Firing Value	Maximum Firing Value
1	6.8	0.1	0.6
2	1.6	0.3	0.7
3	5.3	0.2	0.5
4	0.4	0.4	0.8
5	3.1	0.2	0.9
Sorted in ascending order:			
4	0.4	0.4	0.8
2	1.6	0.3	0.7

Table 2. Example aggregate set generation.

Rule	Rule Output Value (f <sub>k</sub> )	Minimum Firing Value	Maximum Firing Value
5	3.1	0.2	0.9
3	5.3	0.2	0.5
1	6.8	0.1	0.6

Table 2. Cont.



**Figure 2.** Exemplary output type-2 fuzzy set for an input vector. To make it easier to understand the structure of the aggregate set, the values for Rule 5 are marked.

Defining the above aggregate type-2 fuzzy output set, the Karnik–Mendel algorithm [34,35] can be applied in order to derive the final output system value. As the complete Takagi–Sugeno type-2 fuzzy inference procedure is well-described in the literature, we will omit further details.

# 3. Fuzzy Information System with Tagaki-Sugeno Reasoning

## 3.1. Type-1 Fuzzy Information System and Fuzzy Rule Induction

This research is based on our interpretation of a fuzzy information system, introduced in [22]. We assume that the values of the system information function are linguistic variables which corresponds to fuzzy sets. This is a granulation information proposal. A classic information system [16] is defined by the elements: (U, A, V, f), where U is a universe, A is a set of attributes, V represents attributes domains:  $V = {}_{df} \bigcup_a V_a$ , with nonempty domain  $V_a$  of the a-th attribute (a  $\in$  A), and f is the information function f: U×A  $\rightarrow$  V,  $\forall x \in$  U, a  $\in$ A f(u, a)  $\in V_a$ . The indiscernibility binary relation (IND), defined over U (IND  $\subseteq$  U<sup>2</sup>), plays in theory an important role. It is an equivalence relation, defined in Equation (7):

$$IND(B) = _{df} \{ (u_i, u_j) \in U ' U: \forall a \in B f(u_i, a) = f(u_j, a) \}, u_i, u_j \in U, B \subseteq A$$
(7)

The above binary relation is used to define the lower and upper approximations of any subset of U. Any such a pair is defined as a rough set [17,18].

In accordance with our interpretation of a fuzzy information system over any attribute domain V<sub>a</sub>, three fuzzy subsets are defined: {low, medium, high}. The information function values are replaced with corresponding fuzzy sets, with respect to the maximum membership value: max{ $\mu_{low}$ (object, attribute),  $\mu_{medium}$ (f(object, attribute),  $\mu_{high}$ (f(object,

attribute)}. The membership function of the medium fuzzy set assumes Gaussian distribution of the attribute data for any  $a \in A$ .

Therefore, the membership functions for the considered fuzzy sets can be easily defined, see Equation (8):

$$\mu_{\text{medium}}(a) =_{df} e^{\frac{-(a-a_0)^2}{2\sigma^2}}, \ \mu_{\text{low}}(a) =_{df} \left\{ \begin{array}{c} 1 - e^{\frac{-(a-a_0)^2}{2\sigma^2}} : a < a_0 \\ 0 : a \ge a_0 \end{array} \right\}, \\ \mu_{\text{high}}(a) =_{df} \left\{ \begin{array}{c} 0 : a < a_0 \\ 1 - e^{\frac{-(a-a_0)^2}{2\sigma^2}} : a \ge a_0 \end{array} \right\}, \ a \in V_a \text{with expected value}(a_0) \\ \text{and standard deviation } (\delta) \end{array}$$
(8)

The above attribute fuzzification introduces a very simple initialization of the information function as well as values' generalization. For example, let us consider the below matrix (Table 3) which represents a fuzzy information system: U = {object<sub>1</sub>, object<sub>2</sub>, object<sub>3</sub>, object<sub>4</sub>}, A = {attribute<sub>1</sub>, attribute<sub>2</sub>, attribute<sub>3</sub>}, V = V<sub>attribute<sub>1</sub></sub>  $\cup$  V<sub>attribute<sub>2</sub></sub>  $\cup$  V<sub>attribute<sub>3</sub></sub> and we assume for each set V<sub>attribute<sub>i</sub></sub> a Gaussian distribution in order to obtain the fuzzy sets {low<sub>i</sub>, medium<sub>i</sub>, high<sub>i</sub>}.

Table 3. Fuzzy information system example.

	$attribute_1$	attribute <sub>2</sub>	attribute <sub>3</sub>
object <sub>1</sub>	low	medium	low
object <sub>2</sub>	medium	high	high
object <sub>3</sub>	medium	high	high
object <sub>4</sub>	low	medium	low

So, for example,  $f(object_2, attribute_2)$  is introduced as *high* because the following inequality is satisfied:  $\mu_{high}(f(object_2, attribute_2)) \ge \max\{ \mu_{low}(f(object_2, attribute_2), \mu_{medium}(f(object_2, attribute_2)\}$  or  $f(object_1, attribute_3)$  is introduced as *low*, as  $\mu_{low}(f(object_1, attribute_3)) \ge \max\{ \mu_{medium}(f(object_1, attribute_3), \mu_{high}(f(object_1, attribute_3))\}$ .

The corresponding partition (P) with respect to IND has two equivalence classes:

*P*/IND{attribute1, attribute2, attribute3} = {{object1, object4}, {object2, object3}}

Next, if a decision attribute (A\*) is added to the set of attributes:  $A = A \cup A^*$  an information system can be represented by a decision table. Assuming such an extension, the Skowron and Suraj rule induction method can be applied [38,39] and next, transformation of the rules induced for each decision into fuzzy rules is introduced.

For example, for pairs: (object<sub>1</sub>, attribute<sub>1</sub>): low, (object<sub>2</sub>, attribute<sub>2</sub>): high, (object<sub>3</sub>, attribute<sub>1</sub>): medium, and a rule for decision D is induced as follows: decision D: (f(object<sub>1</sub>, attribute<sub>1</sub>)  $\land$  f(object<sub>2</sub>, attribute<sub>2</sub>))  $\lor$  f(object<sub>3</sub>, attribute<sub>1</sub>), the above rule can be transformed into the following fuzzy rule: If ((f(object<sub>1</sub>, attribute<sub>1</sub>) is low)  $\otimes$  (f(object<sub>2</sub>, attribute<sub>2</sub>) is high))  $\oplus$  (f(object<sub>3</sub>,

attribute<sub>1</sub>) is medium) Then D.

An explanation of the rule induction procedure and the generation of fuzzy rules directly from a decision table can be found in our previous research in more detail ([22], Sections 2.2.3 and 2.2.4).

The most important issues here are: (1) the possibility of attribute value generalization by applying fuzzy sets directly as values in the decision table. Therefore, the information function values are related directly to membership functions of fuzzy sets. (2) The rule induction procedure and the transformation of the rule induced into type-1 fuzzy rules.

## 3.2. Involving Type-2 Fuzzy Sets

The idea of applying type-2 fuzzy sets is very simple and related to our previous research [22]. It is enough to change the values of the standard deviation of the Gaussian-type membership function of the fuzzy set medium to define the bounds of the FOU for the type-2 fuzzy set medium. In such a way, the assumed fuzzy sets {low, medium, high} can be replaced with the {low, medium, high} type-2 fuzzy sets. Figure 3 clarifies the issue. This transformation is done after rule induction and transformation into type-1 fuzzy rules.



Figure 3. The type-1 and type-2 fuzzification process.

Nevertheless, the assumption of Gaussian distribution for each attribute value is hard to be fulfilled. That is the reason why we assumed the experiment with the fuzzy rule premises in terms of standard derivation values (defining the corresponding FOU) parameter named as *sigma\_offset* and different configurations of membership functions. Meaning, we extend the assumption of using only three type-2 fuzzy sets {low, medium, high}. Without changing our previous research, the configurations assumed are shown in Table 4 below. This concerns the number of Gaussians in the rule premises, the applied *sigma\_offset*, and the expected value for the medium membership function. For more details, see ([22], Section 2.2.6) concerning the fuzzy sets defined in the fuzzy rule premises.

**Table 4.** The IT2 fuzzy sets used (presentation of the concept with respect to a chosen attribute of the blood transfusion dataset, assuming normalization as well).



9 of 17





#### 3.3. Takagi–Sugeno Reasoning with Optimization

Consequently, for each rule premises, configuration of different *sigma\_offset* are considered. Additionally, in terms of first-order Takagi–Sugeno systems, two parameters defining the corresponding linear function for each fuzzy set in the rule premises should be calculated. The assumed varieties of type-2 fuzzy sets along with the *sigma\_offset* value and the linear function parameters define an optimization problem. A finite set of membership functions is considered, generated with respect to the *medium* membership function. The assumed numbers of membership functions are: {3, 5, 7, 9 11}. Considering the membership function configurations and the *sigma\_offset*, the grid search algorithm in a k-fold cross-validation manner for these hyperparameters optimization was applied. The linear function Evolutionary Strategy (CMA-ES) [40], i.e., as each cross-validation step defines a new training set, the induced rules also differ. Therefore, at each cross-validation step, we applied CMA-ES for linear function parameters optimization.

Additionally, in the conducted experiments as possible data pre-processing, a dimension reduction procedure using the well-known principal component analysis method (PCA) and the Random Oversampling (ROS) [41–43] to handle with class imbalance were applied. The oversampling was important in terms of rule induction, as the inconsistency elimination step ([22], Section 2.2.3, algorithm 1) strongly favors classes with large presence of data.

The complete set of hyperparameters to be optimized is given in Table 5. On the other hand, Table 6 presents the hyperparameter values used in the CMA-ES procedure itself. The number of folds for each dataset are given in Table 7.

Hyperparameter	Value
Number of Gaussian functions	3, 5, 7, 9, 11
Whether the std applied to Gaussians are the same	Yes/No
If the mean value of the medium membership function is derived directly from mean value of the corresponding feature	Yes/No
Sigma_offset	[0.5, 0.9] with step of 0.05
Use PCA	Yes/No
Use ROS	Yes/No

Table 5. Tuned hyperparameters of sets of membership functions.

Table 6. CMA-ES hyperparameters.

Hyperparameter	Value
Initialization	Random
Maximum evaluations of fitness function	20,000
Number of restarts	1
Population size increase ratio after restart	2
Standard deviation in each coordinate	0.7
Population size	20
Stagnation tolerance	100 evaluations
Parameters values range	[-400, 400]

Benchmark	k-Value
Blood Transfusion	10
Breast Cancer Data	10
Breast Cancer Wisconsin	10
Data Banknote Authentication	10
Haberman	6
Heart	5
HTRU	10
Immunotherapy	4
Indian Liver Patient	10
Ionosphere	6
Parkinson	4
Pima Indians Diabetes	10
Vertebral	5

Table 7. Number of folds applied in the cross-validation process.

The experimental procedure pipeline in this research was defined as follows:

- 1. Split a considered dataset into k-folds.
- 2. For each set of hyperparameters:
  - a. For each cross-validation step:
    - i. Set one fold as held-out for validation, use the rest for training.
    - ii. Induce the knowledge base using the training set.
    - iii. Infer the crisp value for each sample in the training set with the Takagi–Sugeno model.
    - iv. Classify the sample with the threshold function set at 0. Treat the sample as negative if the crisp value is lower than 0.
    - v. Calculate the F1 metric.
    - vi. Optimize the parameters of linear functions with CMA-ES, defining the F1 metric as the fitness function.
    - vii. Evaluate the model on the validation set when the fitness function is optimized or the maximum number of evaluations passes. Otherwise return to (iii).
  - b. Choose a set of hyperparameters which maximize the mean value of the F1 metric over the validation sets.

In order to evaluate the model for each dataset, we conduct the following procedure:

- (1) Choose the best hyperparameters set for a dataset.
- (2) Optimize the fitness function on the training set.
- (3) Choose the best parameters of linear functions.
- (4) Evaluate the model on a test set.

Repeat the procedure n times (n = 10) to minimize CMA-ES possible local extrema solutions and choose the best result.

#### 4. Binary Classification Results

In Table 8 the best binary classification results achieved for the best hyperparameters setup are presented.

Dataset	F1 Score (%)	Accuracy (%)	Sensitivity (%)	Hyperparameters	ROS	PCA
Blood Transfusion	56.6	69.3	83.3	<11 Gaussian, equal, mean, 0.75>	Yes	No
Breast Cancer Data	97.6	98.2	95.2	<3 Gaussian, equal, center, 0.5>	Yes	Yes
Breast Cancer Wisconsin	96.0	97.1	100.0	<3 Gaussian, progressive, mean, 0.85>	Yes	Yes
Data Banknote Authentication	100.0	100.0	100.0	<3 Gaussian, equal, mean, 0.65>	Yes	No
Haberman	40.0	66.1	43.8	<9 Gaussian, equal, mean, 0.55>	No	No
Heart	85.2	85.2	92.9	<3, progressive, mean, 0.6>	Yes	Yes
HTRU 2	87.8	97.8	85.4	<9 Gaussian, equal, center, 0.5>	No	Yes
Immunotherapy	66.7	83.3	75.0	<9 Gaussian, equal, mean, 0.75>	Yes	No
Indian Liver Patient	57.4	58.1	97.1	<11, Gaussian, equal, center, 0.8>	Yes	Yes
Ionosphere	89.4	93.0	84.0	<7 Gaussian, equal, mean, 0.55>	No	Yes
Parkinson	80.0	89.7	80.0	<5 Gaussian, progressive, mean, 0.8>	Yes	No
Pima Indians Diabetes	66.2	68.8	87.0	<11 Gaussian, equal, center, 0.9>	Yes	Yes
Vertebral	87.1	82.3	88.1	<7 Gaussian, progressive, mean, 0.75>	Yes	No

Table 8. Classification results.

In Table 9, we present the best classification results achieved in comparison with our previous research [22] and other classifiers as well.

Dataset	The Presented Approach (Using the Takagi–Sugeno Model) (%)	Our Previous Approach (Using the Mamdani Model) (%)	Other Classifiers (%)
Breast Cancer Data	97.6	91.2	Immune-inspired semi-supervised Algorithm, introduced in [44]: 97.3
Breast Cancer Wisconsin	96.0	95.7	Extreme Learning Machine Neural Networks, introduced in [45]: 97.8 Immune-inspired semi-supervised algorithm, introduced in [44]: 96.5 Support vector machines combined with Feature Selection, introduced in [46]: 99.7
Data Banknote Authentication	100.0	99.3	Deep Neural Network with PCA and LDA, introduced in [47]: 99 Decision tree approach, introduced in [48]: 99.4 Random Forest approach, introduced in [49]: 94.8 Neural Network-Genetic Algorithm, introduced in [50]: 100
HTRU 2	87.8	89.0	Classical classifiers: C4.5: 74; MLP: 75.2; NB: 69.2; SVM: 78.9 GH-VFDT Algorithm, introduced in [51]: 86.2 A hybrid ensemble method, introduced in [52]: 91.8 (with respect to a voting threshold parameter)
Ionosphere	89.4	88.8	Clustered Bayesian classification, 88.5 [53]

Table 9. Comparison with other methods, concerning the F1 score measure.

## 5. Discussion

This research is a continuation of a previous one, published in [22]. Both papers present our newly introduced definition of a fuzzy information system, which has the following advantages:

- The information function values are interpreted as fuzzy sets, labeled with corresponding linguistic variables. This gives the possibility to generalize information—we do not consider numerical values for pairs (object, attribute), but general descriptions such as 'small', 'medium', and 'high'.
- The decision table used is generated in an automatic manner for a considered data set, as the value 'medium' is assumed as the Gaussian distribution of the data for each attribute. Then, the sets 'low' and 'high' are easy to be defined using the 'medium' membership function. Next, a corresponding label (identifying the corresponding fuzzy set) is given for a pair (object, attribute), by using the maximum membership value.

Additionally, we propose an easy transformation in order to change the applied type-1 fuzzy sets into type-2 fuzzy sets.

Therefore, by defining such a decision table, the rule induction procedure described in [38,39] is applied and next, type-2 fuzzy rules are derived. What is more, different fuzzification strategies are possible (see Table 4).

In this research, the goal was to incorporate our fuzzy information system concept with the Takagi–Sugeno model, as it does not require the model outputs to be defined as fuzzy sets. Such an achievement would create new possibilities, using the advantages of the Takagi–Sugeno model. To incorporate the model, we had to optimize the model parameters with respect to the corresponding benchmark data. We have applied the CMA-ES (Covariance Matrix Adaptation Evolutionary Strategy) for this purpose. We conducted experiments with the same binary classification problems and the same fuzzification experiments, as presented in [22], in order to compare the results.

While performing experiments, we have discovered further promising features and potential disadvantages of the method proposed. The classification process is able to fit training data, but sometimes has issues with generalization. Applying PCA allows us to minimize this issue, making the method proposed more robust. Additionally, performing optimization with an evolutionary strategy allows the model to find very well-fitting decision boundaries. Therefore, one might achieve even better performance by applying additional regularization methods.

To solve the problem of imbalanced data, Random Oversampling was used, which was not beneficial in a few cases. If the imbalance is very high, ROS may produce too much redundancy; therefore, the information system may produce biased rules, ignoring important information contained in the data. In such a scenario, other sampling techniques should be applied instead.

Regarding the attribute fuzzification procedure, only Gaussians were assumed as membership functions. We suppose, there is room for improvement, adjusting membership functions to the attributes' characteristics. Additionally, not only the first-order Takagi– Sugeno model could be applied.

Despite such constraints, good classification results for several datasets were achieved. We believe that the application of the Takagi–Sugeno model in our approach will support the solving of multi-class classification problems. This is because, in the model used, there is no need to define decision classes, contrary to the Mamdani fuzzy model.

Summarizing, a new classification concept was proposed in this research with the following main advantages:

- Defining a decision table with fuzzy values. The fuzzification is provided in an automatic manner directly from a data set.
- Using the rule induction method based on the information system concept, which has a solid mathematical background. Each rule is related to a corresponding class, regarding the classification problem considered.
- Transformation of the induced rules into type-2 fuzzy rules.
- Application of the Takagi–Sugeno model in the classification process. Therefore, there is no need to define the fuzzy rule consequents as fuzzy sets.

#### 6. Conclusions

In this research, a novel classification method, which incorporates our previously introduced concept of a fuzzy information system with the first-order Takagi–Sugeno model, was introduced. A fuzzification procedure with membership function optimization and a corresponding rule induction of type-2 fuzzy rules were proposed as well. The research was a continuation of our previous approach. Here, the aim was to replace the fuzzy model previously applied with the Takagi–Sugeno model. The model parameters optimization was conducted with the CMA-ES evolutionary strategy with respect to the classification problem considered. The achieved classification results proved that the Takagi–Sugeno-type inference is fully implementable with our fuzzy type-2 rule induction concept. Therefore, there is no need to define the fuzzy rule consequents as fuzzy sets, which simplify the whole classification method and potentially support the solving of multi-class classification problems. Nevertheless, some improvements could be applied, such as better feature selection, detection of outliers, or data augmentations in order to increase the classification results. Limitations of the method proposed might occur if the

generalizations applied in the decision table are too high. This means that if there is a low number of fuzzy sets and attributes, the rules induced will not differentiate the decision classes appropriately. Additionally, categorical variables are not appropriate due to the required data fuzzification.

As further research, we consider the possibility to use representation learning in order to generate attribute values for a considered problem and to fuzzify them. Therefore, we intend to involve the deep learning phase in the method proposed. We consider the concept proposed as a general approach, which is applicable to real-world classification problems, meant for data containing vague and incomplete information.

Author Contributions: Conceptualization, M.T.; Methodology, M.T.; Software, A.B.C. and A.R.C.; Validation, A.B.C. and A.R.C.; Formal analysis, M.T.; Investigation, M.T., A.B.C. and A.R.C.; Data curation, A.B.C. and A.R.C.; Writing—original draft, M.T.; Writing—review & editing, A.B.C. and A.R.C.; Supervision, M.T. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the statutory funds of the Department of Artificial Intelligence, Wroclaw University of Science and Technology.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

**Data Availability Statement:** The presented experiments were conducted on publicly available benchmark data.

**Acknowledgments:** All calculations were performed using computational resources provided by WCSS (Wroclaw Centre for Networking and Supercomputing). Additionally, we wish to acknowledge the help provided by Konrad Karanowski, Aleksy Walczak, Mateusz Grzesiuk, Jan Wielgus, Artur Zawisza and Piotr Majchrowski, for collaboration in implementation of software library dedicated to fuzzy techniques, which we have applied in our experiments.

Conflicts of Interest: The authors declare no conflict of interest.

#### References

- 1. Rostam Niakan Kalhori, M.; FazelZarandi, M.H. A new interval type-2 fuzzy reasoning method for classification systems based on normal forms of a possibility-based fuzzy measure. *Inf. Sci.* **2021**, *581*, 567–586. [CrossRef]
- Ghasemi, M.; Kelarestaghi, M.; Eshghi, F.; Sharifi, A. T2-FDL: A robust sparse representation method using adaptive type-2 fuzzy dictionary learning for medical image classification. *Expert Syst. Appl.* 2020, 158, 113500. [CrossRef]
- 3. Xing, H.; He, H.; Hu, D.; Jiang, T.; Yu, X. An interval Type-2 fuzzy sets generation method for remote sensing imagery classification. *Comput. Geosci.* 2019, 133, 104287. [CrossRef]
- Xu, J.; Feng, G.; Zhao, T.; Sun, X.; Zhu, M. Remote sensing image classification based on semi-supervised adaptive interval type-2 fuzzy c-means algorithm. *Comput. Geosci.* 2019, 131, 132–143. [CrossRef]
- 5. Wu, C.; Guo, X. Adaptive enhanced interval type-2 possibilistic fuzzy local information clustering with dual-distance for land cover classification. *Eng. Appl. Artif. Intell.* **2023**, *119*, 105806. [CrossRef]
- Erozan, İ.; Özel, E.; Erozan, D. A two-stage system proposal based on a type-2 fuzzy logic system for ergonomic control of classrooms and offices. *Eng. Appl. Artif. Intell.* 2023, 120, 105854. [CrossRef]
- Rafiei, H.; Salehi, A.; Baghbani, F.; Parsa, P.; Akbarzadeh-T, M.-R. Interval type-2 Fuzzy control and stochastic modeling of COVID-19 spread based on vaccination and social distancing rates. *Comput. Methods Programs Biomed.* 2023, 232, 107443. [CrossRef]
- 8. Bhandari, G.; Raj, R.; Pathak, P.M.; Yang, J.-M. Robust control of a planar snake robot based on interval type-2 Takagi–Sugeno fuzzy control using genetic algorithm. *Eng. Appl. Artif. Intell.* **2022**, *116*, 105437. [CrossRef]
- Precup, R.E.; David, R.C.; Roman, R.C.; Szedlak-Stinean, A.I.; Petriu, E.M. Optimal tuning of interval type-2 fuzzy controllers for nonlinear servo systems using Slime Mould Algorithm. *Int. J. Syst. Sci.* 2021, 1–16. [CrossRef]
- 10. Pozna, C.; Precup, R.E.; Horvath, E.; Petriu, E.M. Hybrid Particle filter-particle swarm optimization algorithm and application to fuzzy controlled servo systems. *IEEE Trans. Fuzzy Syst.* **2022**, *30*, 4286–4297. [CrossRef]
- 11. Cuevas, F.; Castillo, O.; Cortes, P. Optimal Setting of Membership Functions for Interval Type-2 Fuzzy Tracking Controllers Using a Shark Smell Metaheuristic Algorithm. *Int. J. Fuzzy Syst.* 2022, 24, 799–822. [CrossRef]
- 12. Mendel, J.M.; John, R.I. Type-2 fuzzy sets made simple. *IEEE Trans. Fuzzy Syst.* 2002, 10, 117–127. [CrossRef]
- 13. Mendel, J.M. Type-2 Fuzzy Sets and Systems: How to Learn about Them. IEEE Smc Enewsletter 2009, 27, 1–7.
- 14. Hagras, H. Type-2 FLCs: A new generation of fuzzy controllers. IEEE Comput. Intell. Mag. 2007, 2, 30–43. [CrossRef]
- 15. Castillo, O.; Melin, P. Type-2 Fuzzy Logic Theory and Applications; Springer: Berlin, Germany, 2008.

- 16. Pawlak, Z. Information systems-Theoretical foundations. Inf. Syst. 1981, 6, 205-218. [CrossRef]
- 17. Pawlak, Z. *Rough Sets*; Basic notions, Report no 431; Institute of Computer Science, Polish Academy of Sciences: Warsaw, Poland, 1981.
- 18. Pawlak, Z. Rough Sets: Theoretical Aspects of Reasoning about Data, System Theory, Knowledge Engineering and Problem Solving; Kluwer Academic Publishers: Dordrecht, The Netherlands, 1991; Volume 9.
- Hong, T.-P.; Chen, J.-B. Building a concise decision table for fuzzy rule induction. In Proceedings of the 1998 IEEE International Conference on Fuzzy Systems Proceedings. IEEE World Congress on Computational Intelligence (Cat. No.98CH36228), Anchorage, AK, USA, 4–9 May 1998; Volume 2, pp. 997–1002.
- Drwal, G.; Sikora, M. Induction of fuzzy decision rules based upon rough sets theory. In Proceedings of the 2004 IEEE International Conference on Fuzzy Systems (IEEE Cat. No.04CH37542), Budapest, Hungary, 25–29 July 2004; Volume 3, pp. 1391–1395.
- Shen, Q.; Chouchoulas, A. Data-driven fuzzy rule induction and its application to systems monitoring. In Proceedings of the FUZZ-IEEE'99. 1999 IEEE International Fuzzy Systems. Conference Proceedings (Cat. No.99CH36315), Seoul, Republic of Korea, 22–25 August 1999; Volume 2, pp. 928–933.
- 22. Tabakov, M.; Chlopowiec, A.; Chlopowiec, A.; Dlubak, A. Classification with Fuzzification Optimization Combining Fuzzy Information Systems and Type-2 Fuzzy Inference. *Appl. Sci.* **2021**, *11*, 3484. [CrossRef]
- Takagi, T.; Sugeno, M. Fuzzy identification of systems and its applications to modeling and control. *IEEE Trans. Syst. Man Cybern.* 1985, 15, 116–132. [CrossRef]
- 24. Wang, C.; Chen, D.; Hu, Q. Fuzzy information systems and their homomorphisms. Fuzzy Sets Syst. 2014, 249, 128–138. [CrossRef]
- 25. Dubois, D. The role of fuzzy sets in decision sciences: Old techniques and new directions. *Fuzzy Sets Syst.* **2011**, *184*, 3–28. [CrossRef]
- Lee, M.-C.; Chang, T. Rule Extraction Based on Rough Fuzzy Sets in Fuzzy Information Systems. *Trans. Comput. Collect. Intell. III* 2011, 6560, 115–127.
- 27. Cheruku, R.; Edla, D.R.; Kuppili, V.; Dharavath, R. RST-Bat-Miner: A fuzzy rule miner integrating rough set feature selection and bat optimization for detection of diabetes disease. *Appl. Soft. Comput.* **2018**, *67*, 764–780. [CrossRef]
- 28. Wang, W.; Zhan, J.; Zhang, C.; Herrera-Viedma, E.; Kou, G. A regret-theory-based three-way decision method with a priori probability tolerance dominance relation in fuzzy incomplete information systems. *Inf. Fusion* **2023**, *89*, 382–396. [CrossRef]
- 29. Kang, Y.; Yu, B.; Cai, M. Multi-attribute predictive analysis based on attribute-oriented fuzzy rough sets in fuzzy information systems. *Inf. Sci.* 2022, *608*, 931–949. [CrossRef]
- Mamdani, E.H.; Assilian, S. An Experiment in Linguistic Synthesis with a Fuzzy Logic Controller. Int. J. Man-Mach. Stud. 1975, 7, 1–13. [CrossRef]
- 31. Dua, D.; Graff, C. *UCI Machine Learning Repository*; University of California, School of Information and Computer Science: Irvine, CA, USA, 2019; Available online: http://archive.ics.uci.edu/ml (accessed on 31 December 2020).
- 32. Zadeh, L. Fuzzy sets. Inf. Control. 1965, 8, 338–353. [CrossRef]
- 33. Bronstein, I.N.; Semendjajew, K.A.; Musiol, G.; Mühlig, H. *Taschenbuch der Mathematik*; Verlag Harri Deutsch: Frankfurt, Germany, 2001; p. 1258.
- Mendel, J.M. Uncertain Rule-Based Fuzzy Logic Systems: Introduction and New Directions; Prentice-Hall: Upper Saddle River, NJ, USA, 2001.
- 35. Karnik, N.N.; Mendel, J.M.; Liang, Q. Type-2 fuzzy logic systems. IEEE Trans. Fuzzy Syst. 1999, 7, 643–658. [CrossRef]
- 36. Wu, D.; Mendel, J.M. Enhanced Karnik–Mendel algorithms. *IEEE Trans. Fuzzy Syst.* 2009, 17, 923–934.
- Wu, D.; Nie, M. Comparison and practical implementation of type reduction algorithms for type-2 fuzzy sets and systems. In Proceedings of the 2011 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2011), Taipei, Taiwan, 27–30 June 2011; pp. 2131–2138.
- Skowron, A.; Suraj, Z. A Rough Set Approach to Real-Time State Identification for Decision Making; Institute of Computer Science Report 18/93; Warsaw University of Technology: Warsaw, Poland, 1993; p. 27.
- 39. Skowron, A.; Suraj, Z. A Rough Set Approach to Real-Time State Identification. Bull. Eur. Assoc. Comput. Sci. 1993, 50, 264.
- 40. Hansen, N.; Ostermeier, A. Completely derandomized self-adaptation in evolution strategies. *Evol. Comput.* **2001**, *9*, 159–195. [CrossRef]
- 41. Hulse, J.V.; Khoshgoftaar, T.; Napolitano, A. Experimental perspectives on learning from imbalanced data. In Proceedings of the Twenty-Fourth International Conference on Machine Learning (ICML 2007), Corvallis, OR, USA, 20–24 June 2007.
- 42. Shelke, M.; Deshmukh, D.P.R.; Shandilya, V. A review on imbalanced data handling using undersampling and oversampling technique. *Comput. Sci.* 2017, *3*, 444–449.
- Mohammed, R.; Rawashdeh, J.; Abdullah, M. Machine learning with oversampling and undersampling techniques: Overview study and experimental results. In Proceedings of the 2020 11th International Conference on Information and Communication Systems (ICICS), Irbid, Jordan, 7–9 April 2020.
- Peng, L.; Chen, W.; Zhou, W.; Li, F.; Yang, J.; Zhang, J. An immune-inspired semi-supervised algorithm for breast cancer diagnosis. Comput. Methods Programs Biomed. 2016, 134, 259–265. [CrossRef]
- 45. Utomo, C.P.; Kardiana, A.; Yuliwulandari, R. Breast Cancer Diagnosis using Artificial Neural Networks with Extreme Learning Techniques. *Int. J. Adv. Res. Artif. Intell. Ijarai* 2014, *3*, 10–14.

- 46. Akay, A.F. Support vector machines combined with feature selection for breast cancer diagnosis. *Expert Syst. Appl.* **2009**, *36*, 3240–3247. [CrossRef]
- Kumar, G.R.; Nagamani, K. Banknote authentication system utilizing deep neural network with PCA and LDA machine learning techniques. *Int. J. Recent Sci. Res.* 2018, 9, 30036–30038.
- Kumar, C.; Dudyala, A.K. Bank note authentication using decision tree rules and machine learning techniques. In Proceedings of the 2015 International Conference on Advances in Computer Engineering and Applications, Ghaziabad, India, 19–20 March 2015; pp. 310–314.
- 49. Jaiswal, R.; Jaiswal, S. Banknote Authentication using Random Forest Classifier. Int. J. Digit. Appl. Contemp. Res. 2019, 7, 1-4.
- 50. Sarma, S.S. Bank Note Authentication: A Genetic Algorithm Supported Neural based Approach. *Int. J. Adv. Res. Comput. Sci.* **2016**, *7*, 97–102.
- 51. Lyon, R.J.; Stappers, B.W.; Cooper, S.; Brooke, J.M.; Knowles, J.D. Fifty Years of Pulsar Candidate Selection: From simple filters to a new principled real-time classification approach. *Mon. Not. R. Astron. Soc.* **2016**, 459, 1104–1123. [CrossRef]
- Wang, Y.; Pan, Z.; Zheng, J.; Qian, L.; Li, M. A hybrid ensemble method for pulsar candidate classification. *Astrophys. Space Sci.* 2019, *8*, 1–13. [CrossRef]
- Sağlam, F.; Yıldırım, E.; Cengiz, M.A. Clustered Bayesian classification for within-class separation. *Expert Syst. Appl.* 2022, 208, 118152. [CrossRef]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.