PONTIFÍCIA UNIVERSIDADE CATÓLICA
DO RIO DE JANEIRO

Renan Piazzaroli Finotti Amaral

On Interval Type-2 Fuzzy Logic System using the Upper and Lower Method for Supervised Classification Problems

Tese de Doutorado

Thesis presented to the Programa de Pós-graduação em Engenharia Mecânica of PUC-Rio in partial fulfillment of the requirements for the degree of Doutor em Engenharia Mecânica.

Advisor: Prof. Ivan Fabio Mota de Menezes

Rio de Janeiro
September 2021

Renan Piazzaroli Finotti Amaral

On Interval Type-2 Fuzzy Logic System using the Upper and Lower Method for Supervised Classification Problems

Thesis presented to the Programa de Pós-graduação em Engenharia Mecânica of PUC-Rio in partial fulfillment of the requirements for the degree of Doutor em Engenharia Mecânica. Approved by the Examination Committee:

Prof. Ivan Fabio Mota de Menezes
Advisor
Departamento de Engenharia Mecânica – PUC-Rio

Prof. D.Sc. Anderson Pereira
Departamento de Engenharia Mecânica – PUC-Rio

Prof. D.Sc. Helon Vicente Hultmann Ayala
Departamento de Engenharia Mecânica – PUC-Rio

Prof. D.Sc. Leonardo Goliatt da Fonseca
Departamento de Mecânica Aplicada Computacional – UFJF

Prof. D.Sc. Saul de Castro Leite
Centro de Matemática Computação e Cognição – UFABC

Rio de Janeiro, September the 3rd, 2021

Renan Piazzaroli Finotti Amaral

Received the Bachelor of Science degree in Mechanical Engineering from Federal University of Juiz de Fora (UFJF), Brazil, in 2015 and Master of Science in Electrical Engineering from Federal University of Juiz de Fora (UFJF), Brazil, in 2017.

To my parents and brothers.

Acknowledgments

First, I would like to thank God and fate who put good and bad moments in my path. I appreciate the bad moments because the process of overcoming them made me grow, which evolve me personally and professionally. Also, these moments made me remember and know what kind of person I do not want to become. I am grateful for the good moments because they provided me with beautiful memories, bringing good feelings just for remembering them.

Thanks to my parents Cleide and Devailton for their support, encouragement, and affection. To my brothers Rafaelle and Rafael for their brotherhood, messes, companionship, and friendship. I thank Sara Bassoli for the friendship, partnership, and affection throughout this stage in my life. I also thank my friend Isadora, Mr. Rubem (in memoriam), Aunt Sirley, and the Bastos family for the welcome, affection, and support that they gave me during my D. Sc. course. I thank all my friends made at PUC-Rio for all the help and friendship that they gave me. To the colleagues I made during my undergraduate and master's degree who became valued friends. Last but not least, I would like to thank my advisor Ivan for all his help, guidance and for showing me how to be a great person and a great professional.

Abstract

Amaral, Renan Piazzaroli Finotti; Menezes, Ivan Fabio Mota (Advisor). **On Interval Type-2 Fuzzy Logic System using the Upper and Lower Method for Supervised Classification Problems**. Rio de Janeiro, 2021. 76p. Tese de Doutorado – Departamento de Engenharia Mecânica, Pontifícia Universidade Católica do Rio de Janeiro.

Fuzzy logic systems are machine learning techniques that can model mathematically uncertainties. They are divided into type-1 fuzzy, and type-2 fuzzy logic systems. The type-1 fuzzy logic system has been widely applied to solve several problems related to machine learning, such as control, classification, clustering, prediction, among others. However, as it presents a better mathematical modeling of uncertainties, the type-2 fuzzy logic system has received much attention over the years. This modeling improvement is also accompanied by an increase in mathematical and computational effort. Aiming to reduce these issues to solve classification problems, this work presents the development and comparison of two Gaussian membership functions for a type-2 interval fuzzy logic system using the upper and lower method. Gaussian membership functions with uncertainty in the mean and with uncertainty in the standard deviation are used. Both fuzzy models covered in this work are trained by algorithms based on first order information. Furthermore, this work proposes the extension of interval type-2 fuzzy models to present multiple outputs, significantly reducing the computational cost in solving multiclass classification problems. Finally, aiming to contextualize the use of these models in mechanical engineering applications, this work presents the solution of a problem of fault detection in aircraft gas turbines.

Keywords

Interval type-2 fuzzy logic system; Classification problem; Upper and lower method; Uncertain mean; Uncertain standard deviation.

Resumo

Amaral, Renan Piazzaroli Finotti; Menezes, Ivan Fabio Mota. **Sistemas de Inferência Fuzzy Intervalar do Tipo-2 usando o Método Superior e Inferior para Problemas de Classificação Supervisionados**. Rio de Janeiro, 2021. 76p. Tese de Doutorado – Departamento de Engenharia Mecânica, Pontifícia Universidade Católica do Rio de Janeiro.

Os sistemas de inferência fuzzy são técnicas de aprendizado de máquina que possuem a capacidade de modelar incertezas matematicamente. Eles são divididos em sistemas de inferências fuzzy tipo-1 e fuzzy tipo-2. O sistema de inferência fuzzy tipo-1 vem sendo amplamente aplicado na solução de diversos problemas referentes ao aprendizado de máquina, tais como, controle, classificação, clusterização, previsão, dentre outros. No entanto, por apresentar uma melhor modelagem matemática das incertezas, o sistema de inferência fuzzy tipo-2 vem ganhando destaque ao longo dos anos. Esta melhora modelagem vem também acompanhada de um aumento do esforço matemático e computacional. Visando reduzir tais pontos para solucionar problemas de classificação, este trabalho apresenta o desenvolvimento e a comparação de duas funções de pertinência Gaussiana para um sistema de inferência fuzzy tipo-2 intervalar usando o método superior e inferior. São utilizadas as funções de pertinência Gaussiana com incerteza na média e com incerteza no desvio padrão. Ambos os modelos fuzzy abordados neste trabalho são treinados por algoritmos baseados em informações de primeira ordem. Além disso, este trabalho propõe a extensão dos modelos fuzzy tipo-2 intervalar para apresentarem múltiplas saídas, reduzindo significativamente o custo computacional na solução de problemas de classificação multiclasse. Finalmente, visando contextualizar a utilização desses modelos em aplicações de engenharia mecânica, este trabalho apresenta a solução de um problema de detecção de falhas em turbinas a gás, utilizadas em aeronaves.

Palavras-chave

Sistema de inferência fuzzy tipo-2 intervalar; Problema de classificação; Método superior e inferior; Incerteza na média; Incerteza no desvio padrão.

Table of contents

List of figures

List of tables

List of Abreviations

ANFIS – adaptive neuro fuzzy inference system

ANN – artificial neural network

BC – binary classifier

BCP – binary classification problem

CP – classification problem

FBF – fuzzy basis function

FL – fuzzy logic

FLS – fuzzy logic system

FLSMO – fuzzy logic system multiple output

FOU – footprint of uncertainty

GFS – genetic fuzzy system

GMFum – Gaussian membership function with uncertain mean

GMFus – Gaussian membership function with uncertain standard deviation

GT2-FLS – general type-2 fuzzy logic system

IT2-FLS – interval type-2 fuzzy logic system

KM – Karnik-Mendel

LMF – lower MF

MCP – multiclass classification problem

MF – membership function

ML – machine learning

MSE – mean squared error

OvA – one-vs-all

OvO – one-vs-one

SCG – scaled conjugate gradient

SD – steepest descent

T1-FLS – type-1 fuzzy logic system

T2-FLS – type-2 fuzzy logic system

TSK – Takagi-Sugeno-Kang

UL – upper and lower method

UMF – upper MF

Live as if you were to die tomorrow. Learn as if you were to live forever.

Mahatma Gandhi.

1
Introduction

Machine Learning (ML) has increased over the years, and many ML models have been developed, such as Artificial Neural Network (ANN) [1], Support Vector Machine [2], Random Forest [3], Decision Tree [4], and Fuzzy Logic System [5]. Each of them has its characteristics being more or less indicated depending on the application, which can be used in different fields, such as medical, financial, engineering, and many others. In the medical area, the ML was employed for cancer diagnosis [6] using a random forest classification algorithm for diagnosing six cancers using five features from the isoform of microRNA. Mohan et al. [7] performed a heart disease prediction using a hybrid ML model. Moreover, the authors in [8] presented a survey showing the ML algorithm and models that can be used to aid the battling of the covid-19 virus. Regarding financial applications, the authors in [9] applied ML to predict the stock market and regional relative directions based on financial network indicators. Hajek and Henriques [10] used a feature selection and classifiers based on ML to detect financial statement fraud analyzing annual reports. Moreover, in [11] is used a ML called Ant Colony Optimization to predict financial crisis. Focusing on engineering applications the authors in [12] used the ML to diagnose fault for single- and multi-faults in induction motors. Gkerekos et al. [13] applied the ML models for predicting ship main engine fuel oil consumption. Metawa et al. [14] use ML to predict the gas turbine performance which can help power plants to study and quantify performance degradation over time.

The application examples, as mentioned above, use different ML models. However, an ML model that can model uncertainties mathematically is the Fuzzy Logic System (FLS) which is the focus of this work. Developed by Lotfi A. Zadeh in 1965 [15], fuzzy logic (FL) is inspired by the processes of human perception and cognition. Classical logic has only the truth and false while in the FL was developed the mathematical concept of partial truth and partially false. Based on this concept, the FLS emerged by promising to deal with problems owning uncertainties that may not be modeled by well-established theories, such as statistical theory [16]. This capability to model uncertainties has been applied in several problems, such as in [17] which an FLS was used to

reduce the residential load in smart grids. In the same direction, the authors in [18] used an FL control as an intelligent supervisor for the power balance in a hybrid Wind/Photovoltaic/Diesel system with battery storage. Taghavifar and Rakheja [19] applied the FLS together with ANN to obtain the desired path-tacking and lane-keeping of autonomous vehicles. In [20] the authors applied an FLS to classify Short Circuit GMA Welding. Moreover, Wu et al. [21] defined risk factors for ship-bridge collision using an FLS approach. All these applications show that the FLS has wide applicability in different areas.

The core of the FLS is the fuzzy set which is characterized by membership functions (MFs). These fuzzy sets can be of two types namely type-1 (T1-FLS) and type-2 (T2-FLS) [16]. A type-1 fuzzy set is represented by a two-dimensional function, and owing to its simplicity, it has been used in a variety of applications, such as forecasting short-term traffic flow [22], implementation of an autonomous approach in smart grids for residential energy management [17], detection of network anomaly (where the FLS is used to decide whether an instance represents an anomaly or not) [23], and fault detection in grid-connected photovoltaic systems using an FLS classifier [24], among others. In contrast, the T2-FLS uses a type-2 fuzzy set, which is represented by a three-dimensional function. The T2-FLS enhances the capability to handle uncertainties [5], and this search for improvement has been investigated by many researchers over the past few years. Jiang et al. [25] used an interval T2-FLS (IT2-FLS) for stock index forecasting based on a fuzzy time series and a fuzzy logical relationship map. In [26], the application of an IT2-FLS classifier to the motor imagery electroencephalogram discrimination task was investigated, and the results show that IT2-FLS outperforms some state-of-the-art brain-computer interface classifiers. Sabahi et al. [27] designed a combination of T2-FLS and a conventional feedback controller for the load-frequency control of a nonlinear time-delay power system, and showed that the combination was efficient even in the presence of long time delays. In [28], an IT2-FLS was used to solve the lifetime maximization problem of wireless sensor networks, and it was shown that the model improved the network performance and enhanced the network lifetime.

The T2-FLS is composed of five parts, namely fuzzification, rules, inference, type-reducer, and defuzzification. During fuzzification, the MFs have two grades of membership, the primary and secondary, which can have any value between 0 and 1, forming the general T2-FLS (GT2-FLS) [29]. However, the GT2-FLS results in high mathematical and computational effort, making it difficult to use. To overcome this issue, the Interval T2-FLS (IT2-FLS) works as a T2-FLS having the secondary membership grade equal to 1 in all domains

[5]. However, even though IT2-FLS reduces the computational costs, it involves some other challenges, such as the need to determine the number and the type of fuzzy rules to use and performing type-reduction in IT2-FLS.

Concerning the first challenge of IT2-FLS (i.e., determining the rules), there are three main approaches: the first one is known as *Ad Hoc* Data Covering Methods which are based on the process that is guided by covering criteria in the training set [30, 31]. The other two approaches for designing the FLS have been widely investigated, which are the merged model that associates the fuzzy rule-based method with a neural network (ANFIS) [32], with the one that uses an evolutionary optimization method associated with FLS, such as a genetic fuzzy system (GFS) [33]. Another less commonly used approach is to design an FLS using information based on the gradient to update a fixed number of rules created initially by the user [34]. In [35], this approach is implemented using the Steepest Descent (SD) and the Scaled Conjugate Gradient (SCG) training methods, where the former uses first-order information, whereas the latter uses second-order information. The results obtained in [35] indicate that both training methods lead to less computational cost and a tendency to use fewer rules when compared to evolutionary fuzzy approaches. Also, this approach has shown promising results for classification problems applied to diagnostic gas turbines [36], switch machines faults [37], and welding process [20]. On the other hand, the second challenge is related to the type-reducer, which transforms a type-2 output fuzzy set into a type-1 output fuzzy set to be used for computing the crisp output during defuzzification [5]. One of the most popular methods used to perform type-reduction is the Karnik-Mendel (KM) iterative approach [38]. However, this approach is computationally intensive, especially when there are many MFs and the rule base is large [39]. To avoid the use of the KM iterative approach, the authors in [40] proposed a new method called the Upper and Lower for IT2-FLS using a Gaussian MF with uncertain mean (GMFum), which was trained using the SD method. The proposed method reduced the computational costs as well as improved the convergence speed and classification ratio when compared to the KM approach. Despite this, the IT2-FLS using GMFum is still computationally intensive, especially when a gradient-based optimization method is used for training since its membership functions are represented in five different segments [5].

Among the different applications of FLS, this work focuses on classification problems (CPs), which are formulated as discrimination of patterns among classes [41]. The CPs can be divided into binary CP (BCP) and multiclass CP (MCP). The models used for solving MCP can be binary classifiers

(BCs) (i.e., classifiers having a single output, such as a support vector machine [2]) or multiclass classifiers (i.e., classifiers with multiple outputs, such as an artificial neural network [42]). The FLS model belongs to BC [43], which needs a binary decomposition strategy to split the original MCP dataset into several binary classification subsets. The main decomposition strategies are One-versus-One (OvO) [44] and One-versus-All (OvA) [45]. The OvO strategy generates binary classification subsets to classify one class from each of the other classes, requiring a total of $\Upsilon\left(\Upsilon-1\right)/2$ classifiers, where $\Upsilon \in \mathbb{N}^*$ is the number of classes in the problem. In contrast, the OvA strategy performs Υ binary decompositions, where each subset is made to distinguish one class from all the other classes. The authors in [43], show that T1-FLS using the decomposition strategy increases the computational costs in a highly non-linear manner with an increase in the number of classes in an MCP. To avoid the use of this strategy, also in [43], an extension of T1-FLS was developed to support multiple outputs (T1-FLSMO). The results demonstrated that T1-FLSMO achieves the same performance as that of T1-FLS using the OvA strategy, significantly reducing the computational effort of MCPs. Figure 1.1 shows the timeline of the relevant papers that supports the scientific innovation of this dissertation.

In this way, this dissertation attempts to reduce the computational costs of IT2-FLS using the Upper and Lower method with GMFum (IT2-FLS UL-GMFum) for CP. Aiming to accomplish the objective, it is proposed to use the Gaussian MF with uncertain standard deviation (GMFus) in IT2-FLS UL (IT2-FLS UL-GMFus) trained by the SD method. GMFus is differentiable in all domains, thereby requiring fewer equations to perform fuzzification, and consequently to compute its first-order information. The derivatives of the equations for IT2-FLS UL-GMFus as well as their gradient expressions are presented, which enables the use of any training method based on its first-order information. In addition, the concepts of IT2-FLS UL-GMFum and IT2-FLS UL-GMFus are extended for multiple outputs to obtain the equations of their models and compute their gradients.
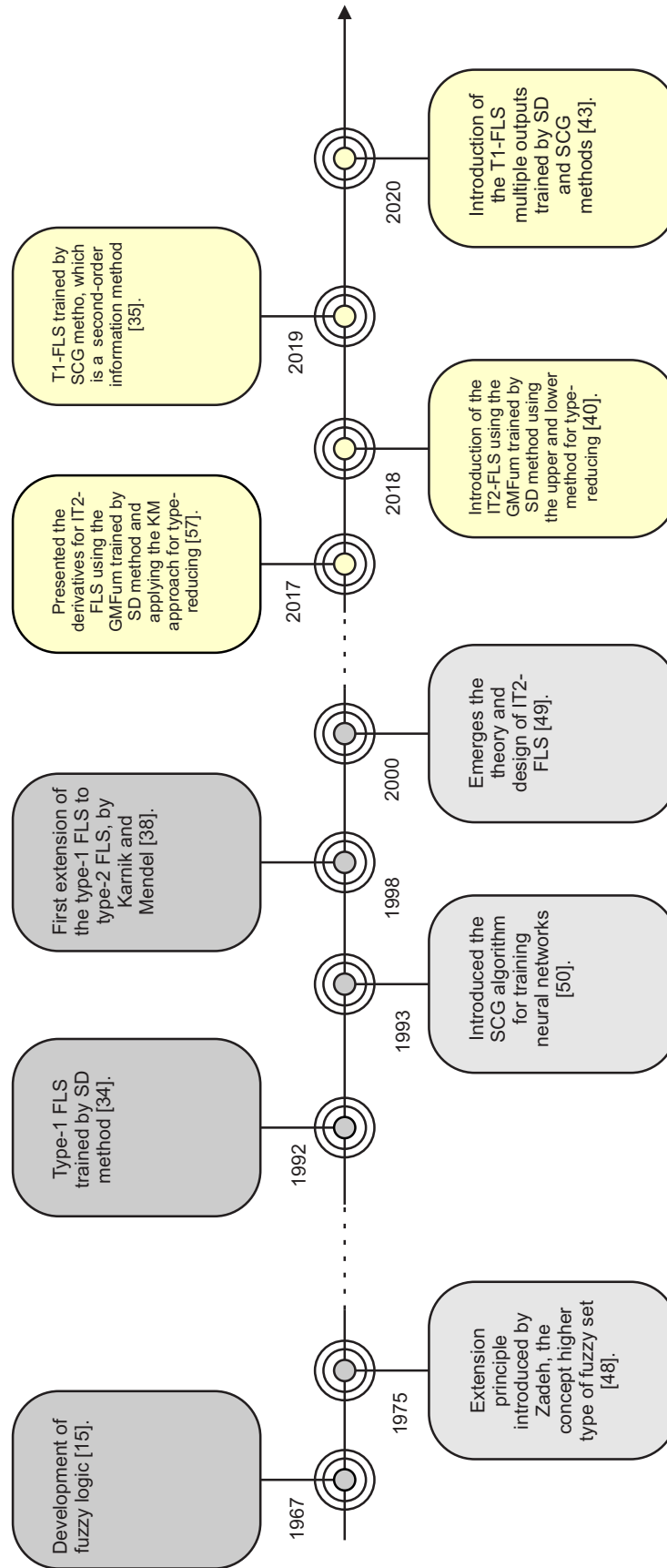
Figure 1.1: Timeline to present the relevant works for this dissertation.

1.1
Objective and main contributions

The objective of this dissertation is to reduce the computational costs of IT2-FLS using the Upper and Lower method trained by gradient-based method for solving CP, and the main contributions can be summarized as follows:

– The derivation of the equations to compute the gradient of IT2-FLS UL-GMFus, reducing the computational costs during training using the steepest descent method;

– The extension of IT2-FLS UL-GMFum and IT2-FLS UL-GMFus to FLSMO, enabling them to handle multiclass classification problems. This avoids the use of binary decomposition strategies, consequently reducing computational costs;

– The derivation of the equations to compute the gradient for both IT2-FLSMOs UL, facilitating the use of first-order information training methods available in the literature [35];

– A comparison of the IT2-FLS models with T1-FLS and T1-FLSMO using the SD and SCG training methods;

– A performance analysis taking into account the accuracy, mean squared error (MSE), time consumed during the training phase, and well-established classification metrics based on the datasets provided by Knowledge Extraction based on Evolutionary Learning [46] and UCI Machine Learning Repository [47].

– A discussion and contextualization of a mechanical engineering application to detect faults in aircraft gas turbine using the IT2-FLS proposed models.

This dissertation is organized as follows. Chapter 2 presents the problem formulation and the main concepts of T1-FLS and T2-FLS; Chapter 3 presents the proposals IT2-FLS using UL-GMFum and UL-GMFus for solving BCP and MCP; Chapter 4 analyzes the experimental results for the benchmark dataset and aims to solve a practical application of the FLS proposed herein; Chapter 5 states the main conclusions and suggests some topics for future investigation; Appendix A and Appendix B outline the equations to compute the gradients of IT2-FLS UL-GMFum and IT2-FLS UL-GMFus, respectively; Appendix C and Appendix D present the equations for computing the gradients of IT2-FLSMO UL-GMFum and IT2-FLSMO UL-GMFus, respectively. Appendix E presents the results obtained using the SD training method with a step size value equal to 0.01, and finally the Appendix F show the papers published during the Doctor of Science course.

2
Problem formulation

2.1
Membership functions

Membership functions (MFs) characterize a fuzzy set, which can be of type-1 or type-2. A type-1 fuzzy set A in a universe X is represented as a set of ordered pairs of a generic element x, where its membership grade is defined as [51]

$$A = \left\{ (x, \mu_A(x)) | \, \forall x \in X \right\}, \tag{2-1}$$

which $\mu_A : X \to [0, 1]$. Figure 2.1 depicts an example of a type-1 MF.
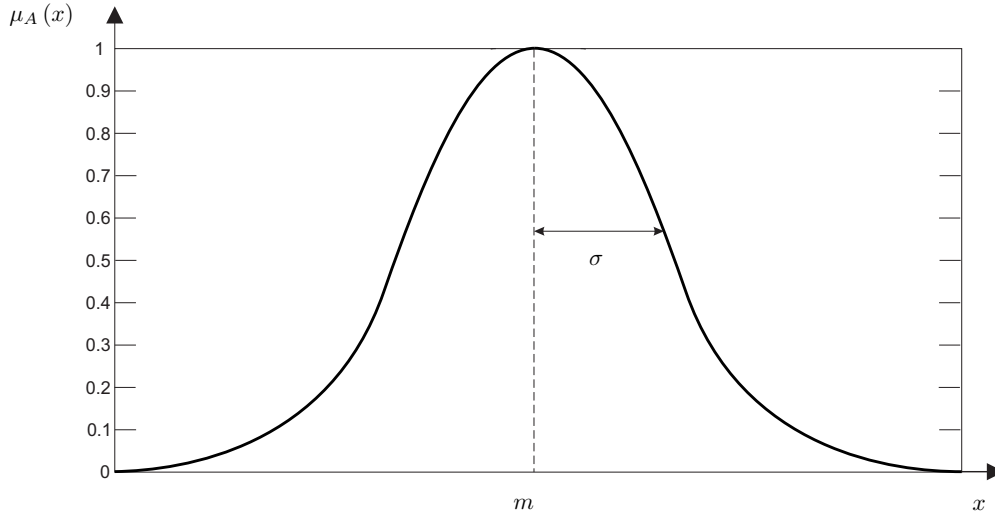
Figure 2.1: Example of a Gaussian type-1 MF.

A type-1 fuzzy set has two dimensions. Using the concept of fuzzy set and extending it to three dimensions, a general type-2 fuzzy set is formed [51]

$$\tilde{A} = \left\{ ((x, u), \mu_{\tilde{A}}(x, u)) | \, \forall x \in X, \forall u \in J_x \subseteq [0, 1] \right\}, \tag{2-2}$$

where u is equal to $\mu_A(x)$, X is the primary domain and J_x is the secondary domain. An example of the general type-2 MF is presented in Figure 2.2, where the fuzzy set models the uncertainties better than type-1 fuzzy set. However, the usage of a general type-2 fuzzy set leads to higher mathematical and computational efforts. Aiming to overcome and to reduce these efforts,
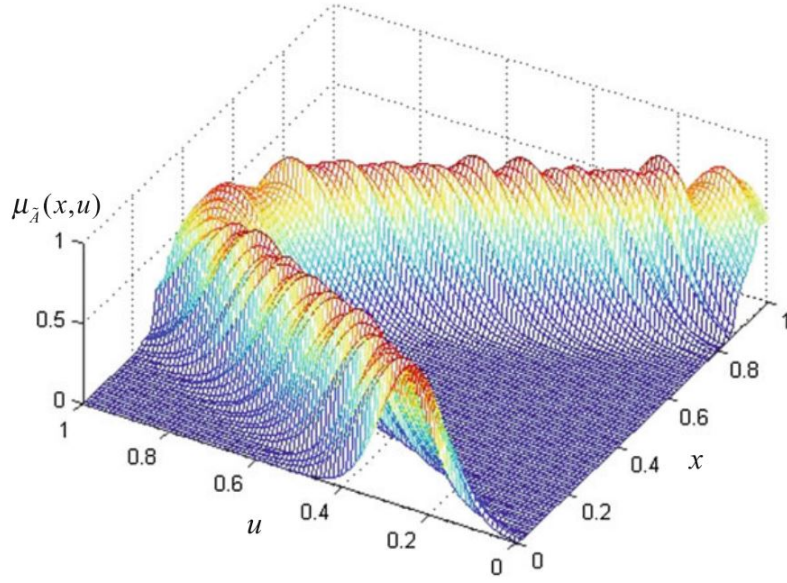
Figure 2.2: Example of a general type-2 MF [5].

there are the interval type-2 fuzzy set $\tilde{A}$, denoted by $\underline{\mu}_{\tilde{A}}(x)$ and $\overline{\mu}_{\tilde{A}}(x)$, representing respectively, the lower and upper membership functions of $\mu_{\tilde{A}}(x)$. The membership grade of an interval type-2 fuzzy set is defined as [51]

$$\tilde{A} = \left\{ ((x, u), 1) \middle| \forall x \in X, \forall u \in J_x \subseteq [0, 1] \right\}. \tag{2-3}$$

It is important to notice in Equation (2-3) that all secondary grades are equal to 1, which simplify the complexity of mathematical expressions for the type-2 fuzzy sets. Figure 2.3 shows an example of an interval type-2 MF.
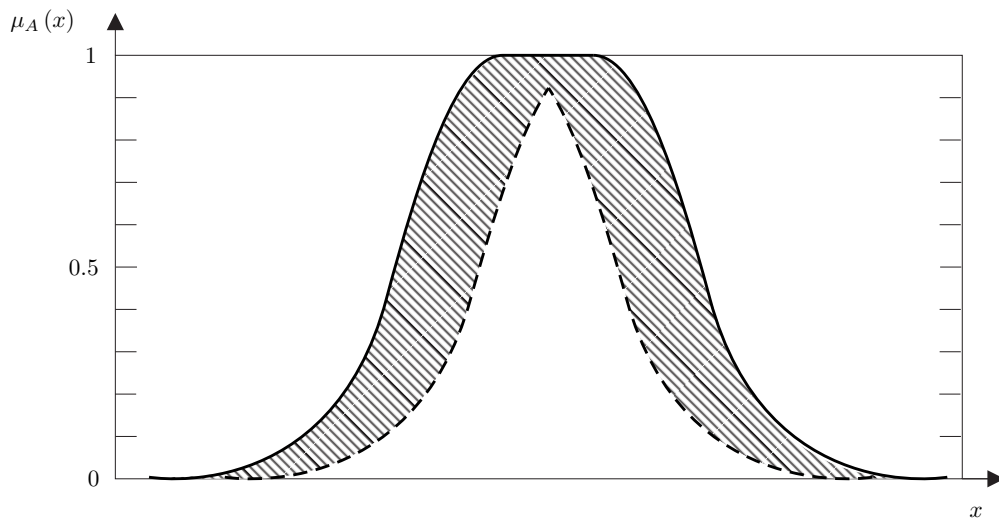
Figure 2.3: Example of an interval type-2 MF.

2.2
Type-1 fuzzy logic system

The T1-FLS is composed of four parts as shown in Figure 2.4, which are known as rules, fuzzifier, inference and defuzzifier. Before the discussion of each part, it is important to notice that there are two models for fuzzy sets, Mamdani or Takagi-Sugeno-Kang (TSK). The difference between the models is in the output of the fuzzy rules. Mamdani model has the output of the fuzzy rules as a fuzzy set, while TSK model has the output of the fuzzy rules as a weighted average or weighted sum (i.e., a function as output of the rules) [5]. Hence, the TSK model is not on the scope of this work. Every FLS which is presented and discussed in this work refers to the Mamdani model.
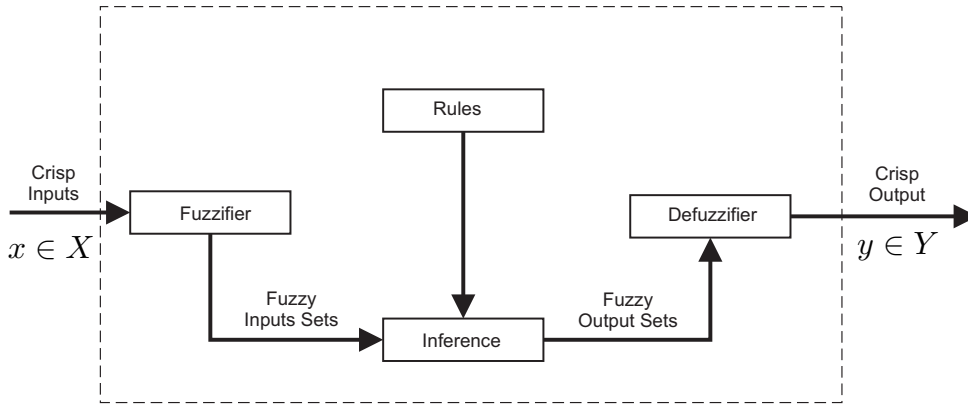
Figure 2.4: Structure of a T1-FLS.

The fuzzy rules use the "IF-THEN" structure and for a T1-FLS having P inputs $x_1 \in X_1, \ldots, x_P \in X_P$ and one output $y \in Y$, the l-th rule of an FLS has the form [16]

$$rule^l : \text{IF } x_1 \text{ is } F_1^l \text{ AND } \ldots \text{ AND } x_P \text{ is } F_P^l \text{ THEN } y \text{ is } Y^l, \qquad (2\text{-}4)$$

where F_P^l is the P-th type-1 fuzzy set associated with l-th rule. The rule presented in Equation (2-4) represents the relationship between the input spaces $X_1 \ldots X_P$ and the output space Y of the FLS. Moreover, the connectives of the fuzzy rules can be "AND"[1], "OR"[2] or their complements forming the "NOT" [16]. The fuzzifier part is responsible to map the crisp inputs $\mathbf{x} = [x_1, \ldots, x_P]^T \in X_1 \times \cdots \times X_P \equiv \mathbf{X}$ into a fuzzy set $A_{\mathbf{x}}$ in $\mathbf{X}$. The fuzzification is done using the MF, which can be Gaussian, trapezoidal piecewise-linear, triangular and others. The inference part uses the fuzzy logic principles to combine the fuzzy "IF-THEN" rules from the fuzzy rule base

---

[1] t-norm

[2] t-conorm

into a mapping from fuzzy inputs sets to fuzzy output sets [16]. Each rule is interpreted as a fuzzy implication, so that Equation (2-4) can be rewritten as

$$rule^l : F_1^l \times \cdots \times F_P^l \rightarrow Y^l = A^l \rightarrow Y^l, \qquad (2\text{-}5)$$

where R^l is described by the MF $\mu_{R^l}(\mathbf{x}, y)$ as

$$\mu_{R^l}(\mathbf{x}, y) = \mu_{A^l \rightarrow Y^l}(\mathbf{x}, y). \qquad (2\text{-}6)$$

The last part of the fuzzy structure is the defuzzifier, which produces the T1-FLS crisp output from the output fuzzy sets. There are many methods for defuzzification, such as, maximum, mean-of-maxima, centroid, center-of-sums, height, modified height and center-of-sets [16].

To design a T1-FLS aiming to attain a good performance for solving a problem requires the right choice for each part of the structure presented in Figure 2.4. There are many alternatives to design an FLS and each of them is more or less suitable depending of the problem to solve (i.e., CP, prediction, clustering). This work focus on CP and all the FLS presented herein will be designed for solving it.

2.3
Classification problems

In the field of ML problems, there are two kinds of problems called supervised or unsupervised problems. The first one is described when the learning associates the input with some output [52], which gives a training set of input-output pairs $(\mathbf{x} : y)$. However, in many cases, the output y may be challenging to collect, forming the unsupervised problems. In these problems, there is no output, and the data set has only 'features' becoming to the users of the ML models to define its objective. In other words, unsupervised problems refer to extract information from the data that do not require human labor to annotate examples [52]. In this sense, the classification problems can be also divided into supervised and unsupervised problems. The first one is called a supervised classification problem, and the second one is commonly called a clustering problem [53]. This work will focus on supervised CP, and all mention of CP will mean to supervised CP.

The CP is stated as a mapping between a vector $\mathbf{x} \in \mathbb{R}^P$ and a class y, where P is the number of features from a given sample with $P \in \mathbb{N}^*$. In other words, the input space is divided into "decision regions" where each region is assigned to one class [41].

2.4
Type-1 fuzzy logic system binary classifier

Adopting the FLS binary classifier, and considering a set of input-output pairs $\left(\mathbf{x}^{(q)} : y^{(q)}\right)$, where $q = 1, \ldots, Q$ and Q is the number of patterns used for training the T1-FLS, the solution of a binary CP can be obtained as the minimization of a given cost function. Typically, a cost function is expressed as [35]

$$J\left(\mathbf{w}^{(\gamma)}\right) = \frac{1}{2}\left[f_{T1}\left(\mathbf{x}^{(q)}\right) - y^{(q)}\right]^2, \tag{2-7}$$

where $\mathbf{w}^{(\gamma)}$ is a vector holding the parameters for T1-FLS in the γ-th iteration, and $f_{T1}(\mathbf{x})$ is the T1-FLS output value. Adopting singleton fuzzification, max-product compositions, product implication[3], and the height defuzzifier, the output of T1-FLS is expressed by [16, 35]

$$f_{T1}(\mathbf{x}) = \frac{\sum\limits_{l=1}^{M} \theta_l \prod\limits_{k=1}^{P} \mu_{F_k^l}(x_k)}{\sum\limits_{l=1}^{M} \prod\limits_{k=1}^{P} \mu_{F_k^l}(x_k)}, \tag{2-8}$$

where x_k is the k-th element of the vector $\mathbf{x}$, $\prod$ denotes the product operator, $\mu_{F_k^l}(x_k)$ is the membership function (MF) associated with the k-th input of the l-th rule, and θ_l is the weight of the height deffuzification associated with the l-th rule, $(l = 1, \cdots, M)$. Equation (2-8) can also be expressed as

$$f_{T1}(\mathbf{x}) = \sum_{l=1}^{M} \theta_l \phi_l(\mathbf{x}), \tag{2-9}$$

where $\phi_l(\mathbf{x})$ is called a fuzzy basis function (FBF) [16] and defined as

$$\phi_l(\mathbf{x}) = \frac{\prod\limits_{k=1}^{P} \mu_{F_k^l}(x_k)}{\sum\limits_{l=1}^{M} \prod\limits_{k=1}^{P} \mu_{F_k^l}(x_k)}. \tag{2-10}$$

Equation (2-8) represents the entire structure of T1-FLS presented in Figure 2.4. As mentioned in Section 2.2, each part of the FLS structure directly influences the final performance. However, the core of the FLSs are the fuzzy rules which have the most impact at the T1-FLS performance, and how to define them is widely researched. These rules can be defined by an specialist through its knowledge about the problem or using some optimization methods for training the rules using the data problem. In addition, it is important to notice that an FLS with high number of rules tends to achieve a better performance than an FLS with low number of rules. But this gain in performance comes with high computational costs, consequently with an

---

[3]t-norm

increased time consuming for training it.

Generally, the main approaches to design the rules of an FLS are the merged model that associates the fuzzy rule-based method with a neural network (ANFIS) [32], and the one which uses an evolutionary optimization method associated with FLS, such as a genetic fuzzy system (GFS) [33]. However, both of these merged models reduce or even end an important property of FLS called interpretability [54]. This property means to the capability of conjugating a complex behavior with a simple description through the fuzzy rules [55]. The importance of the interpretability is to understand why the ML model achieved that output, which features are more relevant for an application, and to avoid the usage of black box ML [56]. In this way, a T1-FLS classifier that uses the first and second-order information for training the fuzzy rule parameters is described in [35], where the fuzzy rules are extracted from the data using only the type "AND" inference. The results in [35] show that the T1-FLS classifier achieved faster convergence rate, better classification performance and interpretability than other fuzzy and non-fuzzy methods due to the need of few rules to achieve good performance. Therefore, this work will use the FLS described in [35] for solving CP, the main idea is to use a gradient optimization method for training the fuzzy rules.

Considering the Gaussian MF which is differentiable at every point of its domain [35] and defined as

$$\mu_{F_k^l}(x_k) = \exp\left\{-\frac{1}{2}\left(\frac{x_k - m_{F_k^l}}{\sigma_{F_k^l}}\right)^2\right\}, \tag{2-11}$$

where $m_{F_k^l}$ and $\sigma_{F_k^l}^2$ are the mean and variance, respectively, the fuzzy rules can be expressed in terms of MF parameters as follows

$$rule^l = \left\{x_1 \text{ is } F_1^l\left(m_{F_1^l}, \sigma_{F_1^l}\right) \; AND \ldots AND \; x_P \text{ is } F_P^l\left(m_{F_P^l}, \sigma_{F_P^l}\right) \; THEN \; Y^l\right\}. \tag{2-12}$$

Hence, the problem consists in the minimization of the cost function (Equation (2-7)) updating the fuzzy parameters, $m_{F_k^l}$, $\sigma_{F_k^l}$ and θ_l [16]. Considering the simplest optimization method called Steepest Descent (SD), the training of the fuzzy rules is performed by searching a minimum by taking unidirectional minimizations at the opposite direction of the gradient of the cost function regarding to each fuzzy parameters. Thus for the cost function presented in Equation (2-7), the SD method optimizes the fuzzy parameters as

$$\mathbf{w}^{(\gamma+1)} = \mathbf{w}^{(\gamma)} - \alpha \nabla \mathbf{J}\left(\mathbf{w}^{(\gamma)}\right) \tag{2-13}$$

where $\nabla \mathbf{J}\left(\mathbf{w}^{(\gamma)}\right)$ is the gradient of the cost function evaluated at point $\mathbf{w} = \mathbf{w}^{(\gamma)}$, α is the step-size, and the vector of fuzzy parameters $\mathbf{w}^{(\gamma)}$ is

expressed as

$$\mathbf{w}^{(\gamma)} = \left[m_{F_1^1}(\gamma), \cdots, m_{F_P^1}(\gamma), \cdots, m_{F_1^M}(\gamma), \cdots, m_{F_P^M}(\gamma), \cdots, \right.$$
$$\sigma_{F_1^1}(\gamma), \cdots, \sigma_{F_P^1}(\gamma), \cdots, \sigma_{F_1^M}(\gamma), \cdots, \sigma_{F_P^M}(\gamma), \cdots, \quad (2\text{-}14)$$
$$\left. \theta_1(\gamma), \cdots, \theta_M(\gamma) \right]^T,$$

and the gradient $\nabla \mathbf{J}\left(\mathbf{w}^{(\gamma)}\right)$ is defined as

$$\nabla \mathbf{J}\left(\mathbf{w}^{(\gamma)}\right) = \left[\frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial m_{F_1^1}(\gamma)}, \cdots, \frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial m_{F_P^1}(\gamma)}, \cdots, \frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial m_{F_1^M}(\gamma)}, \cdots, \frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial m_{F_P^M}(\gamma)}, \cdots, \right.$$
$$\frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial \sigma_{F_1^1}(\gamma)}, \cdots, \frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial \sigma_{F_P^1}(\gamma)}, \cdots, \frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial \sigma_{F_1^M}(\gamma)}, \cdots, \frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial \sigma_{F_P^M}(\gamma)}, \cdots,$$
$$\left. \frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial \theta_1(\gamma)}, \cdots, \frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial \theta_M(\gamma)} \right]^T,$$
$$(2\text{-}15)$$

with the derivatives of $J\left(\mathbf{w}^{(\gamma)}\right)$ with respect to the parameters $m_{F_k^l}(\gamma)$, $\sigma_{F_k^l}(\gamma)$, and $\theta_l(\gamma)$ defined as

$$\frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial m_{F_k^l}(\gamma)} = \left[f_{T1}\left(\mathbf{x}^{(q)}\right) - y(q) \right] \left[\theta_l(\gamma) - f_{T1}\left(\mathbf{x}^{(q)}\right) \right] a_{F_k^l(\gamma)}^{(q)} \phi_l\left(\mathbf{x}^{(q)}\right), \quad (2\text{-}16)$$

$$\frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial \sigma_{F_k^l}(\gamma)} = \left[f_{T1}\left(\mathbf{x}^{(q)}\right) - y(q) \right] \left[\theta_l(q) - f_{T1}\left(\mathbf{x}^{(q)}\right) \right] b_{F_k^l(\gamma)}^{(q)} \phi_l\left(\mathbf{x}^{(q)}\right), \quad (2\text{-}17)$$

$$\frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial \theta_l(\gamma)} = \left[f_{T1}\left(\mathbf{x}^{(q)}\right) - y(q) \right] \phi_l\left(\mathbf{x}^{(q)}\right). \quad (2\text{-}18)$$

where $a_{F_k^l(\gamma)}^{(q)}$ and $b_{F_k^l(\gamma)}^{(q)}$ are created to simplify the above equations as follow

$$a_{F_k^l(\gamma)}^{(q)} = \frac{x_k^{(q)} - m_{F_k^l}(\gamma)}{\sigma_{F_k^l}^2(\gamma)}, \quad (2\text{-}19)$$

$$b_{F_k^l(\gamma)}^{(q)} = \frac{\left(x_k^{(q)} - m_{F_k^l}(\gamma)\right)^2}{\sigma_{F_k^l}^3(\gamma)}. \quad (2\text{-}20)$$

It is important to notice that T1-FLS is a binary classifier and only has one output, which implies a dependence on binary decomposition strategies for solving MCPs. This dependence makes this model less attractive if the number of classes is large. Aiming to solve this issue, the authors in [43] developed a new fuzzy model which has multiple outputs (i.e., the fuzzy output is a vector instead a crisp output). The results in [43] show that the new fuzzy model

attains the same classification performance using the fewer rules than T1-FLS.

2.5
Type-1 fuzzy logic system multiple outputs

The extension of T1-FLS to T1-FLSMO presented in [43] has the structure presented in Figure 2.5.
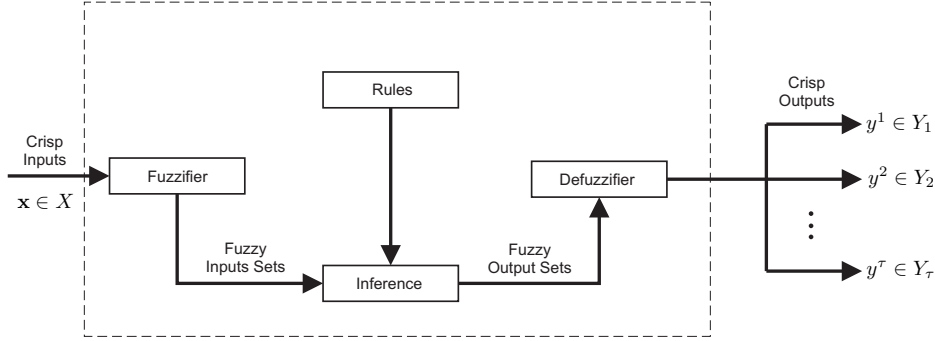
Figure 2.5: Structure of a T1-FLSMO.

The multiple outputs of T1-FLSMO are related to deffuzifier block. Using the height defuzzification which replaces each rule in an output fuzzy set with a singleton (θ_l) at the point having the maximum membership value in that output set. The extension to multiple outputs of T1-FLSMO is performed by calculating the centroid of the T1 set comprised of these singletons, creating a singleton (θ_l) for each output (t) as follows

$$
\Theta = \begin{bmatrix}
\theta_1^1 & \theta_2^1 & \cdots & \theta_M^1 \\
\theta_1^2 & \theta_2^2 & \cdots & \theta_M^2 \\
\vdots & \vdots & \ddots & \vdots \\
\theta_1^\tau & \theta_2^\tau & \cdots & \theta_M^\tau
\end{bmatrix},
\tag{2-21}
$$

where τ is the number of outputs and $t = 1, \cdots, \tau$. Therefore, the outputs of T1-FLSMO using a height deffuzifier are defined as

$$
\mathbf{y}^{(q)} = \mathbf{f}_{mo}\left(\mathbf{x}^{(q)}\right) = \Theta\Phi\left(\mathbf{x}^{(q)}\right),
\tag{2-22}
$$

where $\Phi\left(\mathbf{x}^{(q)}\right)$ is a vector of FBFs, which is expressed as

$$
\Phi\left(\mathbf{x}^{(q)}\right) = \begin{bmatrix} \phi_1\left(\mathbf{x}^{(q)}\right) & \phi_2\left(\mathbf{x}^{(q)}\right) & \cdots & \phi_M\left(\mathbf{x}^{(q)}\right) \end{bmatrix}^T.
\tag{2-23}
$$

In Equation (2-22), the subscript "mo" indicates that is a T1-FLSMO.

The rules of T1-FLS described in Equations (2-8) and (2-9) are extracted from the data using only inference of the type "AND" [35]. Consequently, the rules of T1-FLSMO are the same as those of T1-FLS, which can be represented as

$$rule_t^l : IF\ x_1\ is\ F_1^l\ AND\ \ldots\ AND\ x_P\ is\ F_P^l\ THEN\ y\ is\ Y_t^l. \qquad (2\text{-}24)$$

The structure of the rule "$IF\ x_1\ is\ F_1^l\ AND\ldots\ \ldots\ AND\ x_P\ is\ F_P^l$" corresponds to the antecedent from l-th rule and "$THEN\ y\ is\ Y_t^l$" is the consequent of the l-th rule for the t-th class. Assuming the Gaussian MF again, the fuzzy rules can be expressed in terms of MF parameters as follows

$$rule_t^l = \left\{ x_1\ is\ F_1^l\left(m_{F_1^l}, \sigma_{F_1^l}\right)\ AND\ldots AND\ x_P\ is\ F_P^l\left(m_{F_P^l}, \sigma_{F_P^l}\right)\ THEN\ Y_t^l \right\}.$$
$$(2\text{-}25)$$

Moreover, the cost function presented in Equation (2-7) is extended for T1-FLSMO as follows

$$J\left(\mathbf{w}^{(\gamma)}\right) = \frac{1}{2}\sum_{t=1}^{\tau}\left(f_{mo}^t\left(\mathbf{x}^{(q)}\right) - y^t(q)\right)^2, \qquad (2\text{-}26)$$

where $\mathbf{w}^{(\gamma)}$ denotes the vector parameters of T1-FLSMO, which are expressed as

$$\mathbf{w}^{(\gamma)} = \Big[m_{F_1^1}(\gamma), \cdots, m_{F_P^1}(\gamma), \cdots, m_{F_1^M}(\gamma), \cdots, m_{F_P^M}(\gamma), \cdots,$$
$$\sigma_{F_1^1}(\gamma), \cdots, \sigma_{F_P^1}(\gamma), \cdots, \sigma_{F_1^M}(\gamma), \cdots, \sigma_{F_P^M}(\gamma), \cdots, \qquad (2\text{-}27)$$
$$\theta_1^1(\gamma), \cdots, \theta_M^1(\gamma), \cdots, \theta_1^\tau(\gamma), \cdots, \theta_M^\tau(\gamma)\Big]^T.$$

Additionally, the gradient $\nabla\mathbf{J}\left(\mathbf{w}^{(\gamma)}\right)$ is defined as

$$\nabla\mathbf{J}\left(\mathbf{w}^{(\gamma)}\right) = \Bigg[\frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial m_{F_1^1}(\gamma)}, \cdots, \frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial m_{F_P^1}(\gamma)}, \cdots, \frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial m_{F_1^M}(\gamma)}, \cdots, \frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial m_{F_P^M}(\gamma)}, \cdots,$$
$$\frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial \sigma_{F_1^1}(\gamma)}, \cdots, \frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial \sigma_{F_P^1}(\gamma)}, \cdots, \frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial \sigma_{F_1^M}(\gamma)}, \cdots, \frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial \sigma_{F_P^M}(\gamma)}, \cdots,$$
$$\frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial \theta_1^1(\gamma)}, \cdots, \frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial \theta_M^1(\gamma)} \cdots \frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial \theta_1^\tau(\gamma)}, \cdots, \frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial \theta_M^\tau(\gamma)}\Bigg]^T,$$
$$(2\text{-}28)$$

and the derivatives of $J\left(\mathbf{w}^{(\gamma)}\right)$ with respect to the parameters $m_{F_k^l}(\gamma)$, $\sigma_{F_k^l}(\gamma)$, and $\theta_l^t(\gamma)$, are defined as

$$\frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial m_{F_k^l}(\gamma)} = \sum_{t=1}^{\tau}\left(\left[f_{mo}^t\left(\mathbf{x}^{(q)}\right) - y^t(q)\right]\left[\theta_l^t(\gamma) - f_{mo}^t\left(\mathbf{x}^{(q)}\right)\right]\right)a_{F_k^l(\gamma)}^{(q)}\phi_l\left(\mathbf{x}^{(q)}\right),$$
$$(2\text{-}29)$$

$$\frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial \sigma_{F_k^l}(q)} = \sum_{t=1}^{\tau}\left(\left[f_{mo}^t\left(\mathbf{x}^{(q)}\right) - y^t(q)\right]\left[\theta_l^t(q) - f_{mo}^t\left(\mathbf{x}^{(q)}\right)\right]\right)b_{F_k^l(\gamma)}^{(q)}\phi_l\left(\mathbf{x}^{(q)}\right)$$
$$(2\text{-}30)$$

and

$$\frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial \theta_l^t\left(q\right)} = \left[f_{mo}^t\left(\mathbf{x}^{(q)}\right) - y^t\left(q\right)\right]\phi_l\left(\mathbf{x}^{(q)}\right). \qquad (2\text{-}31)$$

The T1-FLSMO can deal with MCP without any decomposition strategy requiring fewer rules than T1-FLS, consequently reducing significantly the computational costs [43]. Otherwise, the T2-FLS handle with uncertainties better than T1-FLS [16]. In this way, this work proposes to extend the T1-FLSMO to T2-FLSMO aiming to improve the classification performance of the FLS for solving MCP.

2.6
Interval type-2 fuzzy logic system binary classifier

The inference block of a T2-FLS is shown in Figure 2.6, where the main difference with respect to the T1-FLS is the output processing block. In T2-FLS, it is necessary to reduce the output type-2 fuzzy set to a type-1 fuzzy set, and only after that can the defuzzifier method be applied to generate the crisp output [5].
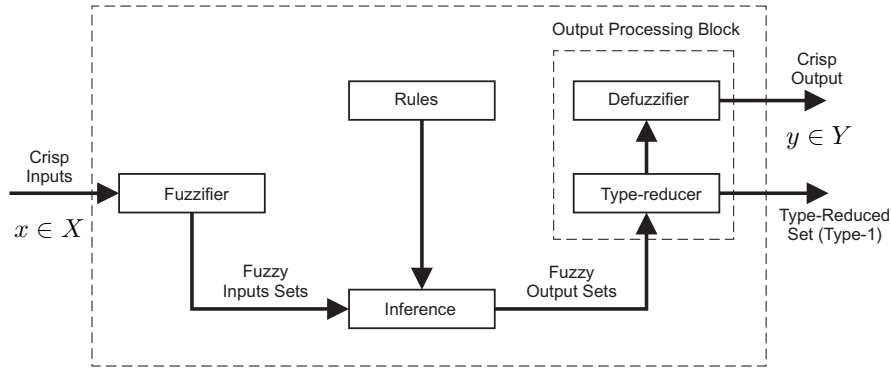
Figure 2.6: Structure of a T2-FLS [16].

2.6.1
Membership functions for interval type-2 fuzzy logic system

As mentioned in Section 2.1 there are the General T2-FLS and the Interval T2-FLS (IT2-FLS). The IT2-FLS uses the type-2 fuzzy set with less computational effort than General T2-FLS, this effort is done by representing the type-2 fuzzy set $\tilde{A}$ by its upper and lower MFs, see Equation (2-3). Considering the Gaussian MF which is differentiable in all domain, two different approaches can be used to characterize an IT2-FLS. The first one is a Gaussian MF with an uncertain mean (GMFum). It can take values in the range $[m_1, m_2]$ with a fixed standard deviation σ [16], and is described as

$$\mu_{\tilde{A}}\left(x\right) = \exp\left[-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2\right], \qquad m \in [m_1, m_2]. \qquad (2\text{-}32)$$

The MF described in Equation (2-32) is presented in Figure 2.7, where the shaded area is called the Footprint Of Uncertainty (FOU). The FOU represents the domain of uncertainty of $\tilde{A}$ [5], and can be expressed as

$$FOU\left(\tilde{A}\right) = \left\{(x,v)|\,\forall x \in X \ \text{ and } \ v \in \left[\underline{\mu}_{\tilde{A}}(x), \overline{\mu}_{\tilde{A}}(x)\right]\right\}, \qquad (2\text{-}33)$$

where $\overline{\mu}_{\tilde{A}}(x)$ and $\underline{\mu}_{\tilde{A}}(x)$ are called upper MF (UMF) and lower MF (LMF) of $FOU\left(\tilde{A}\right)$, as well as the upper and lower (type-1 fuzzy set) bounding functions of the FOU, respectively [5]. The UMF and LMF of the MF described in Equation (2-32) are given by

$$\overline{\mu}_{\tilde{A}}(x) = \begin{cases} \exp\left[-\frac{1}{2}\left(\frac{x-m_1}{\sigma}\right)^2\right] & , & x < m_1 \\ 1 & , & m_1 \leq x \leq m_2 \\ \exp\left[-\frac{1}{2}\left(\frac{x-m_2}{\sigma}\right)^2\right] & , & x > m_2 \end{cases} \qquad (2\text{-}34)$$

and

$$\underline{\mu}_{\tilde{A}}(x) = \begin{cases} \exp\left[-\frac{1}{2}\left(\frac{x-m_2}{\sigma}\right)^2\right] & , & x \leq \frac{m_1+m_2}{2} \\ \exp\left[-\frac{1}{2}\left(\frac{x-m_1}{\sigma}\right)^2\right] & , & x > \frac{m_1+m_2}{2} \end{cases} , \qquad (2\text{-}35)$$

respectively.
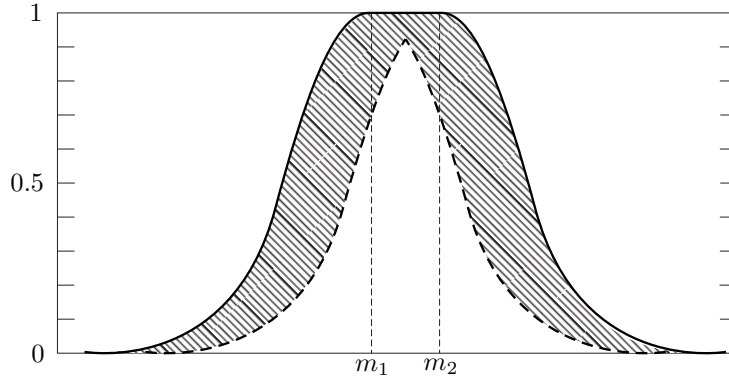
Figure 2.7: FOU of Gaussian MF with uncertain mean [16].

The second Gaussian MF, which characterizes IT2-FLS, is a Gaussian MF with an uncertain standard deviation (GMFus). It can take values in the range $[\sigma_1, \sigma_2]$ with a fixed mean m [16], expressed by

$$\mu_{\tilde{A}}(x) = \exp\left[-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2\right] \quad , \qquad \sigma \in [\sigma_1, \sigma_2], \qquad (2\text{-}36)$$

where the UMF and LMF are expressed as

$$\overline{\mu}_{\tilde{A}}(x) = \exp\left[-\frac{1}{2}\left(\frac{x-m}{\sigma_2}\right)^2\right] \qquad (2\text{-}37)$$

and

$$\underline{\mu}_{\tilde{A}}(x) = \exp\left[-\frac{1}{2}\left(\frac{x-m}{\sigma_1}\right)^2\right], \qquad (2\text{-}38)$$

respectively. Figure 2.8 shows the MF presented in Equation (2-36) and its FOU. It is important to note that the UMF and LMF presented by Equations (2-34) and (2-35) are not differentiable at all points; however, it is piecewise differentiable (i.e., each branch is differentiable over its segment domain) [16]. On the other hand, the UMF and LMF presented by Equations (2-37) and (2-38) are differentiable everywhere in all domains.
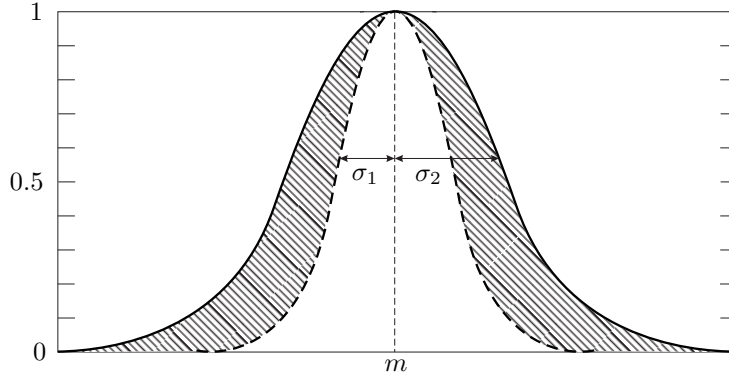
Figure 2.8: FOU of Gaussian MF with uncertain standard deviation [16].

2.6.2
Interval type-2 fuzzy logic system single output

Considering the height type-reducer, the authors in [57] provided the first-order information for training the IT2-FLS binary classifier using the SD method. The approach proposed in [57] involves type reduction with the Karnik and Mendel method [5], which uses an algorithm to compute the right and left end-points of IT2-FLS requiring at most $(M + 1)/2$ iterations. The algorithm determines the rule required to change the UMF to LMF for computing the left end-points and the rule required to change the LMF to UMF for calculating the right end-points [4]. Further, in [40], the same authors of [57] proposed a new method for IT2-FLS called upper and lower T2-FLS (IT2-FLS UL), which does not require the Karnik and Mendel method, thus reducing the computational costs. The idea is to compute the upper and lower end-points, using only the UMF for the upper end-points and the LMF for the lower end-points. The results show that the approach based on upper and lower end-points achieves better performance and convergence rates than the approach based on left and right end-points using the Karnik and Mendel method. In this way, this work proposes to extend the IT2-FLS UL single output to IT2-FLS UL multiple outputs (IT2-FLSMO UL) aiming to improve the classification performance of the fuzzy logic system.

---

[4]A completely description of the KM method is detailed in [16]

Adopting the upper and lower method, singleton fuzzification, max-product composition, product implication[5], height type-reduction using the upper and lower method and a height defuzzifier, the upper and lower outputs of IT2-FLS UL [40] are expressed respectively as

$$Y_{U_{T2}}(\mathbf{x}) = \frac{\sum\limits_{l=1}^{M} \theta_l \prod\limits_{k=1}^{P} \overline{\mu}_{\tilde{F}_k^l}(x_k)}{\sum\limits_{l=1}^{M} \prod\limits_{k=1}^{P} \overline{\mu}_{\tilde{F}_k^l}(x_k)} \qquad (2\text{-}39)$$

and

$$Y_{L_{T2}}(\mathbf{x}) = \frac{\sum\limits_{l=1}^{M} \theta_l \prod\limits_{k=1}^{P} \underline{\mu}_{\tilde{F}_k^l}(x_k)}{\sum\limits_{l=1}^{M} \prod\limits_{k=1}^{P} \underline{\mu}_{\tilde{F}_k^l}(x_k)}, \qquad (2\text{-}40)$$

where $\overline{\mu}_{F_k^l}(x_k)$ and $\underline{\mu}_{F_k^l}(x_k)$ are the UMF and LMF associated with the k-th input of the l-th rule [35]. Moreover, the upper and lower fuzzy basis functions (FBFs) can be defined as

$$\overline{\phi}_l(\mathbf{x}) = \frac{\prod\limits_{k=1}^{P} \overline{\mu}_{\tilde{F}_k^l}(x_k)}{\sum\limits_{l=1}^{M} \prod\limits_{k=1}^{P} \overline{\mu}_{\tilde{F}_k^l}(x_k)} \qquad (2\text{-}41)$$

and

$$\underline{\phi}_l(\mathbf{x}) = \frac{\prod\limits_{k=1}^{P} \underline{\mu}_{\tilde{F}_k^l}(x_k)}{\sum\limits_{l=1}^{M} \prod\limits_{k=1}^{P} \underline{\mu}_{\tilde{F}_k^l}(x_k)}, \qquad (2\text{-}42)$$

respectively.

The difference between T1-FLS and T2-FLS (including IT2-FLS) is related to the nature of the MFs, which is not important when forming the fuzzy rules. In this sense, the structure of T2-FLS rules is the same as that of T1-FLS, where the fuzzy sets involved are of type-2 [5], expressed as

$$rule^l : \text{IF } x_1 \text{ is } \tilde{F}_1^l \text{ AND } \dots \text{ AND } x_P \text{ is } \tilde{F}_P^l \text{ THEN } y \text{ is } \tilde{Y}^l, \qquad (2\text{-}43)$$

where $\tilde{F}_P^l$ is the P-th type-2 fuzzy set associated with l-th rule. The structure of the rule "*IF* x_1 *is* $\tilde{F}_1^l$ *AND* ... ... *AND* x_P *is* $\tilde{F}_P^l$" corresponds to the antecedent from l-th rule and "*THEN* y *is* $\tilde{Y}^l$" is the consequent of the l-th rule.

The IT2-FLS UL output is computed from the mean of the upper and lower outputs, given by

$$f_{UL_{T2}}\left(\mathbf{x}^{(q)}\right) = \frac{Y_{U_{T2}}\left(\mathbf{x}^{(q)}\right) + Y_{L_{T2}}\left(\mathbf{x}^{(q)}\right)}{2}, \qquad (2\text{-}44)$$

and considering the same cost function presented in Equation (2-7), we have

---

[5]t-norm

$$J\left(\mathbf{w}^{(\gamma)}\right) = \frac{1}{2}\left[f_{UL_{T2}}\left(\mathbf{x}^{(q)}\right) - y^{(q)}\right]^2, \tag{2-45}$$

where $\mathbf{w}^{(\gamma)}$ represents the vector of the parameters for IT2-FLS UL in the γ-th iteration. Adopting GMFum for IT2-FLS UL (IT2-FLS UL-GMFum), the rules can be expressed in terms of GMFum parameters as follows

$$rule^l = \left\{x_1 \text{ is } \tilde{F}_1^l\left(m_{1\tilde{F}_1^l}, m_{2\tilde{F}_1^l}, \sigma_{\tilde{F}_1^l}\right) \text{ AND} \ldots \right.$$
$$\left. \ldots \text{ AND } x_P \text{ is } \tilde{F}_P^l\left(m_{1\tilde{F}_P^l}, m_{2\tilde{F}_1^l}, \sigma_{\tilde{F}_P^l}\right) \text{ THEN } \tilde{Y}^l\right\}. \tag{2-46}$$

Moreover, the vector of parameters for IT2-FLS UL-GMFum is given by

$$\mathbf{w}^{(\gamma)} = \left[m_{1\tilde{F}_1^1}\left(\gamma\right), \cdots, m_{1\tilde{F}_P^1}\left(\gamma\right), \cdots, m_{1\tilde{F}_1^M}\left(\gamma\right), \cdots, m_{1\tilde{F}_P^M}\left(\gamma\right), \cdots, \right.$$
$$m_{2\tilde{F}_1^1}\left(\gamma\right), \cdots, m_{2\tilde{F}_P^1}\left(\gamma\right), \cdots, m_{2\tilde{F}_1^M}\left(\gamma\right), \cdots, m_{2\tilde{F}_P^M}\left(\gamma\right), \cdots,$$
$$\sigma_{\tilde{F}_1^1}\left(\gamma\right), \cdots, \sigma_{\tilde{F}_P^1}\left(\gamma\right), \cdots, \sigma_{\tilde{F}_1^M}\left(\gamma\right), \cdots, \sigma_{\tilde{F}_P^M}\left(\gamma\right), \cdots, \tag{2-47}$$
$$\left. \theta_1\left(\gamma\right), \cdots, \theta_M\left(\gamma\right)\right]^T,$$

and its gradient vector is expressed as

$$\nabla\mathbf{J}\left(\mathbf{w}^{(\gamma)}\right) = \left[\frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial m_{1\tilde{F}_1^1}\left(\gamma\right)}, \cdots, \frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial m_{1\tilde{F}_P^1}\left(\gamma\right)}, \cdots, \frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial m_{1\tilde{F}_1^M}\left(\gamma\right)}, \cdots, \frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial m_{1\tilde{F}_P^M}\left(\gamma\right)}, \cdots, \right.$$
$$\frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial m_{2\tilde{F}_1^1}\left(\gamma\right)}, \cdots, \frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial m_{2\tilde{F}_P^1}\left(\gamma\right)}, \cdots, \frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial m_{2\tilde{F}_1^M}\left(\gamma\right)}, \cdots, \frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial m_{2\tilde{F}_P^M}\left(\gamma\right)}, \cdots,$$
$$\frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial \sigma_{\tilde{F}_1^1}\left(\gamma\right)}, \cdots, \frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial \sigma_{\tilde{F}_P^1}\left(\gamma\right)}, \cdots, \frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial \sigma_{\tilde{F}_1^M}\left(\gamma\right)}, \cdots, \frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial \sigma_{\tilde{F}_P^M}\left(\gamma\right)}, \cdots,$$
$$\left. \frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial \theta_1\left(\gamma\right)}, \cdots, \frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial \theta_M\left(\gamma\right)}\right]^T. \tag{2-48}$$

The equations for computing $\nabla\mathbf{J}\left(\mathbf{w}^{(\gamma)}\right)$ in Equation (2-48) were provided in [40] and are presented in Appendix A. Analyzing the domain pieces of UMF and LMF, and the equations of $\nabla\mathbf{J}\left(\mathbf{w}^{(\gamma)}\right)$, it can be seen that IT2-FLS UL-GMFum results in a significant increase of the computational costs in comparison with T1-FLS. This is because IT2-FLS UL-GMFum is piecewise differentiable, which results in one derivative equation for each MF interval.

3
Proposed interval type-2 fuzzy logic systems

As mentioned in the previous chapter, the IT2-FLS UL-GMFum results in higher computational costs than T1-FLS. In this way, to reduce the computational costs of the IT2-FLS UL-GMFum, this chapter addresses the work proposals. The first one is to provide the derivatives of the IT2-FLS UL using the GMFus. In addition, the second one is to extend the IT2-FLS using both GMFs to the IT2-FLSMO enabling to use them for solving MCPs without applying any decomposition strategy.

3.1
Gaussian membership function with uncertain standard deviation

The idea to propose the IT2-FLS UL-GMFus is to have an IT2-FLS UL differentiable in all domains, thereby reducing the equations of its first-order information. Therefore, applying Equations (2-37) and (2-38) for an input $\mathbf{x}^{(q)}$, the grades of UMF and LMF are expressed by

$$\overline{\mu}_{\tilde{F}_k^l}(x) = \exp\left[-\frac{1}{2}\left(\frac{x_k - m_{\tilde{F}_k^l}(\gamma)}{\sigma_{2\tilde{F}_k^l}(\gamma)}\right)^2\right] \tag{3-1}$$

and

$$\underline{\mu}_{\tilde{F}_k^l}(x) = \exp\left[-\frac{1}{2}\left(\frac{x_k - m_{\tilde{F}_k^l}(\gamma)}{\sigma_{1\tilde{F}_k^l}(\gamma)}\right)^2\right], \tag{3-2}$$

respectively. The replacement of the MF in IT2-FLS UL does not change the cost function (i.e., the cost function is already defined in Equation (2-45)). However, the vector $\mathbf{w}^{(\gamma)}$ is now described as

$$\begin{aligned}
\mathbf{w}^{(\gamma)} = \Big[&m_{\tilde{F}_1^1}(\gamma), \cdots, m_{\tilde{F}_P^1}(\gamma), \cdots, m_{\tilde{F}_1^M}(\gamma), \cdots, m_{\tilde{F}_P^M}(\gamma), \cdots, \\
&\sigma_{1\tilde{F}_1^1}(\gamma), \cdots, \sigma_{1\tilde{F}_P^1}(\gamma), \cdots, \sigma_{1\tilde{F}_1^M}(\gamma), \cdots, \sigma_{1\tilde{F}_P^M}(\gamma), \cdots, \\
&\sigma_{2\tilde{F}_1^1}(\gamma), \cdots, \sigma_{2\tilde{F}_P^1}(\gamma), \cdots, \sigma_{2\tilde{F}_1^M}(\gamma), \cdots, \sigma_{2\tilde{F}_P^M}(\gamma), \cdots, \\
&\qquad\qquad\qquad\qquad\qquad \theta_1(\gamma), \cdots, \theta_M(\gamma)\Big]^T.
\end{aligned} \tag{3-3}$$

Additionally, its gradient vector $\nabla \mathbf{J}\left(\mathbf{w}^{(\gamma)}\right)$ is expressed as

$$
\nabla \mathbf{J}\left(\mathbf{w}^{(\gamma)}\right) = \left[\frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial m_{\tilde{F}_1^1}(\gamma)}, \cdots, \frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial m_{\tilde{F}_P^1}(\gamma)}, \cdots, \frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial m_{\tilde{F}_1^M}(\gamma)}, \cdots, \frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial m_{\tilde{F}_P^M}(\gamma)}, \cdots, \right.
$$

$$
\frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial \sigma_{1\tilde{F}_1^1}(\gamma)}, \cdots, \frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial \sigma_{1\tilde{F}_P^1}(\gamma)}, \cdots, \frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial \sigma_{1\tilde{F}_1^M}(\gamma)}, \cdots, \frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial \sigma_{1\tilde{F}_P^M}(\gamma)}, \cdots,
$$

$$
\frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial \sigma_{2\tilde{F}_1^1}(\gamma)}, \cdots, \frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial \sigma_{2\tilde{F}_P^1}(\gamma)}, \cdots, \frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial \sigma_{2\tilde{F}_1^M}(\gamma)}, \cdots, \frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial \sigma_{2\tilde{F}_P^M}(\gamma)}, \cdots,
$$

$$
\left. \frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial \theta_1(\gamma)}, \cdots, \frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial \theta_M(\gamma)} \right]^T . \quad (3\text{-}4)
$$

The equations for computing the gradient $\nabla \mathbf{J}\left(\mathbf{w}^{(\gamma)}\right)$ are provided in Appendix B. As expected, the number of equations for obtaining the gradient of IT2-FLS UL-GMFus reduces from thirteen to four, reducing the computational costs compared to the model using IT2-FLS UL-GMFum.

Although the computational costs is reduced, IT2-FLS UL has only one output, which significantly increases the computational effort when dealing with MCP [43]. The FLS model needs to use a decomposition strategy for solving MCP, such as OvO or OvA. Moreover, it was shown in [43] that the number of rules required for solving MCP using an FLS with single output increases in a nonlinear manner according to the number of classes.

3.2
The interval type-2 fuzzy logic systems multiple outputs

Applying the concept of multiple outputs from T1-FLSMO to IT2-FLS, a new inference block is formed, as shown in Figure 3.1.
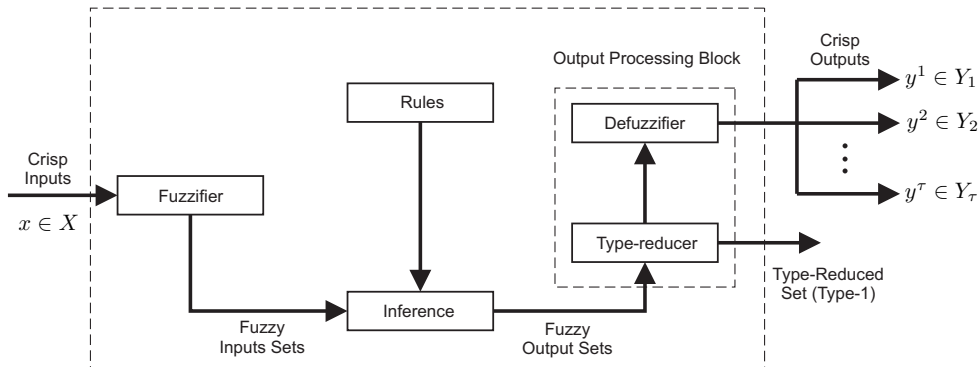
Figure 3.1: Structure of T2-FLSMO.

Considering IT2-FLS UL and extending it to multiple outputs, its output vector can be expressed as

$$\mathbf{f}_{ULT2_{mo}}^{(q)} = \frac{\mathbf{Y}_{U_{T2mo}}\left(\mathbf{x}^{(q)}\right) + \mathbf{Y}_{L_{T2mo}}\left(\mathbf{x}^{(q)}\right)}{2}, \tag{3-5}$$

where $\mathbf{Y}_{U_{T2mo}}\left(\mathbf{x}^{(q)}\right)$ represents the upper output vector and $\mathbf{Y}_{L_{T2mo}}\left(\mathbf{x}^{(q)}\right)$ represents the lower output vector, as follows

$$\mathbf{Y}_{U_{T2mo}}\left(\mathbf{x}^{(q)}\right) = \Theta\overline{\Phi}\left(\mathbf{x}^{(q)}\right) \tag{3-6}$$

and

$$\mathbf{Y}_{L_{T2mo}}\left(\mathbf{x}^{(q)}\right) = \Theta\underline{\Phi}\left(\mathbf{x}^{(q)}\right). \tag{3-7}$$

The Θ in Equations (3-6) and (3-7) is the same as that used in T1-FLSMO (see Equation (2-21)). Additionally, it is important to define the upper vector of FBFs $\overline{\Phi}\left(\mathbf{x}^{(q)}\right)$ and lower vector of FBFs $\underline{\Phi}\left(\mathbf{x}^{(q)}\right)$, as

$$\overline{\Phi}\left(\mathbf{x}^{(q)}\right) = \left[\begin{array}{cccc} \overline{\phi}_1\left(\mathbf{x}^{(q)}\right) & \overline{\phi}_2\left(\mathbf{x}^{(q)}\right) & \cdots & \overline{\phi}_M\left(\mathbf{x}^{(q)}\right) \end{array} \right]^T \tag{3-8}$$

and

$$\underline{\Phi}\left(\mathbf{x}^{(q)}\right) = \left[\begin{array}{cccc} \underline{\phi}_1\left(\mathbf{x}^{(q)}\right) & \underline{\phi}_2\left(\mathbf{x}^{(q)}\right) & \cdots & \underline{\phi}_M\left(\mathbf{x}^{(q)}\right) \end{array} \right]^T \tag{3-9}$$

respectively. In addition, extending the cost function presented in Equation (2-45), we obtain

$$J\left(\mathbf{w}^{(\gamma)}\right) = \frac{1}{2}\sum_{t=1}^{\tau}\left(f_{ULT2mo}^t\left(\mathbf{x}^{(q)}\right) - y^t\left(q\right)\right)^2, \tag{3-10}$$

where the vector of parameters $\mathbf{w}^{(\gamma)}$ for IT2-FLSMO UL-GMFum is expressed as

$$
\begin{aligned}
\mathbf{w}^{(\gamma)} = \Big[& m_{1\tilde{F}_1^1}\left(\gamma\right), \cdots, m_{1\tilde{F}_P^1}\left(\gamma\right), \cdots, m_{1\tilde{F}_1^M}\left(\gamma\right), \cdots, m_{1\tilde{F}_P^M}\left(\gamma\right), \cdots, \\
& m_{2\tilde{F}_1^1}\left(\gamma\right), \cdots, m_{2\tilde{F}_P^1}\left(\gamma\right), \cdots, m_{2\tilde{F}_1^M}\left(\gamma\right), \cdots, m_{2\tilde{F}_P^M}\left(\gamma\right), \cdots, \\
& \sigma_{\tilde{F}_1^1}\left(\gamma\right), \cdots, \sigma_{\tilde{F}_P^1}\left(\gamma\right), \cdots, \sigma_{\tilde{F}_1^M}\left(\gamma\right), \cdots, \sigma_{\tilde{F}_P^M}\left(\gamma\right), \cdots, \\
& \theta_1^1\left(\gamma\right), \cdots, \theta_M^1\left(\gamma\right), \cdots, \theta_1^\tau\left(\gamma\right), \cdots, \theta_M^\tau\left(\gamma\right) \Big]^T,
\end{aligned}
\tag{3-11}
$$

and that for IT2-FLSMO UL-GMFus as

$$
\begin{aligned}
\mathbf{w}^{(\gamma)} = \Big[& m_{\tilde{F}_1^1}\left(\gamma\right), \cdots, m_{\tilde{F}_P^1}\left(\gamma\right), \cdots, m_{\tilde{F}_1^M}\left(\gamma\right), \cdots, m_{\tilde{F}_P^M}\left(\gamma\right), \cdots, \\
& \sigma_{1\tilde{F}_1^1}\left(\gamma\right), \cdots, \sigma_{1\tilde{F}_P^1}\left(\gamma\right), \cdots, \sigma_{1\tilde{F}_1^M}\left(\gamma\right), \cdots, \sigma_{1\tilde{F}_P^M}\left(\gamma\right), \cdots, \\
& \sigma_{2\tilde{F}_1^1}\left(\gamma\right), \cdots, \sigma_{2\tilde{F}_P^1}\left(\gamma\right), \cdots, \sigma_{2\tilde{F}_1^M}\left(\gamma\right), \cdots, \sigma_{2\tilde{F}_P^M}\left(\gamma\right), \cdots, \\
& \theta_1^1\left(\gamma\right), \cdots, \theta_M^1\left(\gamma\right), \cdots, \theta_1^\tau\left(\gamma\right), \cdots, \theta_M^\tau\left(\gamma\right) \Big]^T.
\end{aligned}
\tag{3-12}
$$

Therefore, the expression of $\nabla\mathbf{J}\left(\mathbf{w}^{(\gamma)}\right)$ for IT2-FLSMO UL-GMFum and IT2-

FLSMO UL-GMFus are respectively given by

$$
\nabla \mathbf{J}\left(\mathbf{w}^{(\gamma)}\right) = \left[\frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial m_{1\tilde{F}_1^1}(\gamma)}, \cdots, \frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial m_{1\tilde{F}_P^1}(\gamma)}, \cdots, \frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial m_{1\tilde{F}_1^M}(\gamma)}, \cdots, \frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial m_{1\tilde{F}_P^M}(\gamma)}, \cdots, \right.
$$

$$
\frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial m_{2\tilde{F}_1^1}(\gamma)}, \cdots, \frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial m_{2\tilde{F}_P^1}(\gamma)}, \cdots, \frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial m_{2\tilde{F}_1^M}(\gamma)}, \cdots, \frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial m_{2\tilde{F}_P^M}(\gamma)}, \cdots,
$$

$$
\frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial \sigma_{\tilde{F}_1^1}(\gamma)}, \cdots, \frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial \sigma_{\tilde{F}_P^1}(\gamma)}, \cdots, \frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial \sigma_{\tilde{F}_1^M}(\gamma)}, \cdots, \frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial \sigma_{\tilde{F}_P^M}(\gamma)}, \cdots,
$$

$$
\left. \frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial \theta_1^1(\gamma)}, \cdots, \frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial \theta_M^1(\gamma)} \cdots \frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial \theta_1^\tau(\gamma)}, \cdots, \frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial \theta_M^\tau(\gamma)} \right]^T
$$

$$
(3\text{-}13)
$$

and

$$
\nabla \mathbf{J}\left(\mathbf{w}^{(\gamma)}\right) = \left[\frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial m_{\tilde{F}_1^1}(\gamma)}, \cdots, \frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial m_{\tilde{F}_P^1}(\gamma)}, \cdots, \frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial m_{\tilde{F}_1^M}(\gamma)}, \cdots, \frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial m_{\tilde{F}_P^M}(\gamma)}, \cdots, \right.
$$

$$
\frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial \sigma_{1\tilde{F}_1^1}(\gamma)}, \cdots, \frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial \sigma_{1\tilde{F}_P^1}(\gamma)}, \cdots, \frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial \sigma_{1\tilde{F}_1^M}(\gamma)}, \cdots, \frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial \sigma_{1\tilde{F}_P^M}(\gamma)}, \cdots,
$$

$$
\frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial \sigma_{2\tilde{F}_1^1}(\gamma)}, \cdots, \frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial \sigma_{2\tilde{F}_P^1}(\gamma)}, \cdots, \frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial \sigma_{2\tilde{F}_1^M}(\gamma)}, \cdots, \frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial \sigma_{2\tilde{F}_P^M}(\gamma)}, \cdots,
$$

$$
\left. \frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial \theta_1^1(\gamma)}, \cdots, \frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial \theta_M^1(\gamma)} \cdots \frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial \theta_1^\tau(\gamma)}, \cdots, \frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial \theta_M^\tau(\gamma)} \right]^T.
$$

$$
(3\text{-}14)
$$

The partial derivatives of the cost function $J\left(\mathbf{w}^{(\gamma)}\right)$ for each fuzzy parameter using GMFum are provided in Appendix C, and those using GMFus are provided in Appendix D. It is important to mention that for solving BCPs the number of rules in both methods (T1-FLS and T1-FLSMO) are the same, as T1-FLSMO turns into T1-FLS with a single output [43].

4
Experimental results

The performance analyses presented in this section consider the datasets provided by Knowledge Extraction based on Evolutionary Learning [46] and UCI Machine Learning Repository [47]. Information regarding these datasets is provided in Table 4.1, which includes nine datasets for BCPs and seven datasets for MCPs. To compare the proposed strategies, we implemented T1-FLS trained by the SD (SD T1-FLS) and SCG (SCG T1-FLS) methods [35]. In addition, for MCPs, we implemented T1-FLS using the OvA decomposition strategy (SD T1-FLS OvA and SCG T1-FLS OvA) and T1-FLSMO trained by the SD and SCG methods (indicated as SD T1-FLSMO and SCG T1-FLSMO, respectively) [43].

The classifier's performance presented herein considers 33 executions on each dataset, together with the 5-fold cross-validation [58] for all datasets containing less than 1000 samples. Regarding the remaining datasets, it was used 15% of the samples as a test set. Moreover, the validation set was performed using 30% of each training set, where $\mathbf{w}^{(\gamma)}$ is the parameter vector that attained the best accuracy metric for the validation set among all epochs. The input variables were normalized to a range of $[-1, 1]$, and the batch learning was applied [35]. The performance metrics used here were accuracy, MSE, Cohen's kappa [59], F-score [60], and best training epoch.

Table 4.1: Details of datasets.

Dataset	Input features	Number of samples								Total of samples
		Classes								
		1	2	3	4	5	6	7	8	
1. Appendicitis[1]	7	21	85	–	–	–	–	–	–	106
2. Balance Scale[2]	4	288	49	288	–	–	–	–	–	625
3. Car Evaluation[2]	6	1210	384	69	65	–	–	–	–	1728
4. Contraceptive[2]	9	629	333	511	–	–	–	–	–	1473
5. Ecoli[2]	7	143	77	2	2	35	20	5	52	336
6. Glass[2]	9	70	76	17	13	9	29	–	–	214
7. Haberman[1]	3	225	81	–	–	–	–	–	–	306
8. Ionosphere[2]	34	225	126	–	–	–	–	–	–	351
9. Iris[2]	4	50	50	50	–	–	–	–	–	150
10. Liver Disorders[2]	6	145	200	–	–	–	–	–	–	345
11. Monk2[1]	6	228	204	–	–	–	–	–	–	432
12. Parkinson[2]	22	147	48	–	–	–	–	–	–	195
13. Pima[2]	8	268	500	–	–	–	–	–	–	768
14. Sonar[2]	60	97	111	–	–	–	–	–	–	208
15. South Africa[1]	9	160	302	–	–	–	–	–	–	462
16. Wine[2]	13	59	71	48	–	–	–	–	–	178

The number of fuzzy rules was heuristically defined as follows: two rules for each class in FLSMO and each FLS OvA binary classifier leading to a total of $M = 2\Upsilon$ and $M = 2 \times 2\Upsilon$ rules, respectively. For T1-FLSMO, the first rule for each class was created using the mean and variance of all inputs in the entire training set for that class [35], which can be summarized as

$$
\begin{aligned}
rule_t^1 = \Big\{ x_1 \text{ is } F_1^1 \left(m_{F_1^1}, \sigma_{F_1^1} \right) \ AND \ldots \\
\ldots \ AND \ x_P \text{ is } F_P^1 \left(m_{F_P^1}, \sigma_{F_P^1} \right) \ THEN \ Y_t^1 \Big\},
\end{aligned}
\tag{4-1}
$$

where each rule '$rule_t^1$' contains the parameter values of the membership functions related to the first rule of the t-th class. The second rule for each class consists of the MF parameters from the first rule scaled by random values, $\{C \in \mathbb{R} | -0.8 \leq C \leq 1.2\}$ and $\{V \in \mathbb{R} | -0.8 \leq V \leq 1.2\}$, which can be expressed as

$$
\begin{aligned}
rule_t^2 = \Big\{ x_1 \text{ is } F_1^2 \left(Cm_{F_1^2}, V\sigma_{F_1^2} \right) \ AND \ldots \\
\ldots \ AND \ x_P \text{ is } F_P^2 \left(Cm_{F_P^2}, V\sigma_{F_P^2} \right) \ THEN \ Y_t^2 \Big\}.
\end{aligned}
\tag{4-2}
$$

The first rule of each class for IT2-FLSMO UL was created in a similar way. However, the mean and variance parameters of GMFum and GMFus were defined heuristically as 80% and 120% of the mean and variance of all inputs (i.e., in GMFum, the mean parameters were $m_{1\tilde{F}_k^l} = 0.8 m_{F_k^l}$ and $m_{2\tilde{F}_k^l} = 1.2 m_{F_k^l}$, and for GMFus, the variance parameters were $\sigma_{1\tilde{F}_k^l} = 0.8\sigma_{F_k^l}$ and $\sigma_{2\tilde{F}_k^l} = 1.2\sigma_{F_k^l}$), respectively. The first rules of GMFum and GMFus are respectively expressed as

$$
\begin{aligned}
rule_t^1 = \Big\{ x_1 \text{ is } \tilde{F}_1^1 \left(m_{1\tilde{F}_1^1}, m_{2\tilde{F}_1^1}, \sigma_{\tilde{F}_1^1} \right) \ AND \ldots \\
\ldots \ AND \ x_P \text{ is } \tilde{F}_P^1 \left(m_{1\tilde{F}_P^1}, m_{2\tilde{F}_P^1}, \sigma_{\tilde{F}_P^1} \right) \ THEN \ \tilde{Y}_t^1 \Big\}
\end{aligned}
\tag{4-3}
$$

and

$$
\begin{aligned}
rule_t^1 = \Big\{ x_1 \text{ is } \tilde{F}_1^1 \left(m_{\tilde{F}_1^1}, \sigma_{1\tilde{F}_1^1}, \sigma_{2\tilde{F}_1^1} \right) \ AND \ldots \\
\ldots \ AND \ x_P \text{ is } \tilde{F}_P^1 \left(m_{\tilde{F}_P^1}, \sigma_{1\tilde{F}_P^1}, \sigma_{2\tilde{F}_P^1} \right) \ THEN \ \tilde{Y}_t^1 \Big\}.
\end{aligned}
\tag{4-4}
$$

Regarding the second rule of each class in IT2-FLSMO UL, they were created by scaling the mean and variance parameters of the first rules by C and V, respectively, analogous to Equation (4-2).

However, the training samples related to the class for which each binary classifier was designed, must be used (e.g., an MCP with Υ classes in OvA decomposition has Υ BCs, where the first BC has two classes; the first class is designed using the samples of class $t = 1$ and the second class is designed

---

[1] Datasets provided by Knowledge Extraction based on Evolutionary Learning (KEEL) Repository [46]

[2] Datasets provided by UCI Machine Learning Repository [47]

using samples of classes $t = \{2, \ldots, \Upsilon\}$) [43]. Table 4.2 lists the total number of rules for the investigated approaches for each MCP dataset. The matrix Θ of the height defuzzifier is initialized with $\theta_l^t = 1$ if a rule corresponds to a class for which the output "t" is positive and with $\theta_l^t = \frac{-1}{\Upsilon - 1}$ for the remaining rules [43]. The output vector for each class was codified by a length-τ vector of negative unit values where the element that presents the class is replaced by 1 [43]. Moreover, the threshold assigns the output that is less than zero to class -1, while an output greater or equal to zero is assigned to class 1.

Table 4.2: Number of rules for each MCP dataset used in FLS with OvA and FLSMO.

Dataset	Total number of rules	
	FLSOvA	FLSMO
1. Balance Scale	12	6
2. Car Evaluation	16	8
3. Contraceptive	12	6
4. Ecoli	12	6
5. Glass	28	14
6. Iris	12	6
7. Wine	12	6

Two hundred epochs were considered for the training phase, and no stop criteria was set up. The step size adopted for the SD training method was initially defined as $\alpha = 0.01$. However, some preliminarily results, which are presented in Appendix E, show a slow convergence rate for IT2-FLS models considering only 200 epochs of training. This is because IT2-FLS has a better capability to model uncertainties than T1-FLS, leading to an optimization surface with less attenuations and consequently allowing one to increase the value of step size to achieve the convergence within a defined number of training epochs. Therefore, a new step size of $\alpha = 0.1$ was adopted for SD method. Considering the SCG training method, it was defined the same parameters used in [35], the initial value of the Lagrange coefficient was $\gamma_1 = 10^{-15}$, and the value for restarting the update direction was $N = 10$, with the infinitesimal increment set as $\epsilon = 10^{-5}$.

4.1
## Performance analysis for binary classification problems

Table 4.3 lists the performance metrics considering only BCPs for the proposed models and the T1-FLS model presented in [35]. Also, Figures 4.1 – 4.3 show the average of accuracy and MSE during the training phase for all 33 executions of the relevant datasets. In almost all the datasets, SCG T1-FLS achieved the highest training accuracy and the lowest best epoch.

This demonstrates the superiority of the SCG method, which uses second-order information, in comparison with the SD method, even when using IT2-FLS. However, regarding the test metrics, all models were found to achieve similar results, with the exception of the Ionosphere dataset, in which IT2-FLSs performed better than the other approaches. This improvement may be explained by the nature of uncertainties in this dataset, which is probably better handled using IT2-FLS than T1-FLS irrespective of the training method applied. In addition, the T1-FLSs models achieved better results than IT2-FLSs in the Liver dataset. The better results obtained using SCG T1-FLS can be explained due to the use of second-order information, which has more chances to avoid poor local minima than first-order information method. The performance of SD T1-FLS in this dataset can be explained by analyzing the MSE in Figure 4.2, in which the behavior of the curve becomes unstable after epoch 40. This behavior indicates that the step size adopted was large for SD T1-FLS, thus avoiding the local minimum many times and achieving good test accuracy due to the validation phase, once the fuzzy parameters of the best validation accuracy were used in the test phase. Moreover, it can be seen in Figures 4.1 – 4.3, that the performance of IT2-FLS GMFum is always slightly better than that of GMFus during the training phase.

Table 4.3: BCPs Performance comparison in terms of the mean and standard deviation.

Dataset	Method	Training accuracy [%]	Test accuracy [%]	Training MSE	Test MSE	Training Kappa	Test Kappa	Training F-score	Test F-score	Best epoch
Appendicitis	IT2-FLS UL-GMFum	92.49 (±2.47)	85.55 (±6.81)	0.25 (±0.07)	0.46 (±0.21)	0.65 (±0.13)	0.53 (±0.22)	0.82 (±0.07)	0.76 (±0.11)	100 (±75)
	IT2-FLS UL-GMFus	92.52 (±2.44)	85.32 (±6.90)	0.26 (±0.07)	0.46 (±0.19)	0.65 (±0.12)	0.54 (±0.21)	0.82 (±0.06)	0.77 (±0.11)	97 (±73)
	SD T1-FLS [35]	93.29 (±3.25)	84.79 (±6.44)	0.22 (±0.08)	0.51 (±0.21)	0.72 (±0.11)	0.49 (±0.21)	0.86 (±0.06)	0.74 (±0.11)	100 (±68)
	SCG T1-FLS [35]	95.57 (±2.11)	84.98 (±6.63)	0.15 (±0.07)	0.51 (±0.21)	0.84 (±0.10)	0.49 (±0.22)	0.92 (±0.05)	0.74 (±0.12)	39 (±58)
Haberman	IT2-FLS UL-GMFum	78.68 (±2.11)	74.22 (±3.58)	0.62 (±0.04)	0.74 (±0.07)	0.25 (±0.10)	0.15 (±0.13)	0.60 (±0.06)	0.55 (±0.08)	92 (±67)
	IT2-FLS UL-GMFus	78.53 (±2.14)	74.47 (±3.51)	0.63 (±0.04)	0.74 (±0.07)	0.24 (±0.10)	0.16 (±0.12)	0.60 (±0.06)	0.55 (±0.07)	88 (±72)
	SD T1-FLS [35]	78.38 (±2.24)	74.39 (±4.18)	0.62 (±0.04)	0.74 (±0.08)	0.29 (±0.08)	0.20 (±0.12)	0.63 (±0.04)	0.59 (±0.07)	67 (±63)
	SCG T1-FLS [35]	81.34 (±2.38)	74.06 (±3.82)	0.57 (±0.05)	0.75 (±0.09)	0.44 (±0.09)	0.19 (±0.12)	0.71 (±0.05)	0.58 (±0.07)	25 (±42)
Ionosphere	IT2-FLS UL-GMFum	98.16 (±1.28)	91.81 (±3.23)	0.09 (±0.03)	0.28 (±0.09)	0.95 (±0.04)	0.82 (±0.07)	0.97 (±0.02)	0.91 (±0.04)	168 (±40)
	IT2-FLS UL-GMFus	98.32 (±0.75)	93.03 (±2.73)	0.09 (±0.02)	0.23 (±0.08)	0.96 (±0.02)	0.85 (±0.06)	0.98 (±0.01)	0.92 (±0.03)	191 (±18)
	SD T1-FLS [35]	75.86 (±2.92)	73.66 (±4.43)	0.59 (±0.05)	0.65 (±0.06)	0.52 (±0.06)	0.49 (±0.09)	0.75 (±0.03)	0.73 (±0.05)	189 (±29)
	SCG T1-FLS [35]	95.81 (±4.83)	87.46 (±6.48)	0.13 (±0.12)	0.39 (±0.16)	0.91 (±0.10)	0.73 (±0.13)	0.95 (±0.05)	0.86 (±0.07)	91 (±60)
Liver	IT2-FLS UL-GMFum	73.23 (±3.43)	65.31 (±5.73)	0.73 (±0.06)	0.90 (±0.08)	0.38 (±0.12)	0.26 (±0.13)	0.68 (±0.07)	0.62 (±0.07)	152 (±60)
	IT2-FLS UL-GMFus	73.34 (±3.21)	66.34 (±5.71)	0.74 (±0.06)	0.87 (±0.07)	0.41 (±0.10)	0.28 (±0.13)	0.70 (±0.06)	0.63 (±0.07)	178 (±44)
	SD T1-FLS [35]	75.60 (±4.40)	69.21 (±5.05)	0.68 (±0.13)	0.85 (±0.09)	0.50 (±0.10)	0.35 (±0.11)	0.74 (±0.06)	0.67 (±0.06)	82 (±45)
	SCG T1-FLS [35]	85.97 (±2.12)	69.87 (±4.99)	0.47 (±0.05)	0.83 (±0.09)	0.70 (±0.05)	0.37 (±0.11)	0.85 (±0.02)	0.68 (±0.06)	25 (±21)
Monk2	IT2-FLS UL-GMFum	97.92 (±1.15)	97.65 (±1.67)	0.09 (±0.04)	0.10 (±0.06)	0.96 (±0.02)	0.95 (±0.03)	0.98 (±0.01)	0.98 (±0.02)	199 (±4)
	IT2-FLS UL-GMFus	98.34 (±1.18)	97.97 (±1.73)	0.08 (±0.03)	0.09 (±0.05)	0.97 (±0.02)	0.96 (±0.03)	0.98 (±0.01)	0.98 (±0.02)	199 (±4)
	SD T1-FLS [35]	97.73 (±0.98)	97.36 (±1.52)	0.10 (±0.03)	0.12 (±0.06)	0.95 (±0.02)	0.95 (±0.03)	0.98 (±0.01)	0.97 (±0.02)	199 (±6)
	SCG T1-FLS [35]	99.10 (±1.22)	98.74 (±1.83)	0.02 (±0.03)	0.03 (±0.04)	0.98 (±0.03)	0.97 (±0.04)	0.99 (±0.01)	0.99 (±0.02)	189 (±24)
Parkinson	IT2-FLS UL-GMFum	92.60 (±3.28)	84.32 (±6.08)	0.22 (±0.09)	0.45 (±0.15)	0.76 (±0.12)	0.57 (±0.17)	0.88 (±0.06)	0.78 (±0.09)	165 (±44)
	IT2-FLS UL-GMFus	93.28 (±3.64)	84.58 (±5.68)	0.20 (±0.08)	0.43 (±0.12)	0.78 (±0.12)	0.58 (±0.16)	0.89 (±0.06)	0.79 (±0.08)	170 (±44)
	SD T1-FLS [35]	94.43 (±3.47)	86.74 (±5.56)	0.20 (±0.10)	0.42 (±0.16)	0.80 (±0.10)	0.61 (±0.18)	0.90 (±0.05)	0.80 (±0.09)	134 (±52)
	SCG T1-FLS [35]	95.27 (±2.36)	85.80 (±5.88)	0.17 (±0.07)	0.45 (±0.17)	0.86 (±0.07)	0.58 (±0.18)	0.93 (±0.04)	0.78 (±0.09)	100 (±64)
Pima	IT2-FLS UL-GMFum	80.33 (±1.62)	76.41 (±3.17)	0.55 (±0.03)	0.64 (±0.06)	0.51 (±0.04)	0.46 (±0.07)	0.76 (±0.02)	0.73 (±0.04)	111 (±61)
	IT2-FLS UL-GMFus	80.14 (±1.50)	76.62 (±3.32)	0.56 (±0.03)	0.63 (±0.06)	0.52 (±0.04)	0.46 (±0.08)	0.76 (±0.02)	0.73 (±0.04)	136 (±50)
	SD T1-FLS [35]	81.90 (±1.41)	76.63 (±3.29)	0.51 (±0.03)	0.63 (±0.07)	0.53 (±0.04)	0.47 (±0.07)	0.77 (±0.02)	0.73 (±0.04)	56 (±51)
	SCG T1-FLS [35]	86.43 (±1.75)	76.56 (±3.20)	0.44 (±0.03)	0.64 (±0.08)	0.69 (±0.04)	0.47 (±0.07)	0.85 (±0.02)	0.73 (±0.04)	13 (±11)
Sonar	IT2-FLS UL-GMFum	98.55 (±1.20)	79.90 (±6.36)	0.06 (±0.04)	0.63 (±0.19)	0.89 (±0.10)	0.60 (±0.13)	0.94 (±0.05)	0.80 (±0.06)	69 (±58)
	IT2-FLS UL-GMFus	98.21 (±1.24)	80.03 (±6.60)	0.07 (±0.04)	0.62 (±0.19)	0.91 (±0.08)	0.60 (±0.13)	0.95 (±0.04)	0.80 (±0.07)	90 (±67)
	SD T1-FLS [35]	99.29 (±2.76)	79.78 (±6.19)	0.03 (±0.08)	0.62 (±0.19)	0.88 (±0.14)	0.59 (±0.13)	0.94 (±0.08)	0.79 (±0.06)	44 (±55)
	SCG T1-FLS [35]	99.69 (±0.51)	79.72 (±6.57)	0.01 (±0.02)	0.63 (±0.21)	0.99 (±0.01)	0.59 (±0.13)	1.00 (±0.01)	0.79 (±0.07)	34 (±52)
South Africa	IT2-FLS UL-GMFum	78.13 (±1.97)	70.27 (±3.90)	0.62 (±0.04)	0.77 (±0.08)	0.43 (±0.07)	0.32 (±0.09)	0.71 (±0.04)	0.65 (±0.05)	103 (±62)
	IT2-FLS UL-GMFus	78.20 (±2.17)	70.67 (±4.26)	0.61 (±0.04)	0.76 (±0.08)	0.43 (±0.07)	0.33 (±0.10)	0.71 (±0.03)	0.66 (±0.05)	109 (±61)
	SD T1-FLS [35]	80.59 (±2.12)	71.08 (±3.96)	0.56 (±0.04)	0.76 (±0.08)	0.46 (±0.05)	0.34 (±0.09)	0.73 (±0.03)	0.67 (±0.05)	69 (±55)
	SCG T1-FLS [35]	88.14 (±2.15)	71.15 (±3.96)	0.42 (±0.05)	0.77 (±0.09)	0.72 (±0.05)	0.34 (±0.09)	0.86 (±0.03)	0.67 (±0.05)	12 (±15)
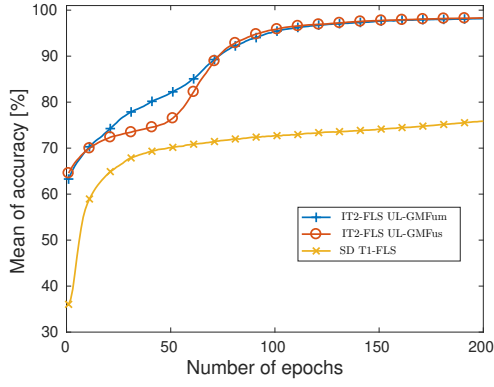
4.1(a): Accuracy.
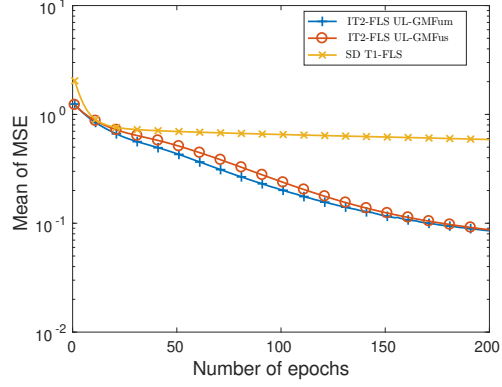
4.1(b): MSE.

Figure 4.1: Ionosphere dataset: average performance of FLSs.

4.2(a): Accuracy.

4.2(b): MSE.

Figure 4.2: Liver dataset: average performance of FLSs.

4.3(a): Accuracy.

4.3(b): MSE.

Figure 4.3: Parkinson dataset: average performance of FLSs.

4.2
Performance analysis for multiclass classification problems

Table 4.4 lists the performance metrics for the MCPs of the proposed models as well as the T1-FLS, and T1-FLSMO models presented in [43]. Figures 4.4 – 4.6 show the average of accuracy and MSE during the training phase for all executions of the most relevant datasets. In addition, it should be noted that the best epoch metric was not computed for all approaches because there can be no fair comparison between the average best epoch of the Υ classifiers required for OvA decomposition strategies and the best epoch achieved by the FLSMO models, which uses only one classifier. With respect to the mean and standard deviation of the test metrics, the models using OvA and FLSMO approaches achieve the same performance for Balance, Contraceptive, Ecoli, Iris, and Wine datasets. Regarding the results for the Car dataset, the FLSMO approaches were slightly better than FLS with OvA models, and the T1-FLSMO trained by SCG attained the best results. On the other hand, in the Glass dataset, the FLS models using OvA strategy were better than FLSMO models, except for SCG T1-FLSMO, which achieved similar results for the test metrics. However, the FLSs using the OvA approach required twice the number of fuzzy rules in comparison with FLSMO, thereby increasing the computational costs during the training phase. Moreover, the training MSE of IT2-FLSMO models always achieved the lowest values in all datasets.
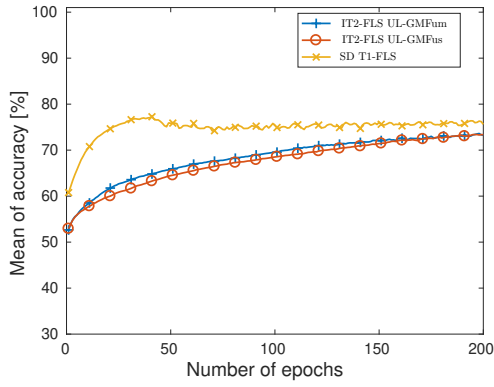

| 4.4(a): Accuracy. | 4.4(b): MSE. |

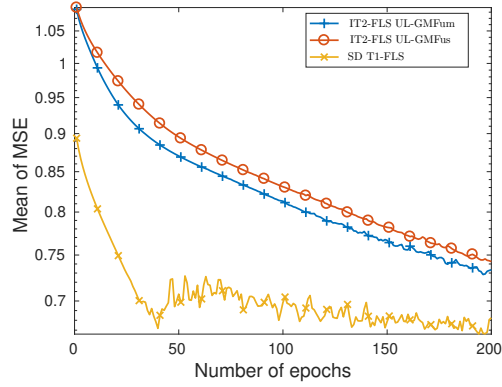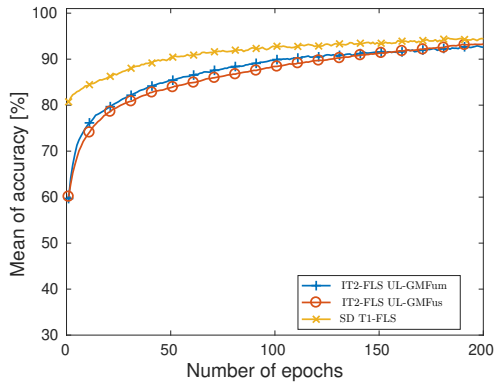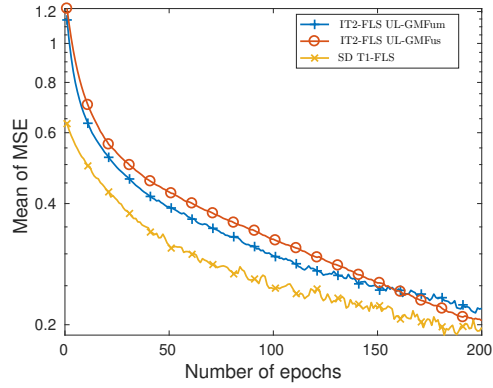Figure 4.4: Balance dataset: average performance of FLSs.

4.5(a): Accuracy.

4.5(b): MSE.

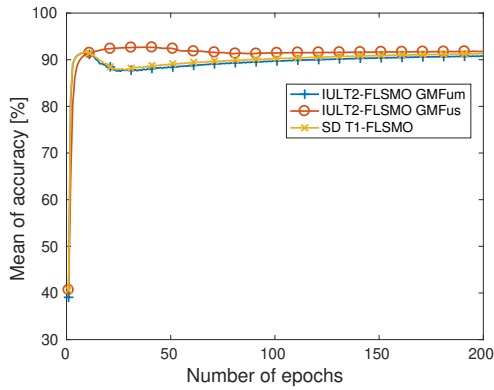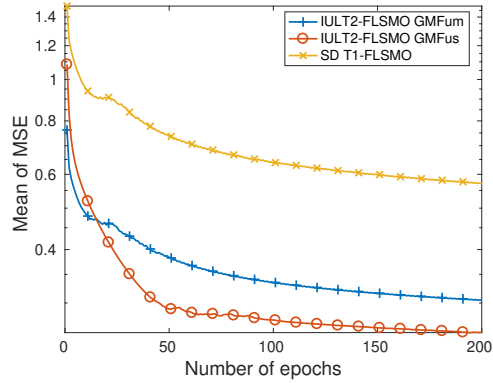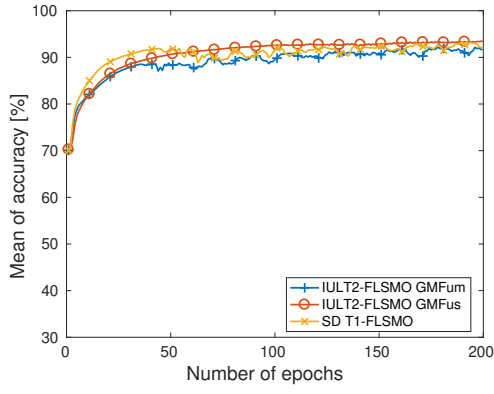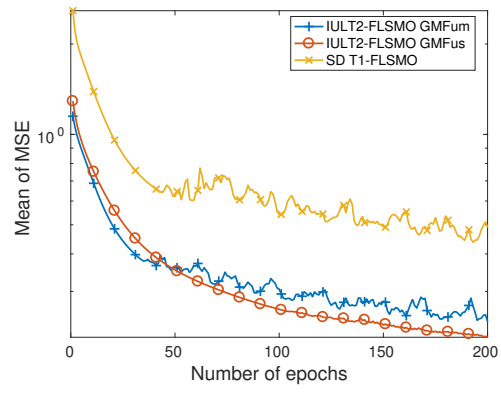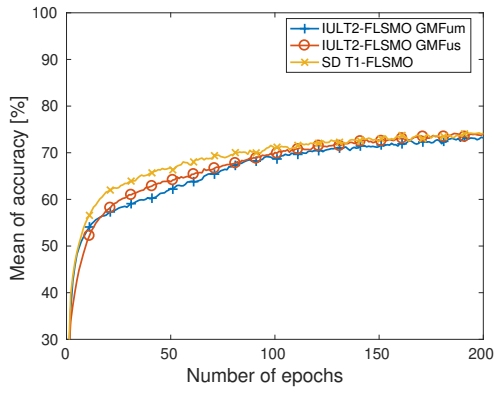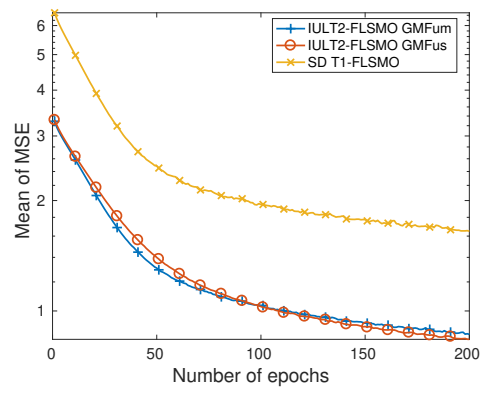Figure 4.5: Car dataset: average performance of FLSs.

4.6(a): Accuracy.

4.6(b): MSE.

Figure 4.6: Glass dataset: average performance of FLSs.

Table 4.4: MCPs performance comparison in terms of the mean and standard deviation.

Dataset	Method	Training accuracy [%]	Test accuracy [%]	Training MSE	Test MSE	Training Kappa	Test Kappa	Training F-score	Test F-score	Best epoch
Balance	IT2-FLS UL-GMFum OvA	91.48 (±1.41)	91.84 (±0.82)	0.57 (±0.05)	0.51 (±0.04)	0.84 (±0.03)	0.85 (±0.02)	0.66 (±0.05)	0.65 (±0.03)	–
	IT2-FLS UL-GMFus OvA	91.97 (±0.80)	91.70 (±0.77)	0.49 (±0.03)	0.53 (±0.03)	0.85 (±0.01)	0.85 (±0.01)	0.65 (±0.02)	0.64 (±0.02)	–
	SD T1-FLS OvA [43]	91.75 (±1.49)	91.90 (±0.93)	0.55 (±0.04)	0.50 (±0.04)	0.85 (±0.03)	0.85 (±0.02)	0.67 (±0.06)	0.66 (±0.04)	–
	IT2-FLSMO UL-GMFum	90.73 (±0.97)	89.47 (±1.88)	0.30 (±0.02)	0.70 (±0.07)	0.83 (±0.02)	0.81 (±0.03)	0.69 (±0.03)	0.64 (±0.03)	191 (±29)
	IT2-FLSMO UL-GMFus	91.76 (±0.95)	90.83 (±1.56)	0.26 (±0.02)	0.60 (±0.06)	0.86 (±0.02)	0.83 (±0.03)	0.72 (±0.04)	0.66 (±0.04)	111 (±55)
	SD T1-FLSMO [43]	91.29 (±1.03)	89.18 (±1.90)	0.57 (±0.05)	0.71 (±0.07)	0.84 (±0.02)	0.80 (±0.03)	0.71 (±0.04)	0.64 (±0.03)	188 (±27)
	SCG T1-FLSMO [43]	92.88 (±1.17)	91.01 (±1.22)	0.43 (±0.06)	0.54 (±0.08)	0.86 (±0.02)	0.83 (±0.02)	0.69 (±0.06)	0.64 (±0.02)	138 (±56)
Car	IT2-FLS UL-GMFum OvA	89.93 (±0.63)	88.18 (±1.83)	0.70 (±0.02)	0.74 (±0.08)	0.77 (±0.02)	0.73 (±0.04)	0.73 (±0.04)	0.67 (±0.07)	–
	IT2-FLS UL-GMFus OvA	90.00 (±0.61)	88.58 (±2.14)	0.69 (±0.02)	0.74 (±0.08)	0.78 (±0.01)	0.74 (±0.04)	0.73 (±0.04)	0.69 (±0.06)	–
	SD T1-FLS OvA [43]	91.41 (±0.66)	89.26 (±2.12)	0.65 (±0.02)	0.71 (±0.08)	0.81 (±0.02)	0.76 (±0.04)	0.77 (±0.04)	0.71 (±0.06)	–
	IT2-FLSMO UL-GMFum	91.78 (±3.18)	90.80 (±1.87)	0.24 (±0.04)	0.51 (±0.07)	0.85 (±0.02)	0.80 (±0.04)	0.81 (±0.06)	0.75 (±0.08)	191 (±11)
	IT2-FLSMO UL-GMFus	93.38 (±1.01)	91.73 (±1.78)	0.21 (±0.02)	0.49 (±0.07)	0.86 (±0.02)	0.82 (±0.04)	0.86 (±0.04)	0.80 (±0.06)	196 (±7)
	SD T1-FLSMO [43]	91.85 (±3.98)	91.80 (±2.07)	0.49 (±0.16)	0.48 (±0.08)	0.86 (±0.02)	0.82 (±0.04)	0.87 (±0.03)	0.82 (±0.06)	194 (±7)
	SCG T1-FLSMO [43]	97.05 (±1.09)	94.89 (±1.69)	0.23 (±0.06)	0.33 (±0.09)	0.93 (±0.03)	0.89 (±0.03)	0.95 (±0.02)	0.90 (±0.05)	183 (±28)
Contraceptive	IT2-FLS UL-GMFum OvA	58.79 (±1.22)	55.09 (±2.61)	2.11 (±0.03)	2.24 (±0.07)	0.35 (±0.02)	0.30 (±0.04)	0.56 (±0.01)	0.52 (±0.03)	–
	IT2-FLS UL-GMFus OvA	58.59 (±1.19)	55.07 (±2.70)	2.13 (±0.03)	2.25 (±0.06)	0.35 (±0.02)	0.30 (±0.04)	0.56 (±0.01)	0.52 (±0.03)	–
	SD T1-FLS OvA [43]	59.52 (±1.33)	55.42 (±2.23)	2.08 (±0.04)	2.24 (±0.07)	0.36 (±0.02)	0.30 (±0.03)	0.57 (±0.02)	0.52 (±0.02)	–
	IT2-FLSMO UL-GMFum	57.86 (±1.37)	55.08 (±3.17)	1.07 (±0.02)	2.27 (±0.09)	0.33 (±0.02)	0.30 (±0.05)	0.54 (±0.02)	0.52 (±0.03)	169 (±43)
	IT2-FLSMO UL-GMFus	59.21 (±1.01)	54.55 (±3.08)	1.04 (±0.02)	2.29 (±0.09)	0.35 (±0.02)	0.29 (±0.05)	0.56 (±0.02)	0.52 (±0.03)	168 (±35)
	SD T1-FLSMO [43]	59.38 (±1.37)	54.76 (±3.29)	2.12 (±0.04)	2.29 (±0.08)	0.35 (±0.03)	0.29 (±0.05)	0.56 (±0.02)	0.52 (±0.04)	145 (±46)
	SCG T1-FLSMO [43]	62.11 (±2.39)	54.52 (±3.16)	2.02 (±0.06)	2.30 (±0.11)	0.35 (±0.04)	0.29 (±0.05)	0.56 (±0.04)	0.51 (±0.04)	49 (±39)
Ecoli	IT2-FLS UL-GMFum OvA	90.46 (±1.61)	86.45 (±3.46)	0.69 (±0.08)	1.02 (±0.19)	0.87 (±0.02)	0.81 (±0.05)	0.75 (±0.08)	0.72 (±0.09)	–
	IT2-FLS UL-GMFus OvA	89.47 (±1.76)	86.28 (±3.49)	0.81 (±0.09)	1.06 (±0.19)	0.85 (±0.03)	0.81 (±0.05)	0.73 (±0.08)	0.72 (±0.09)	–
	SD T1-FLS OvA [43]	90.82 (±1.62)	86.28 (±3.50)	0.67 (±0.09)	1.03 (±0.18)	0.87 (±0.02)	0.81 (±0.05)	0.76 (±0.09)	0.72 (±0.09)	–
	IT2-FLSMO UL-GMFum	90.11 (±2.02)	85.10 (±3.47)	0.45 (±0.07)	1.17 (±0.18)	0.86 (±0.03)	0.79 (±0.05)	0.74 (±0.05)	0.63 (±0.07)	195 (±9)
	IT2-FLSMO UL-GMFus	89.84 (±2.17)	85.48 (±3.55)	0.48 (±0.09)	1.16 (±0.18)	0.86 (±0.03)	0.80 (±0.05)	0.74 (±0.06)	0.64 (±0.07)	198 (±3)
	SD T1-FLSMO [43]	90.74 (±1.83)	85.36 (±3.85)	0.85 (±0.13)	1.16 (±0.16)	0.87 (±0.03)	0.80 (±0.05)	0.75 (±0.05)	0.64 (±0.06)	193 (±12)
	SCG T1-FLSMO [43]	93.21 (±2.49)	84.72 (±4.14)	0.52 (±0.13)	1.20 (±0.57)	0.87 (±0.03)	0.79 (±0.06)	0.76 (±0.05)	0.60 (±0.02)	66 (±48)
Glass	IT2-FLS UL-GMFum OvA	77.04 (±2.95)	65.08 (±6.71)	1.47 (±0.12)	2.17 (±0.26)	0.68 (±0.04)	0.51 (±0.09)	0.69 (±0.06)	0.52 (±0.10)	–
	IT2-FLS UL-GMFus OvA	75.95 (±2.97)	65.10 (±6.93)	1.54 (±0.11)	2.16 (±0.25)	0.67 (±0.04)	0.52 (±0.09)	0.67 (±0.06)	0.53 (±0.10)	–
	SD T1-FLS OvA [43]	77.73 (±3.65)	65.84 (±6.95)	1.41 (±0.15)	2.15 (±0.27)	0.69 (±0.05)	0.52 (±0.10)	0.71 (±0.06)	0.54 (±0.09)	–
	IT2-FLSMO UL-GMFum	73.19 (±5.28)	62.56 (±7.59)	0.86 (±0.11)	2.20 (±0.29)	0.62 (±0.07)	0.47 (±0.11)	0.59 (±0.07)	0.42 (±0.08)	182 (±25)
	IT2-FLSMO UL-GMFus	74.07 (±3.94)	61.61 (±6.61)	0.84 (±0.07)	2.23 (±0.25)	0.63 (±0.06)	0.46 (±0.09)	0.59 (±0.04)	0.41 (±0.08)	186 (±19)
	SD T1-FLSMO [43]	74.17 (±4.44)	62.60 (±6.86)	1.65 (±0.15)	2.17 (±0.27)	0.63 (±0.07)	0.47 (±0.10)	0.60 (±0.05)	0.43 (±0.08)	178 (±25)
	SCG T1-FLSMO [43]	82.04 (±4.89)	64.45 (±6.64)	1.25 (±0.21)	2.27 (±0.41)	0.68 (±0.08)	0.50 (±0.09)	0.60 (±0.04)	0.42 (±0.04)	79 (±58)
Iris	IT2-FLS UL-GMFum OvA	97.80 (±1.28)	94.97 (±4.03)	0.18 (±0.05)	0.37 (±0.23)	0.97 (±0.02)	0.92 (±0.06)	0.98 (±0.01)	0.95 (±0.04)	–
	IT2-FLS UL-GMFus OvA	97.54 (±1.25)	95.23 (±3.74)	0.18 (±0.05)	0.34 (±0.20)	0.96 (±0.02)	0.93 (±0.06)	0.98 (±0.01)	0.95 (±0.04)	–
	SD T1-FLS OvA [43]	97.94 (±1.46)	95.11 (±3.90)	0.16 (±0.08)	0.36 (±0.22)	0.97 (±0.02)	0.93 (±0.06)	0.98 (±0.01)	0.95 (±0.04)	–
	IT2-FLSMO UL-GMFum	97.84 (±3.50)	94.97 (±3.94)	0.07 (±0.11)	0.38 (±0.22)	0.95 (±0.05)	0.92 (±0.06)	0.97 (±0.04)	0.95 (±0.04)	144 (±46)
	IT2-FLSMO UL-GMFus	98.87 (±1.13)	95.35 (±3.73)	0.03 (±0.02)	0.34 (±0.30)	0.97 (±0.03)	0.93 (±0.06)	0.98 (±0.02)	0.95 (±0.04)	144 (±47)
	SD T1-FLSMO [43]	97.92 (±3.26)	95.01 (±3.77)	0.14 (±0.17)	0.33 (±0.25)	0.96 (±0.05)	0.93 (±0.06)	0.97 (±0.04)	0.95 (±0.04)	137 (±47)
	SCG T1-FLSMO [43]	99.52 (±0.74)	94.79 (±3.97)	0.04 (±0.05)	0.44 (±0.32)	0.97 (±0.03)	0.92 (±0.06)	0.98 (±0.02)	0.95 (±0.04)	59 (±64)
Wine	IT2-FLS UL-GMFum OvA	99.64 (±0.53)	95.84 (±3.30)	0.04 (±0.03)	0.36 (±0.17)	0.99 (±0.01)	0.94 (±0.05)	1.00 (±0.01)	0.96 (±0.03)	–
	IT2-FLS UL-GMFus OvA	99.59 (±0.55)	96.15 (±3.17)	0.05 (±0.03)	0.33 (±0.17)	0.99 (±0.01)	0.94 (±0.05)	1.00 (±0.01)	0.96 (±0.03)	–
	SD T1-FLS OvA [43]	99.76 (±0.47)	95.95 (±3.29)	0.03 (±0.03)	0.35 (±0.18)	1.00 (±0.01)	0.94 (±0.05)	1.00 (±0.00)	0.96 (±0.03)	–
	IT2-FLSMO UL-GMFum	99.56 (±0.67)	94.60 (±3.98)	0.02 (±0.02)	0.35 (±0.28)	0.98 (±0.05)	0.92 (±0.06)	0.99 (±0.01)	0.95 (±0.04)	96 (±62)
	IT2-FLSMO UL-GMFus	99.24 (±0.79)	94.11 (±3.57)	0.03 (±0.03)	0.40 (±0.23)	0.98 (±0.02)	0.91 (±0.05)	0.99 (±0.01)	0.94 (±0.04)	114 (±61)
	SD T1-FLSMO [43]	99.59 (±0.66)	94.99 (±3.52)	0.03 (±0.05)	0.33 (±0.23)	0.99 (±0.01)	0.92 (±0.05)	0.99 (±0.01)	0.95 (±0.04)	109 (±59)
	SCG T1-FLSMO [43]	99.05 (±1.43)	93.80 (±4.50)	0.07 (±0.10)	0.46 (±0.30)	0.98 (±0.03)	0.91 (±0.07)	0.98 (±0.02)	0.94 (±0.05)	65 (±57)

4.3
Statistical analysis

To evaluate the results presented in Tables 4.3 and 4.4, a two-sample t-test was performed on the test accuracy metric. The two-sample t-test can infer a statistical comparison from two independent data samples [35]. This statistical test is expressed as

$$T = \frac{\overline{\mathcal{G}}_1 - \overline{\mathcal{G}}_2}{\sqrt{\frac{\sigma_{\mathcal{G}_1}^2}{L_{\mathcal{G}_1}} + \frac{\sigma_{\mathcal{G}_2}^2}{L_{\mathcal{G}_2}}}}, \qquad (4\text{-}5)$$

where $\overline{\mathcal{G}}_1$, $\overline{\mathcal{G}}_2$, $\sigma_{\mathcal{G}_1}^2$ and $\sigma_{\mathcal{G}_2}^2$ are the mean and standard deviation values of samples belonging to $\mathcal{G}_1$ and $\mathcal{G}_2$, respectively. In addition, $L_{\mathcal{G}_1} = \#\{\mathcal{G}_1\}$ and $L_{\mathcal{G}_2} = \#\{\mathcal{G}_2\}$, where $\#$ denotes the cardinality operator. The degree of freedom is defined as $L_{\mathcal{G}_1} + L_{\mathcal{G}_2} - 2$. In addition to determining T, it becomes important to infer the following hypotheses

$$\begin{cases} \mathcal{H}_0 : \overline{\mathcal{G}}_1 = \overline{\mathcal{G}}_2 \\ \mathcal{H}_1 : \overline{\mathcal{G}}_1 \neq \overline{\mathcal{G}}_2 \end{cases}. \qquad (4\text{-}6)$$

Given a significance level ω (typically around 0.05), the p-value is calculated from T and it represents the lowest value of ω required to non reject the null hypothesis ($\mathcal{H}_0$). Therefore, p-values below ω indicates that the null hypothesis is rejected [61].

The statistical analyses performed herein were obtained from 33 experiments, having 64 degrees of freedom for datasets with less than 1000 samples and 328 degrees of freedom for datasets requiring 5-fold cross-validation, thereby allowing us to avoid verifying the normality of error distributions [61]. Tables 4.5 and 4.6 list the statistical analysis results, which compare the FLS and FLSMO models for IT2 using GMFus, IT2 using GMFum, and T1 trained by the SCG method [35, 43]. The rejection of the null hypothesis is expressed by the letters "W" and "L", which indicate wins and losses for the tested models, respectively. The non-rejection of the null hypothesis is expressed by the letter "E" which indicates equality between the tested models.

Table 4.5: Statistical analyses performed by two-sample t-test for the test accuracy metric considering BCPs.

Data set	*Set of sample$_1$*	*Set of sample$_2$*	p-value	Lower boundary	Upper boundary	Null hypothesis
Appendicitis	IT2-FLS UL-GMFus	IT2-FLS UL-GMFum	0.76	-1.72	1.25	E
		SCG T1-FLS	0.65	-1.13	1.80	E
Haberman	IT2-FLS UL-GMFus	IT2-FLS UL-GMFum	0.53	-0.52	1.01	E
		SCG T1-FLS	0.31	-0.38	1.21	E
Ionosphere	IT2-FLS UL-GMFus	IT2-FLS UL-GMFum	0.00	0.58	1.87	W
		SCG T1-FLS	0.00	4.50	6.65	W
Liver	IT2-FLS UL-GMFus	IT2-FLS UL-GMFum	0.10	-0.20	2.27	E
		SCG T1-FLS	0.00	-4.69	-2.37	L
Monk2	IT2-FLS UL-GMFus	IT2-FLS UL-GMFum	0.09	-0.05	0.69	E
		SCG T1-FLS	0.00	-1.15	-0.38	L
Parkinson	IT2-FLS UL-GMFus	IT2-FLS UL-GMFum	0.68	-1.01	1.54	E
		SCG T1-FLS	0.06	-2.46	0.04	E
Pima	IT2-FLS UL-GMFus	IT2-FLS UL-GMFum	0.56	-0.49	0.91	E
		SCG T1-FLS	0.86	-0.64	0.77	E
Sonar	IT2-FLS UL-GMFus	IT2-FLS UL-GMFum	0.85	-1.27	1.54	E
		SCG T1-FLS	0.67	-1.12	1.73	E
South Africa	IT2-FLS UL-GMFus	IT2-FLS UL-GMFum	0.37	-0.48	1.29	E
		SCG T1-FLS	0.30	-1.36	0.42	E

The results of the statistical test show that in 13 datasets, the IT2 fuzzy models (namely FLS and FLSMO) using GMFus achieved the same performance as that of IT2 using GMFum, whereas in three datasets, the IT2 models using GMFus performed better than those using GMFum. Additionally, the IT2 fuzzy models using GMFus obtained the same results as those of the SCG T1 fuzzy models in 11 datasets, presented a better performance in the Ionosphere dataset, and the worse performance in the remaining four datasets. These results show that the IT2 fuzzy models can obtain the same performance as that of a T1-FLS trained by a method that uses second-order information, even attaining better results in some cases.

Table 4.6: Statistical analyses performed by two-sample t-test for the test accuracy metric considering MCPs.

Data set	Set of sample$_1$	Set of sample$_2$	p-value	Lower boundary	Upper boundary	Null hypothesis
Balance	IT2-FLSMO UL-GMFus	IT2-FLSMO UL-GMFum	0.00	0.99	1.74	W
		SCG T1-FLSMO	0.25	-0.48	0.12	E
Car	IT2-FLSMO UL-GMFus	IT2-FLSMO UL-GMFum	0.04	0.03	1.82	W
		SCG T1-FLSMO	0.00	-4.01	-2.31	L
Contraceptive	IT2-FLSMO UL-GMFus	IT2-FLSMO UL-GMFum	0.49	-2.07	1.00	E
		SCG T1-FLSMO	0.97	-1.51	1.56	E
Ecoli	IT2-FLSMO UL-GMFus	IT2-FLSMO UL-GMFum	0.33	-0.38	1.14	E
		SCG T1-FLSMO	0.08	-0.08	1.59	E
Glass	IT2-FLSMO UL-GMFus	IT2-FLSMO UL-GMFum	0.22	-2.50	0.59	E
		SCG T1-FLSMO	0.00	-4.28	-1.41	L
Iris	IT2-FLSMO UL-GMFus	IT2-FLSMO UL-GMFum	0.36	-0.45	1.21	E
		SCG T1-FLSMO	0.18	-0.27	1.40	E
Wine	IT2-FLSMO UL-GMFus	IT2-FLSMO UL-GMFum	0.24	-1.31	0.32	E
		SCG T1-FLSMO	0.50	-0.58	1.18	E

4.4
Computational costs analysis

The computational costs was analyzed by computing the time consumed during the training phase without the validation set using the SD method. The following five datasets were selected: Haberman, Monk2, South Africa, Ionosphere, and Sonar datasets having 3, 6, 9, 34, and 60 input features, respectively. Moreover, each analysis was performed 33 times, during which 100 samples were selected randomly from each dataset normalized between $[-1, 1]$, considering 200 epochs of training. The analyses were performed in MATLAB 2019a environment on a computer with Intel Core i7-2630QM 2.00 GHz CPU, 6144MB DDR3 666.7 MHz RAM, and 64-bit Windows 10 operating system.

Table 4.7 lists the mean and standard deviation for each investigated dataset, and Figure 4.7 depicts the corresponding graphical results. Additionally, it should be noted that each dataset is listed in Table 4.7 according to the number of their input features, since they have been normalized to have the same number of samples and training epochs. Based on the results, we can state that the proposed method, namely IT2-FLS UL-GMFus, is up to 7 times faster than IT2-FLS UL-GMFum and slightly slower than T1-FLS. This is because IT2-FLS ULGMFum requires more equations than IT2-FLS UL-GMFus to perform fuzzification, and consequently to compute the gradient vector of the cost function. This result can be extended to FLSMO models since its fuzzification procedure is the same as that in FLS.

Table 4.7: Time performance comparison in terms of the mean and standard deviation.

Number of features	Time consumed [s]		
	IT2-FLS UL-GMFum	IT2-FLS UL-GMFus	T1-FLS
3	12.84 ($\pm$ 0.31)	1.66 ($\pm$ 0.06)	1.69 ($\pm$ 0.07)
6	12.99 ($\pm$ 0.28)	1.87 ($\pm$ 0.05)	1.78 ($\pm$ 0.04)
9	13.56 ($\pm$ 0.34)	2.02 ($\pm$ 0.04)	1.90 ($\pm$ 0.09)
34	15.29 ($\pm$ 0.24)	3.43 ($\pm$ 0.11)	2.60 ($\pm$ 0.09)
60	17.48 ($\pm$ 0.24)	4.77 ($\pm$ 0.07)	3.41 ($\pm$ 0.06)
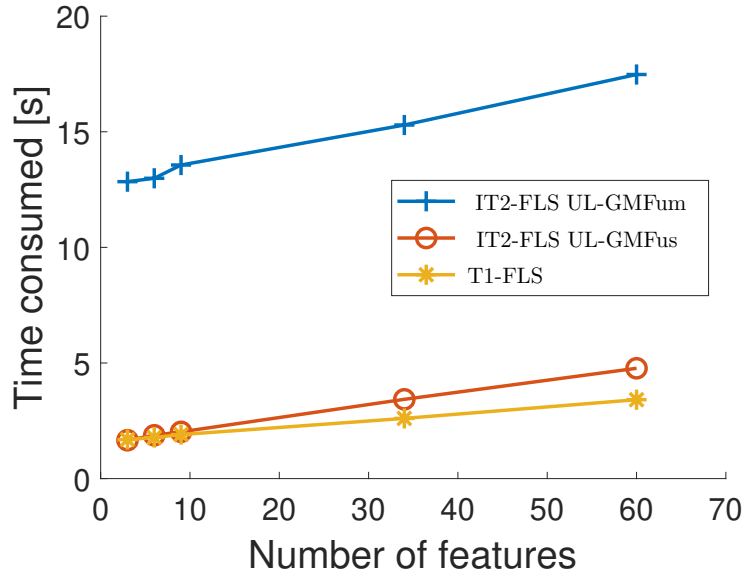
Figure 4.7: Time consumed during the training phase for IT2-FLS approaches and T1-FLS considering datasets with different numbers of input features.

All the models proposed herein used benchmark datasets for their validation, and, therefore, the that dataset was not explained in detail. In this way, aiming to contextualize an practical engineering application of the proposed models the next chapter attempts to solve a detection problem of fault in aircraft gas turbines.

4.5
Practical application: aircraft gas turbine fault detection

The gas turbine (GT) is a combustion engine that converts natural gas or other liquid fuels to mechanical energy. It has found increasing service in the last decades in the power industry, due to its compactness, lightweight and multiple fuel application [62]. Nowadays, there are GTs which run on natural gas, diesel fuel, naphtha, methane, crude, low-BTU gases, vaporized fuel oils, and biomass gases [62]. The above-mentioned characteristics also turn out the

GTs are the most common engine used in aircraft. Accordingly to AIRBUS global market forecasting done in 2019 [63], the number of aircraft will increase around 84% until 2038. Also, the International Air Transport Association (IATA) forecast predicts 8.2 billion passengers worldwide by the year of 2037 [64], even after the revision due to the Covid-19 pandemic [65]. Considering this scenario, it is important to deal with the aircraft's GT reliability and maintenance. The GT components operate in a hostile environment that creates degradation mechanisms in the structure of the parts, which increase the risk of mechanical failures [66]. Aiming to ensure the reliability of the GT components the engines are supported by a rigorous maintenance schedule [36]. This maintenance schedule considers the jet engine thermal cycles and the time it is in the air [67]. Even having a rigorous maintenance schedule for preventing failures, due to the high number of aircrafts the aeronautical companies need to be well prepared to monitor and diagnose GT conditions for preventing major accidents and losses that may occur between overhaul [36]. In this way, this chapter aims to exemplify a mechanical engineering application using the proposed model IT2-FLS for detecting aircraft engine gas path faults.

4.5.1
Gas turbine problem formulation

Engine Health Management (EHM) is used by Aircraft Engine Manufacturers in order to maintain an engine operative through a reduction of operational events [68]. The EHM is described as the process of diagnosing and preventing system failures, whilst predicting the reliability and remaining useful life of its components [69]. In order to stimulate the development of a model fully capable of diagnosing a gas turbine engine, NASA created the Propulsion diagnostic method evaluation strategy (ProDiMES) [70], where a software that simulates real data signal of aeronautical gas turbine engines. The ProDiMES allows users to develop diagnostic solutions through the evaluation of sensor signals which are simulated by an environment called Engine Fleet Simulator (EFS) [36]. The EFS uses a modified steady-state version of the NASA Commercial Modular Aero-Propulsion System Simulation (C-MAPSS) which was validated against the original C-MAPSS model [70]. Table 4.5.1 lists the EFS output parameters and Figure 4.8 illustrates stations numbers, modules, and sensors of the C-MAPSS steady-state.

The EFS users can select some properties of the simulated data set, which are: number of engines in the fleet, the number of flights of data will be collected, eighteen different gas path fault types (listed in Table 4.5.1), the fault initiation flight and the fault evolution rate. The faults are divided into

Table 4.8: EFS output parameters.

Index	Symbol	Description
1	Nf	Physical fan speed [rpm]
2	Nc	Physical core speed [rpm]
3	P24	Total pressure at LPC outlet [psia]
4	Ps30	Static pressure at HPC outlet [psia]
5	T24	Total temperature at LPC outlet [°R]
6	T30	Total temperature at HPC outlet [°R]
7	T48	Total temperature at HPT outlet [°R]
8	Wf	Fuel flow [pps]
9	P2	Total pressure at fan inlet [psia]
10	T2	Total temperature at fan inlet [°R]
11	Pamb	Ambient pressure [psia]
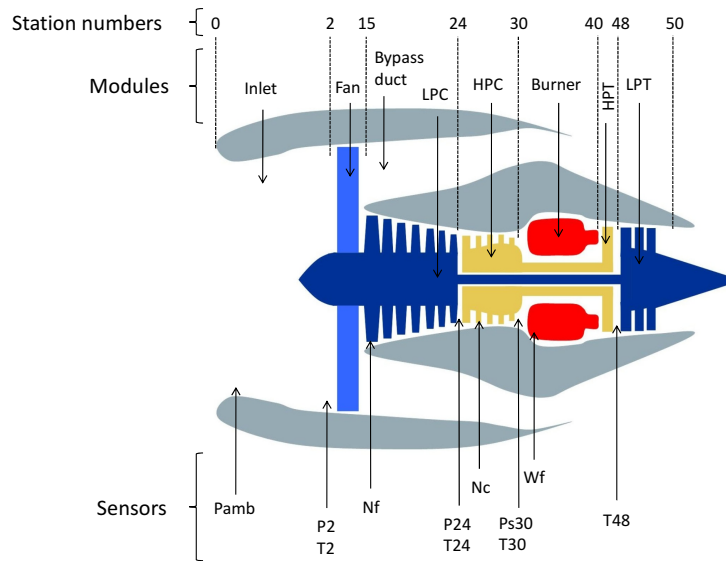
Figure 4.8: C-MAPSS Steady-State station numbers, modules, and sensors [70].

two categories, abrupt faults and rapid faults. The first one evolves to its final magnitude after a number of flights while the second one emerges in its final magnitude. This faults magnitude are divided into three groups: small faults, medium faults, and large faults. In addition to all these properties, the EFS can insert noise fluctuations in the simulated sensor's signals. After the EFS generates the data, the historical sensed parameters can be used to develop a supervised diagnostic model. Thus, the problem consists of design an IT2-FLS UL-GMFus and IT2-FLS UL-GMFum using the EFS generated data to detect faults in GT.

Table 4.9: EFS fault types.

Index	Fault description	Fault component
0	No-fault	–
1	Fan fault	Turbomachinery
2	LPC fault	Turbomachinery
3	HPC fault	Turbomachinery
4	HPT fault	Turbomachinery
5	LPT fault	Turbomachinery
6	VSV fault	Actuator
7	VBV fault	Actuator
8	Nf sensor fault	Sensor
9	Nc sensor fault	Sensor
10	P24 sensor fault	Sensor
11	Ps30 sensor fault	Sensor
12	T24 sensor fault	Sensor
13	T30 sensor fault	Sensor
14	T48 sensor fault	Sensor
15	Wf sensor fault	Sensor
16	P2 sensor fault	Sensor
17	T2 sensor fault	Sensor
18	Pamb sensor fault	Sensor

4.5.2
Pre-processing stage

According to the ProDiMES manual [70] the pre-processing stage will be applied to the data generated by EFS. The pre-processing stage is composed of parameter correction, trend monitoring, anomaly detection, and normalization as illustrated in Figure 4.9. This pre-processing stage was proposed in [70] and further adopted in [36, 71, 72, 73]. At the pre-processing stage, the raw data $\mathbf{R}$ generated by EFS turns out to be seven ones $\mathbf{X}_{proc}$ as recommended by ProDiMES.
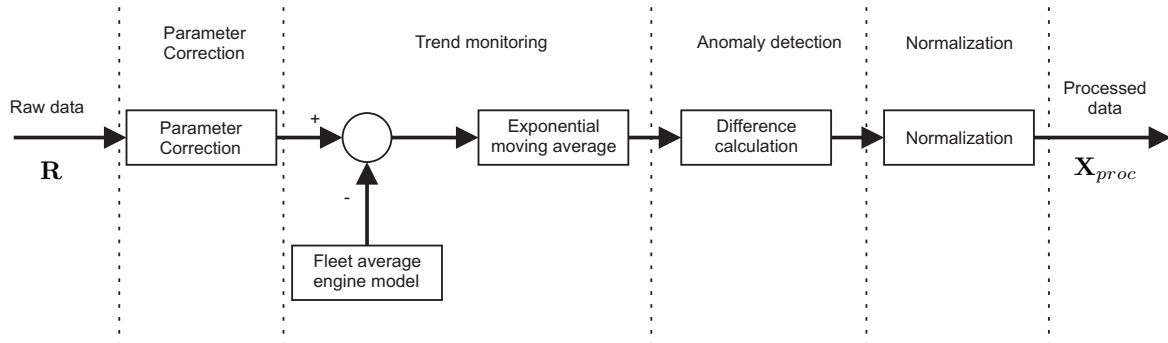

Figure 4.9: Pre-processing diagram block of the EFS generated data [36].

The pre-processing stage has four steps; the first one is called parameter

correction, where all engine parameters are corrected to standard day operating conditions, applying the following equation

$$\text{Parameter}_{\text{corrected}} = \frac{\text{Parameter}}{\theta^a \delta^b}, \tag{4-7}$$

which the temperature θ and pressure correction δ terms, are defined as

$$\theta = \frac{T2\,(^\circ R)}{518.67} \tag{4-8}$$

and

$$\theta = \frac{P2\,(psia)}{14.696}. \tag{4-9}$$

The second step is called trend monitoring, applied to capture gradual performance changes in the form of residuals, or measurements deltas, relative to a fleet average engine [70]. Considering an individual engine corrected data collected in each flight, it is referenced against the fleet average engine model to calculate measurement deltas as [70]

$$\Delta y_i\,(k) = y_i\,(k) - y_{i\_\text{baseline}}\,(k), \tag{4-10}$$

where $y_i\,(k)$ is the corrected value of the *i-th* measurement collected during the *k-th* flight, and $y_{i\_\text{baseline}}\,(k)$ is the fleet average engine value for the *i-th* measurement at the corresponding pressure altitude, Mach number, and corrected fan speed values of the *k-th* flight. Further, the calculated measurement delta values are trended over time by applying an exponential moving average approach given as [70]

$$\Delta y_{i\_ema}\,(k) = \alpha y_{i\_ema}\,(k-1) + (1-\alpha)\,\Delta y_i\,(k) \tag{4-11}$$

$\Delta y_{i\_ema}\,(k)$ is the exponential moving average of the *i-th* measurement delta on flight k. In addition, the moving average weighting between previous and current data is established by a constant $\alpha \in \mathbb{R}\,[0,1]$.

The third step of the pre-processing stage is the anomaly detection, applied to detect discrete events causing a rapid shift in observed measurement delta and to extract the corresponding measurement delta signature of the events [70]. The anomaly detection uses a backward difference calculation of the exponential moving average (EMA) of each measurement delta shown as

$$\Delta\Delta y_{i\_ema}\,(k) = \Delta y_{i\_ema}\,(k) - \Delta y_{i\_ema}\,(k-\beta), \tag{4-12}$$

where $\Delta\Delta y_{i\_ema}\,(k)$ is the change in the EMA of the *i-th* measurement delta between a flight k and some previous flight $k-\beta$.

The last step of the pre-processing stage is the normalization, which gathers the sensors data in a scale standardized by the standard deviation of the equivalent sensor [36][70] using the following equation

$$N\left(\Delta\Delta y_{i\_ema}\left(k\right)\right) = \frac{\Delta\Delta y_{i\_ema}\left(k\right)}{\sigma_i}, \qquad (4\text{-}13)$$

where $N\left(\Delta\Delta y_{i\_ema}\left(k\right)\right)$ is the normalized data and σ_i is the standard deviation of the corresponding measurements delta.

4.5.3
Practical application results

The experimental results consider a dataset composed of 3600 engines performing 50 flights each, where 1800 engines develop one of the 18 fault types with the fault initiation occurring at any flight. These 1800 engines are equally divided to have 100 occurrences of the same fault. In addition for training the fuzzy model, the dataset is divided into 50% as test set, 15% as the validation set, and 35% as the training set. Moreover, the dataset features are composed of sensors values of take-off and cruise operation which after the pre-processing totalizing 14 features [36, 73]. For both investigated models (IT2-FLS UL-GMFus and IT2-FLS UL-GMFum) it was considered $\gamma = 500$, $\alpha = 0.1$, normalization of the dataset to a $[-1, 1]$ range, the SD training method with batch learning, and the threshold equal to zero (i.e., outputs higher or equal to zero are assigned as fault and less than zero as non-fault). In addition, the fuzzy rules are defined as the same as Equation (4-4) with $M = 4$ (two rules for the non-fault and two rules for fault classes).

4.5.3.1
Detecting performance

Table 4.10 lists the confusion matrix of the test set for both IT2-FLS proposed models. Same as noted for benchmarking data sets, both models achieve similar results with test accuracy equal to 81.4%. However, attempting to maintain a level of uniformity in the diagnostic solutions applied for ProDiMES, a target false positive detection rate (false alarm rate) of once per 1,000 flights has to be satisfied [70]. In this scenario, it was defined the threshold equal to 0.61 for both models, attaining 1038 and 1076 flights per false alarm for IT2-FLS UL-GMFum and IT2-FLS UL-GMFus, respectively. Table 4.11 lists the confusion matrix of this new scenario in which IT2-FLS GMFus detected correctly only 18 flights higher than IT2-FLS GMFum, obtaining the same test accuracy equal to 78.5%. Moreover, it is important to notice that the data set is unbalanced. Even attaining around 78% of correct accuracy, both models miss classifying around one-quarter when only the samples predicted as 'non-fault' is considered. In the same way, analyzing only the target samples of 'fault', the models classify correctly only 34%.

Table 4.10: Confusion matrix of the test dataset for both investigated IT2-FLS models, where F and NF means fault and non-fault respectively.

		Predicted						Predicted		
		F	**NF**					**F**	**NF**	
Target	**F**	31299	27241	53.5%		**Target**	**F**	31604	26936	54.0%
		17.4%	15.2%	46.5%				17.6%	15.0%	46.0%
	NF	6072	114938	95.0%			**NF**	6354	114656	94.7%
		3.4%	64.0%	5.0%				3.5%	63.8%	5.3%
		83.8%	80.8%	81.4%				83.3%	81.0%	81.5%
		16.2%	19.2%	18.6%				16.7%	19.0%	18.5%

(a) IT2-FLS UL-GMFum (b) IT2-FLS UL-GMFus

Table 4.11: Confusion matrix of both IT2-FLS models considering at least 1000 flights per false alarm for the test dataset, where F and NF means fault and non-fault respectively.

		Predicted						Predicted		
		F	**NF**					**F**	**NF**	
Target	**F**	20177	38363	34.5%		**Target**	**F**	20189	38351	34.5%
		11.2%	21.3%	65.5%				11.2%	21.3%	65.5%
	NF	173	120837	99.9%			**NF**	167	120843	99.9%
		0.1%	67.3%	0.1%				0.1%	67.3%	0.1%
		99.1%	75.9%	78.5%				99.2%	75.9%	78.5%
		0.9%	24.1%	21.5%				0.8%	24.1%	21.5%

(a) IT2-FLS UL-GMFum (b) IT2-FLS UL-GMFus

5
Conclusions

In this work, we proposed the equations required to use GMFus in an IT2-FLS UL binary classifier trained by the steepest descent method, significantly reducing the computational costs in comparison with IT2-FLS UL-GMFum. In addition, we extended the IT2-FLS UL models (adopting any Gaussian MF) to IT2-FLSMO and provided the necessary equations, simplifying the use of IT2-FLS UL without adopting any decomposition strategies for handling MCP.

The objective of the dissertation was entirely fulfilled by reducing the computational cost of the IT2-FLS with GMFs using the upper and lower method. The experimental results indicated that the approach using the Gaussian MBF with uncertain standard deviation, i.e., IT2-FLS UL-GMFus and IT2-FLSMO UL-GMFus, achieved the same or better performance in comparison with the approach using the Gaussian MBF with uncertain mean, i.e., IT2 FLS UL-GMFum and IT2-FLSMO UL-GMFum. Moreover, the fuzzy models using GMFus were up to 7 times faster than the models using GMFum, which is explained by GMFus requiring fewer equations to compute the gradient vector of the cost function than GMFum. In addition, the IT2-FLSs UL-GMFus attained the competitive performance with the SCG T1-FLSs, which used the second-order information for training the fuzzy rules. These results of the SCG T1-FLSs support the superiority of the second-order training method even when we compare it with the IT2-FLS classifier, which presents better modeling of uncertainties. On the other hand, the single output proposed model was also applied for solving a practical engineering application regarding to a detection fault in aircraft gas turbine. The results shows that the proposed model can detect correctly 81.4% of the faults at the flight and 78.5% even in a worse scenario.

Future research will focus on modifying the IT2-FLSMO classifier to handle additional supervised learning problems, such as the prediction of time series. Also, we will attempt to apply the SCG training method for IT2-FLS UL-GMFus to improve the performance of the model.

Bibliography

[1] D. Graupe, Principles of artificial neural networks, World Scientific, 7 (2013).

[2] A. Ben-Hur, J. Weston, A user's guide to support vector machines, in: Data mining techniques for the life sciences, Springer, 223–239.

[3] Y. L. Pavlov, Random Forests, De Gruyter (2000).

[4] C. Sheppard, Tree-based Machine Learning Algorithms: Decision Trees, Random Forests, and Boosting, CreateSpace Independent Publishing Platform (2017).

[5] J. M. Mendel, Uncertain rule-based fuzzy systems, Springer (2017).

[6] Z. Liao, D. Li, X. Wang, L. Li, Q. Zou, Cancer diagnosis through isomir expression with machine learning method, Current Bioinformatics 13 (1) (2018) 57–63.

[7] S. Mohan, C. Thirumalai, G. Srivastava, Effective heart disease prediction using hybrid machine learning techniques, IEEE Access 7 (2019) 81542–81554.

[8] O. Shahid, M. Nasajpour, S. Pouriyeh, R. M. Parizi, M. Han, M. Valero, F. Li, M. Aledhari, Q. Z. Sheng, Machine learning research towards combating covid-19: Virus detection, spread prevention, and medical assistance, Journal of Biomedical Informatics 117 (2021) 103751.

[9] T. K. Lee, J. H. Cho, D. S. Kwon, S. Y. Sohn, Global stock market investment strategies based on financial network indicators using machine learning techniques, Expert Systems with Applications 117 (2019) 228–242.

[10] P. Hajek, R. Henriques, Mining corporate annual reports for intelligent detection of financial statement fraud – a comparative study of machine learning methods, Knowledge-Based Systems 128 (2017) 139–152.

[11] U. J, N. Metawa, K. Shankar, S. Lakshmanaprabu, Financial crisis prediction model using ant colony optimization, International Journal of Information Management 50 (2020) 538–556.

[12] M. Z. Ali, M. N. S. K. Shabbir, X. Liang, Y. Zhang, T. Hu, Machine learning-based fault diagnosis for single- and multi-faults in induction motors using measured stator currents and vibration signals, IEEE Transactions on Industry Applications 55 (3) (2019) 2378–2391.

[13] C. Gkerekos, I. Lazakis, G. Theotokatos, Machine learning models for predicting ship main engine fuel oil consumption: A comparative study, Ocean Engineering 188 (2019) 106282.

[14] Z. Liu, I. A. Karimi, Gas turbine performance prediction via machine learning, Energy 192 (2020) 116627.

[15] L. Zadeh, Fuzzy sets, Information and Control 8 (3) (1965) 338–353.

[16] J. M. Mendel, Uncertain rule-based fuzzy logic system: introduction and new directions, Prentice–Hall PTR (2001).

[17] A. Keshtkar, S. Arzanpour, An adaptive fuzzy logic system for residential energy management in smart grid environments, Applied Energy 186 (2017) 68–81.

[18] Z. Roumila, D. Rekioua, T. Rekioua, Energy management based fuzzy logic controller of hybrid system wind/photovoltaic/diesel with storage battery, International Journal of Hydrogen Energy 42 (30) (2017) 19525–19535.

[19] H. Taghavifar, S. Rakheja, Path-tracking of autonomous vehicles using a novel adaptive robust exponential-like-sliding-mode fuzzy type-2 neural network controller, Mechanical Systems and Signal Processing 130 (2019) 41–55.

[20] R. A. Campos, R. P. F. Amaral, I. F. M. Menezes, L. G. da Fonseca, M. L. Lagares, E. P. Aguiar, Classification of short circuit gma welding using type-1 and singleton fuzzy logic system, in: 2018 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), IEEE ((2018)), 1 – 7.

[21] B. Wu, T. L. Yip, X. Yan, C. Guedes Soares, Fuzzy logic based approach for ship-bridge collision alert system, Ocean Engineering 187 (2019) 106152.

[22] Y. Zhang, Z. Ye, Short-term traffic flow forecasting using fuzzy logic system methods, Journal of Intelligent Transportation Systems 12 (3) (2008) pp. 102 – 112.

[23] A. H. Hamamoto, L. F. Carvalho, L. D. H. Sampaio, T. Abrão, M. L. Proença, Network anomaly detection system using genetic algorithm and fuzzy logic, Expert Systems with Applications 92 (2018) pp. 390 – 402.

[24] M. Dhimish, V. Holmes, B. Mehrdadi, M. Dales, P. Mather, Photovoltaic fault detection algorithm based on theoretical curves modelling and fuzzy classification system, Energy 140 (2017) pp. 276 – 290.

[25] J. Jiang, C. Syue, C. Wang, J. Wang, J. Shieh, An interval type-2 fuzzy logic system for stock index forecasting based on fuzzy time series and a fuzzy logical relationship map, IEEE Access 6 (2018) pp. 69107 – 69119.

[26] P. A. Herman, G. Prasad, T. M. McGinnity, Designing an interval type-2 fuzzy logic system for handling uncertainty effects in brain–computer interface classification of motor imagery induced eeg patterns, IEEE Transactions on Fuzzy Systems 25 (1) (2017) pp. 29 – 42.

[27] K. Sabahi, S. Ghaemi, M. Badamchizadeh, Designing an adaptive type-2 fuzzy logic system load frequency control for a nonlinear time-delay power system, Applied Soft Computing 43 (2016) pp. 97 – 106.

[28] W. Peng, C. Li, G. Zhang, J. Yi, Interval type-2 fuzzy logic based transmission power allocation strategy for lifetime maximization of wsns, Engineering Applications of Artificial Intelligence 87 (2020) pp. 103269.

[29] J. M. Mendel, General type-2 fuzzy logic systems made simple: A tutorial, IEEE Transactions on Fuzzy Systems 22 (5) (2014) pp. 1162 – 1182.

[30] R. Alcalá, O. Cordón, J. Casillas, F. Herrera, S. Zwir, 29 - learning and tuning fuzzy rule-based systems for linguistic modeling, in: C. T. Leondes (Ed.), Knowledge-Based Systems, Academic Press, San Diego, 889 – 941.

[31] L.-X. Wang, J. M. Mendel, Generating fuzzy rules by learning from examples, IEEE Transactions on systems, man, and cybernetics 22 (6) (1992) 1414 – 1427.

[32] D. Karaboga, E. Kaya, Adaptive network based fuzzy inference system (anfis) training approaches: a comprehensive survey, Artificial Intelligence Review 52 (4) (2019) pp. 2263 – 2293.

[33] F. Herrera, Genetic fuzzy systems: taxonomy, current research trends and prospects, Evolutionary Intelligence 1 (1) (2008) pp. 27 – 46.

[34] L. Wang, J. M. Mendel, Back-propagation fuzzy system as nonlinear dynamic system identifiers, in: 1992 Proc. IEEE International Conference on Fuzzy Systems (1992), 1409 – 1418.

[35] R. P. F. Amaral, M. V. Ribeiro, E. P. Aguiar, Type-1 and singleton fuzzy logic system trained by a fast scaled conjugate gradient methods for dealing with binary classification problems, Neurocomputing Vol. 355 (2019) pp. 57 − 70.

[36] P. H. S. Calderano, M. G. C. Ribeiro, R. P. F. Amaral, M. M. B. R. Vellasco, R. Tanscheit, E. P. Aguiar, An enhanced aircraft engine gas path diagnostic method based on upper and lower singleton type-2 fuzzy logic system, Journal of the Brazilian Society of Mechanical Sciences and Engineering Vol. 41 (2) (2019) pp. 70.

[37] E. P. Aguiar, F. M. A. Nogueira, R. P. F. Amaral, D. F. Fabri, S. C. Rossignoli, J. G. Ferreira, M. M. B. R. Vellasco, R. Tanscheit, P. C. S. Vellasco, M. V. Ribeiro, Eann 2014: A fuzzy logic system trained by conjugate gradient methods for fault classification in a switch machine, Neural Computing and Applications Vol. 27 (5) (2016) pp. 1175–1189.

[38] N. N. Karnik, J. M. Mendel, Introduction to type-2 fuzzy logic systems, in: 1998 IEEE international conference on fuzzy systems proceedings. IEEE world congress on computational intelligence, IEEE, 2 ((1998)), 915 − 920.

[39] W.-W. Tan, D. Wu, Design of type-reduction strategies for type-2 fuzzy logic systems using genetic algorithms (2007) pp. 169 − 187.

[40] E. P. Aguiar, R. P. F. Amaral, M. M. Vellasco, M. V. Ribeiro, An enhanced singleton type-2 fuzzy logic system for fault classification in a railroad switch machine, Electric Power Systems Research Vol. 158 (2018) pp. 195 − 206.

[41] C. M. Bishop, Pattern recognition and machine learning, springer ((2006)).

[42] G. Daniel, Principles of artificial neural networks, World Scientific, 7 ((2013)).

[43] R. P. Finotti Amaral, I. F. Menezes, M. V. Ribeiro, An extension of the type-1 and singleton fuzzy logic system trained by scaled conjugate gradient methods for multiclass classification problems, Neurocomputing Vol. 411 (2020) pp. 149 − 163.

[44] Y. Liu, J.-W. Bi, Z.-P. Fan, A method for multi-class sentiment classification based on an improved one-vs-one (ovo) strategy and the support vector machine (svm) algorithm, Information Sciences Vol. 394 - 395 (2017) pp. 38 − 52.

[45] M. Galar, A. Fernández, E. Barrenechea, H. Bustince, F. Herrera, An overview of ensemble methods for binary classifiers in multi-class problems: Experimental study on one-vs-one and one-vs-all schemes, Pattern Recognition Vol. 44 (8) (2011) pp. 1761 − 1776.

[46] J. Alcalá-Fdez, A. Fernández, J. Luengo, J. Derrac, S. García, L. Sánchez, F. Herrera, Keel data-mining software tool: data set repository, integration of algorithms and experimental analysis framework., Journal of Multiple-Valued Logic & Soft Computing 17.

[47] M. Lichman, UCI machine learning repository, http://archive.ics.uci.edu/ml (2013).

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

[49] Q. Liang, J. Mendel, Interval type-2 fuzzy logic systems: theory and design, IEEE Transactions on Fuzzy Systems 8 (5) (2000) 535–550.

[50] M. F. Møller, A scaled conjugate gradient algorithm for fast supervised learning, Neural Networks 6 (4) (1993) 525–533.

[51] O. Castillo, L. Amador-Angulo, J. R. Castro, M. Garcia-Valdez, A comparative study of type-1 fuzzy logic systems, interval type-2 fuzzy logic systems and generalized type-2 fuzzy logic systems in control problems, Information Sciences Vol. 354 (2016) pp. 257 – 274.

[52] I. Goodfellow, Y. Bengio, A. Courville, Deep learning, MIT press (2016).

[53] S. Theodoridis, K. Koutroumbas, Pattern Recognition (Fourth Edition), Academic Press, Boston (2009).

[54] M. Pota, M. Esposito, G. Pietro, Interpretability indexes for fuzzy classification in cognitive systems, in: 2016 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE) (2016), 24–31.

[55] C. Mencar, Interpretability of fuzzy systems, in: Fuzzy Logic and Applications, Springer International Publishing (2013), 22–35.

[56] C. Rudin, Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead, Nature Machine Intelligence 1 (5) (2019) 206–215.

[57] E. P. Aguiar, R. P. F. Amaral, M. M. B. R. Vellasco, M. V. Ribeiro, Computing derivatives in interval type-2 fuzzy logic systems trained by steepest descent method for fault classification in a switch machine, in: 2017 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE) ((2017)), 1 – 6.

[58] R. Kohavi, A study of cross-validation and bootstrap for accuracy estimation and model selection, in: 14th international joint conference on artificial intelligence (IJCAI), 14 (2) ((1995)), 1137 − 1145.

[59] S. Stehman, Estimating the kappa coefficient and its variance under stratified random sampling, Photogrammetric Engineering and Remote Sensing Vol. 62 (4) (1996) pp. 401 − 407.

[60] M. Sokolova, N. Japkowicz, S. Szpakowicz, Beyond accuracy, f-score and roc: A family of discriminant measures for performance evaluation, in: Australasian joint conference on artificial intelligence, Springer Berlin Heidelberg ((2006)), 1015 − 1021.

[61] D. S. Moore, S. Kirkland, The basic practice of statistics, WH Freeman New York, 2 ((2007)).

[62] M. P. Boyce, Gas turbine engineering handbook, Elsevier (2011).

[63] S. Airbus, Global market forecast 2019–2038, Future Journeys, Blagnac, France.

[64] I. IATA, 2036 forecast reveals air passengers will nearly double to 7.8 billion (2019).

[65] I. IATA, 20 year passenger forecast (2020).

[66] H. Hanachi, C. Mechefske, J. Liu, A. Banerjee, Y. Chen, Performance-based gas turbine health monitoring, diagnostics, and prognostics: A survey, IEEE Transactions on Reliability 67 (3) (2018) 1340–1363.

[67] J. E. Hershey, J. F. Ackerman, V. K. M. Hanagandi, A. V. Aragones, B. E. Osborn, N. W. Chbat, R. A. Korkosz, Method and apparatus for determining an effective jet engine maintenance schedule, uS Patent 6,473,677 (oct 2002).

[68] G. Bastard, J. Lacaille, J. Coupard, Y. Stouky, Engine health management in safran aircraft engines, in: Annual Conference of the PHM Society, 8 (1) (2016).

[69] J.-H. Shin, H.-B. Jun, On condition based maintenance policy, Journal of Computational Design and Engineering 2 (2) (2015) 119–127.

[70] D. L. Simon, Propulsion diagnostic method evaluation strategy (ProDiMES) user's guide, Citeseer (2010).

[71] D. L. Simon, S. Borguet, O. Léonard, X. Zhang, Aircraft engine gas path diagnostic methods: public benchmarking results, American Society of Mechanical Engineers, 55188 (2013).

[72] T. Teixeira, R. Tanscheit, M. Ribeiro, Sistema de inferência fuzzy para diagnóstico de desempenho de turbinas a gás aeronáuticas, in: Proceedings of fourth Brazilian conference on fuzzy systems (2016).

[73] M. G. Castro Ribeiro, P. H. S. Calderano, R. P. F. Amaral, I. F. M. Menezes, R. Tanscheit, M. M. B. R. Vellasco, E. P. Aguiar, Detection and classification of faults in aeronautical gas turbine engine: a comparison between two fuzzy logic systems, in: 2018 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), IEEE (2018), 1–7.

Appendices

A
Partial derivatives of $\nabla J\left(\mathbf{w}^{(\gamma)}\right)$ for the IT2-FLS UL-GMFum

Using the cost function presented in Equation (2-45) the gradient vector $\nabla J\left(\mathbf{w}^{(\gamma)}\right)$ is formed by partial derivatives from $J\left(\mathbf{w}^{(\gamma)}\right)$ regarding to each fuzzy parameter $m_{1\tilde{F}_k^l}(\gamma)$, $m_{2\tilde{F}_k^l}(\gamma)$, $\sigma_{\tilde{F}_k^l}(\gamma)$ and $\theta_l(\gamma)$. Aiming to simplify the partial derivatives equations the following terms have been used

$$e^{(q)} = f_{UL_{T2}}\left(\mathbf{x}^{(q)}\right) - y^{(q)}, \tag{A-1}$$

$$\overline{e}_{\theta_l}^{(q)} = \theta_l\left(\gamma\right) - Y_{U_{T2}}\left(\mathbf{x}^{(q)}\right), \tag{A-2}$$

$$\underline{e}_{\theta_l}^{(q)} = \theta_l\left(\gamma\right) - Y_{L_{T2}}\left(\mathbf{x}^{(q)}\right), \tag{A-3}$$

$$a_{1\tilde{F}_k^l(\gamma)}^{(q)} = \frac{x_k^{(q)} - m_{1\tilde{F}_k^l}\left(\gamma\right)}{\sigma_{\tilde{F}_k^l}^2\left(\gamma\right)}, \tag{A-4}$$

$$a_{2\tilde{F}_k^l(\gamma)}^{(q)} = \frac{x_k^{(q)} - m_{2\tilde{F}_k^l}\left(\gamma\right)}{\sigma_{\tilde{F}_k^l}^2\left(\gamma\right)}, \tag{A-5}$$

$$b_{1\tilde{F}_k^l(\gamma)}^{(q)} = \frac{\left(x_k^{(q)} - m_{1\tilde{F}_k^l}\left(\gamma\right)\right)^2}{\sigma_{\tilde{F}_k^l}^3\left(\gamma\right)} \tag{A-6}$$

and

$$b_{2\tilde{F}_k^l(\gamma)}^{(q)} = \frac{\left(x_k^{(q)} - m_{2\tilde{F}_k^l}\left(\gamma\right)\right)^2}{\sigma_{\tilde{F}_k^l}^3\left(\gamma\right)}. \tag{A-7}$$

GMFum has distinct equations in each part of its domain. Thus the partial derivatives of IT2-FLS UL have to be applied to each domain part.

A.1
Partial derivatives with respect to $m_{1\tilde{F}_k^l}\left(\gamma\right)$:

If $x_k^{(q)} < m_{1\tilde{F}_k^l}(\gamma)$:

$$\frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial m_{1\tilde{F}_k^l}\left(\gamma\right)} = \frac{e^{(q)}}{2}\left(\overline{e}_{\theta_l}^{(q)}\overline{\phi}_l\left(\mathbf{x}^{(q)}\right)a_{1\tilde{F}_k^l(\gamma)}^{(q)}\right) \tag{A-8}$$

If $m_{1\tilde{F}_k^l}(\gamma) \leq x_k^{(q)} \leq m_{2\tilde{F}_k^l}(\gamma)$ and $x_k^{(q)} \leq \frac{m_{1\tilde{F}_k^l}(\gamma) + m_{2\tilde{F}_k^l}(\gamma)}{2}$:

$$\frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial m_{1\tilde{F}_k^l}(\gamma)} = 0. \tag{A-9}$$

If $m_{1\tilde{F}_k^l}(\gamma) \leq x_k^{(q)} \leq m_{2\tilde{F}_k^l}(\gamma)$ and $x_k^{(q)} > \frac{m_{1\tilde{F}_k^l}(\gamma) + m_{2\tilde{F}_k^l}(\gamma)}{2}$:

$$\frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial m_{1\tilde{F}_k^l}(\gamma)} = \frac{e^{(q)}}{2}\left(\underline{e}_{\theta_l}^{(q)}\underline{\phi}_l\left(\mathbf{x}^{(q)}\right)a_{1\tilde{F}_k^l(\gamma)}^{(q)}\right) \tag{A-10}$$

If $x_k^{(\gamma)} > m_{2\tilde{F}_k^l}(\gamma)$:

$$\frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial m_{1\tilde{F}_k^l}(\gamma)} = \frac{e^{(q)}}{2}\left(\underline{e}_{\theta_l}^{(q)}\underline{\phi}_l\left(\mathbf{x}^{(q)}\right)a_{1\tilde{F}_k^l(\gamma)}^{(q)}\right) \tag{A-11}$$

A.2
Partial derivatives with respect to $m_{2\tilde{F}_k^l}(q)$:

If $x_k^{(\gamma)} < m_{1\tilde{F}_k^l}(\gamma)$:

$$\frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial m_{2\tilde{F}_k^l}(\gamma)} = \frac{e^{(q)}}{2}\left(\underline{e}_{\theta_l}^{(q)}\underline{\phi}_l\left(\mathbf{x}^{(q)}\right)a_{2\tilde{F}_k^l(\gamma)}^{(q)}\right) \tag{A-12}$$

If $m_{1\tilde{F}_k^l}(\gamma) \leq x_k^{(q)} \leq m_{2\tilde{F}_k^l}(\gamma)$ and $x_k^{(q)} \leq \frac{m_{1\tilde{F}_k^l}(\gamma) + m_{2\tilde{F}_k^l}(\gamma)}{2}$:

$$\frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial m_{2\tilde{F}_k^l}(\gamma)} = \frac{e^{(q)}}{2}\left(\underline{e}_{\theta_l}^{(q)}\underline{\phi}_l\left(\mathbf{x}^{(q)}\right)a_{2\tilde{F}_k^l(\gamma)}^{(q)}\right) \tag{A-13}$$

If $m_{1\tilde{F}_k^l}(\gamma) \leq x_k^{(q)} \leq m_{2\tilde{F}_k^l}(\gamma)$ and $x_k^{(q)} > \frac{m_{1\tilde{F}_k^l}(\gamma) + m_{2\tilde{F}_k^l}(\gamma)}{2}$:

$$\frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial m_{2\tilde{F}_k^l}(\gamma)} = 0. \tag{A-14}$$

If $x_k^{(q)} > m_{2\tilde{F}_k^l}(\gamma)$:

$$\frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial m_{2\tilde{F}_k^l}(\gamma)} = \frac{e^{(q)}}{2}\left(\overline{e}_{\theta_l}^{(q)}\overline{\phi}_l\left(\mathbf{x}^{(q)}\right)a_{2\tilde{F}_k^l(\gamma)}^{(q)}\right) \tag{A-15}$$

A.3
Partial derivatives with respect to $\sigma_{\tilde{F}_k^l}(q)$:

If $x_k^{(q)} < m_{1\tilde{F}_k^l}^{(q)}$:

$$\frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial \sigma_{\tilde{F}_k^l}(\gamma)} = \frac{e^{(q)}}{2}\left(\overline{e}_{\theta_l}^{(q)}\overline{\phi}_l\left(\mathbf{x}^{(q)}\right)b_{1\tilde{F}_k^l(\gamma)}^{(q)} + \underline{e}_{\theta_l}^{(q)}\underline{\phi}_l\left(\mathbf{x}^{(q)}\right)b_{2\tilde{F}_k^l(\gamma)}^{(q)}\right) \tag{A-16}$$

If $m_{1\tilde{F}_k^l}^{(q)} \leq x_k^{(q)} \leq m_{2\tilde{F}_k^l}^{(q)}$ and $x_k^{(q)} \leq \frac{m_{1\tilde{F}_k^l}^{(q)}+m_{2\tilde{F}_k^l}^{(q)}}{2}$:

$$\frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial \sigma_{\tilde{F}_k^l}(\gamma)} = \frac{e^{(q)}}{2}\left(\underline{e}_{\theta_l}^{(q)}\underline{\phi}_l\left(\mathbf{x}^{(q)}\right)b_{2\tilde{F}_k^l(\gamma)}^{(q)}\right) \tag{A-17}$$

If $m_{1\tilde{F}_k^l}^{(q)} \leq x_k^{(q)} \leq m_{2\tilde{F}_k^l}^{(q)}$ and $x_k^{(q)} > \frac{m_{1\tilde{F}_k^l}^{(q)}+m_{2\tilde{F}_k^l}^{(q)}}{2}$:

$$\frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial \sigma_{\tilde{F}_k^l}(\gamma)} = \frac{e^{(q)}}{2}\left(\underline{e}_{\theta_l}^{(q)}\underline{\phi}_l\left(\mathbf{x}^{(q)}\right)b_{1\tilde{F}_k^l(\gamma)}^{(q)}\right) \tag{A-18}$$

If $x_k^{(q)} > m_{2\tilde{F}_k^l}^{(q)}$:

$$\frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial \sigma_{\tilde{F}_k^l}(\gamma)} = \frac{e^{(q)}}{2}\left(\overline{e}_{\theta_l}^{(q)}\overline{\phi}_l\left(\mathbf{x}^{(q)}\right)b_{2\tilde{F}_k^l(\gamma)}^{(q)}+\underline{e}_{\theta_l}^{(q)}\underline{\phi}_l\left(\mathbf{x}^{(q)}\right)b_{1\tilde{F}_k^l(\gamma)}^{(q)}\right) \tag{A-19}$$

A.4
Partial derivatives with respect to $\theta_l(\gamma)$:

$$\frac{\partial J\left(\mathbf{x}^{(\gamma)}\right)}{\partial \theta_l(\gamma)} = \frac{e^{(q)}}{2}\left(\overline{\phi}_l\left(\mathbf{x}^{(q)}\right) + \underline{\phi}_l\left(\mathbf{x}^{(q)}\right)\right) \tag{A-20}$$

B
Partial derivatives of $\nabla J\left(\mathbf{w}^{(\gamma)}\right)$ for the IT2-FLS UL-GMFus

The same cost function used for IT2-FLS UL-GMFum presented in Equation (2-45) is applied for IT2-FLS UL-GMFus, then the gradient vector $\nabla J\left(\mathbf{w}^{(\gamma)}\right)$ is formed by partial derivatives from $J\left(\mathbf{w}^{(\gamma)}\right)$ regarding to each fuzzy parameter $m_{\tilde{F}_k^l}(\gamma)$, $\sigma_{1\tilde{F}_k^l}(\gamma)$, $\sigma_{2\tilde{F}_k^l}(\gamma)$ and $\theta_l(\gamma)$. Aiming to simplify the partial derivatives equations the following terms have been introduced

$$c_{1\tilde{F}_k^l(\gamma)}^{(q)} = \frac{x_k^{(q)} - m_{\tilde{F}_k^l}(\gamma)}{\sigma_{1\tilde{F}_k^l}^2(\gamma)}, \tag{B-1}$$

$$c_{2\tilde{F}_k^l(\gamma)}^{(q)} = \frac{x_k^{(q)} - m_{\tilde{F}_k^l}(\gamma)}{\sigma_{2\tilde{F}_k^l}^2(\gamma)}, \tag{B-2}$$

$$d_{1\tilde{F}_k^l(\gamma)}^{(q)} = \frac{\left(x_k^{(q)} - m_{\tilde{F}_k^l}(\gamma)\right)^2}{\sigma_{1\tilde{F}_k^l}^3(\gamma)} \tag{B-3}$$

and

$$d_{2\tilde{F}_k^l(\gamma)}^{(q)} = \frac{\left(x_k^{(q)} - m_{\tilde{F}_k^l}(\gamma)\right)^2}{\sigma_{2\tilde{F}_k^l}^3(\gamma)}. \tag{B-4}$$

The partial derivatives of $\nabla J\left(\mathbf{w}^{(\gamma)}\right)$ are expressed as follows

$$\frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial m_{\tilde{F}_k^l}(\gamma)} = \frac{e^{(q)}}{2}\left(\overline{e}_{\theta_l}^{(q)}\overline{\phi}_l\left(\mathbf{x}^{(q)}\right)c_{2\tilde{F}_k^l(\gamma)}^{(q)} + \underline{e}_{\theta_l}^{(q)}\underline{\phi}_l\left(\mathbf{x}^{(q)}\right)c_{1\tilde{F}_k^l(\gamma)}^{(q)}\right), \tag{B-5}$$

$$\frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial \sigma_{1\tilde{F}_k^l}(\gamma)} = \frac{e^{(q)}}{2}\left(\underline{e}_{\theta_l}^{(q)}\underline{\phi}_l\left(\mathbf{x}^{(q)}\right)d_{1\tilde{F}_k^l(\gamma)}^{(q)}\right), \tag{B-6}$$

$$\frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial \sigma_{2\tilde{F}_k^l}(\gamma)} = \frac{e^{(q)}}{2}\left(\overline{e}_{\theta_l}^{(q)}\overline{\phi}_l\left(\mathbf{x}^{(q)}\right)d_{2\tilde{F}_k^l(\gamma)}^{(q)}\right) \tag{B-7}$$

and

$$\frac{\partial J\left(\mathbf{x}^{(\gamma)}\right)}{\partial \theta_l(\gamma)} = \frac{e^{(q)}}{2}\left(\overline{\phi}_l\left(\mathbf{x}^{(q)}\right) + \underline{\phi}_l\left(\mathbf{x}^{(q)}\right)\right). \tag{B-8}$$

C
Partial derivatives of $\nabla \mathbf{J}\left(\mathbf{w}^{(\gamma)}\right)$ for the IT2-FLSMO UL-GMFum

Using the cost function presented in Equation (3-10) the gradient vector $\nabla J\left(\mathbf{w}^{(\gamma)}\right)$ is formed by partial derivatives from $J\left(\mathbf{w}^{(\gamma)}\right)$ with respect to each fuzzy parameter $m_{1\tilde{F}_k^l}(\gamma)$, $m_{2\tilde{F}_k^l}(\gamma)$, $\sigma_{\tilde{F}_k^l}(\gamma)$ and $\theta_l^t(\gamma)$. Aiming to simplify the partial derivatives equations the following terms have been introduced

$$e_t^{(q)} = f_{ULT2mo}^t\left(\mathbf{x}^{(q)}\right) - y^t(q), \tag{C-1}$$

$$\overline{e}_{\theta_l}^t(q) = \theta_l^t\left(\gamma\right) - Y_{U_{T2mo}}^t\left(\mathbf{x}^{(q)}\right) \tag{C-2}$$

and

$$\underline{e}_{\theta_l}^t(q) = \theta_l^t\left(\gamma\right) - Y_{L_{T2mo}}^t\left(\mathbf{x}^{(q)}\right). \tag{C-3}$$

Following the same explanation presented in A, the partial derivatives of IT2-FLSMO UL have to be applied in each domain part of GMFum.

C.1
Partial derivatives with respect to $m_{1\tilde{F}_k^l}(\gamma)$

If $x_k^{(q)} < m_{1\tilde{F}_k^l}(\gamma)$:

$$\frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial m_{1\tilde{F}_k^l}(\gamma)} = \sum_{t=1}^{\tau} \frac{e_t^{(q)}}{2}\left(\overline{e}_{\theta_l}^t(q)\overline{\phi}_l^t\left(\mathbf{x}^{(q)}\right)a_{1\tilde{F}_k^l(\gamma)}^{(q)}\right) \tag{C-4}$$

If $m_{1\tilde{F}_k^l}(\gamma) \le x_k^{(q)} \le m_{2\tilde{F}_k^l}(\gamma)$ and $x_k^{(q)} \le \frac{m_{1\tilde{F}_k^l}(\gamma)+m_{2\tilde{F}_k^l}(\gamma)}{2}$:

$$\frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial m_{1\tilde{F}_k^l}(\gamma)} = 0. \tag{C-5}$$

If $m_{1\tilde{F}_k^l}(\gamma) \le x_k^{(q)} \le m_{2\tilde{F}_k^l}(\gamma)$ and $x_k^{(q)} > \frac{m_{1\tilde{F}_k^l}(\gamma)+m_{2\tilde{F}_k^l}(\gamma)}{2}$:

$$\frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial m_{1\tilde{F}_k^l}(\gamma)} = \sum_{t=1}^{\tau} \frac{e_t^{(q)}}{2}\left(\underline{e}_{\theta_l}^t(q)\underline{\phi}_l^t\left(\mathbf{x}^{(q)}\right)a_{1\tilde{F}_k^l(\gamma)}^{(q)}\right) \tag{C-6}$$

If $x_k^{(\gamma)} > m_{2\tilde{F}_k^l}(\gamma)$:

$$\frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial m_{1\tilde{F}_k^l}(\gamma)} = \sum_{t=1}^{\tau} \frac{e_t^{(q)}}{2}\left(\underline{e}_{\theta_l}^t(q)\underline{\phi}_l^t\left(\mathbf{x}^{(q)}\right)a_{1\tilde{F}_k^l(\gamma)}^{(q)}\right) \tag{C-7}$$

C.2
Partial derivatives with respect to $m_{2\tilde{F}_k^l}(\gamma)$

If $x_k^{(\gamma)} < m_{1\tilde{F}_k^l}(\gamma)$:

$$\frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial m_{2\tilde{F}_k^l}(\gamma)} = \sum_{t=1}^{\tau} \frac{e_t^{(q)}}{2}\left(\underline{e}_{\theta_l}^t(q)\underline{\phi}_l^t\left(\mathbf{x}^{(q)}\right)a_{2\tilde{F}_k^l(\gamma)}^{(q)}\right) \tag{C-8}$$

If $m_{1\tilde{F}_k^l}(\gamma) \leq x_k^{(q)} \leq m_{2\tilde{F}_k^l}(\gamma)$ and $x_k^{(q)} \leq \frac{m_{1\tilde{F}_k^l}(\gamma)+m_{2\tilde{F}_k^l}(\gamma)}{2}$:

$$\frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial m_{2\tilde{F}_k^l}(\gamma)} = \sum_{t=1}^{\tau} \frac{e_t^{(q)}}{2}\left(\underline{e}_{\theta_l}^t(q)\underline{\phi}_l^t\left(\mathbf{x}^{(q)}\right)a_{2\tilde{F}_k^l(\gamma)}^{(q)}\right) \tag{C-9}$$

If $m_{1\tilde{F}_k^l}(\gamma) \leq x_k^{(q)} \leq m_{2\tilde{F}_k^l}(\gamma)$ and $x_k^{(q)} > \frac{m_{1\tilde{F}_k^l}(\gamma)+m_{2\tilde{F}_k^l}(\gamma)}{2}$:

$$\frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial m_{2\tilde{F}_k^l}(\gamma)} = 0. \tag{C-10}$$

If $x_k^{(q)} > m_{2\tilde{F}_k^l}(\gamma)$:

$$\frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial m_{2\tilde{F}_k^l}(\gamma)} = \sum_{t=1}^{\tau} \frac{e_t^{(q)}}{2}\left(\overline{e}_{\theta_l}^t(q)\overline{\phi}_l^t\left(\mathbf{x}^{(q)}\right)a_{2\tilde{F}_k^l(\gamma)}^{(q)}\right) \tag{C-11}$$

C.3
Partial derivatives with respect to $\sigma_{\tilde{F}_k^l}(\gamma)$

If $x_k^{(q)} < m_{1\tilde{F}_k^l}(\gamma)$:

$$\frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial \sigma_{\tilde{F}_k^l}(\gamma)} = \sum_{t=1}^{\tau} \frac{e_t^{(q)}}{2}\left(\overline{e}_{\theta_l}^t(q)\overline{\phi}_l^t\left(\mathbf{x}^{(q)}\right)b_{1\tilde{F}_k^l(\gamma)}^{(q)}+\underline{e}_{\theta_l}^t(q)\underline{\phi}_l^t\left(\mathbf{x}^{(q)}\right)b_{2\tilde{F}_k^l(\gamma)}^{(q)}\right) \tag{C-12}$$

If $m_{1\tilde{F}_k^l}(\gamma) \leq x_k^{(q)} \leq m_{2\tilde{F}_k^l}(\gamma)$ and $x_k^{(q)} \leq \frac{m_{1\tilde{F}_k^l}(\gamma)+m_{2\tilde{F}_k^l}(\gamma)}{2}$:

$$\frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial \sigma_{\tilde{F}_k^l}(\gamma)} = \sum_{t=1}^{\tau} \frac{e_t^{(q)}}{2}\left(\underline{e}_{\theta_l}^t(q)\underline{\phi}_l^t\left(\mathbf{x}^{(q)}\right)b_{2\tilde{F}_k^l(\gamma)}^{(q)}\right) \tag{C-13}$$

If $m_{1\tilde{F}_k^l}(\gamma) \leq x_k^{(q)} \leq m_{2\tilde{F}_k^l}(\gamma)$ and $x_k^{(q)} > \frac{m_{1\tilde{F}_k^l}(\gamma)+m_{2\tilde{F}_k^l}(\gamma)}{2}$:

$$\frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial \sigma_{\tilde{F}_k^l}(\gamma)} = \sum_{t=1}^{\tau} \frac{e_t^{(q)}}{2}\left(\underline{e}_{\theta_l}^t(q)\underline{\phi}_l^t\left(\mathbf{x}^{(q)}\right)b_{1\tilde{F}_k^l(\gamma)}^{(q)}\right) \tag{C-14}$$

If $x_k^{(q)} > m_{2\tilde{F}_k^l}(\gamma)$:

$$\frac{\partial J\left(\mathbf{w}^{(\gamma)}\right)}{\partial \sigma_{\tilde{F}_k^l}(\gamma)} = \sum_{t=1}^{\tau} \frac{e_t^{(q)}}{2}\left(\overline{e}_{\theta_l}^t(q)\overline{\phi}_l^t\left(\mathbf{x}^{(q)}\right)b_{2\tilde{F}_k^l(\gamma)}^{(q)}+\underline{e}_{\theta_l}^t(q)\underline{\phi}_l^t\left(\mathbf{x}^{(q)}\right)b_{1\tilde{F}_k^l(\gamma)}^{(q)}\right) \tag{C-15}$$

C.4
Partial derivatives with respect to $\theta_l^t(\gamma)$

$$\frac{\partial J\left(\mathbf{x}^{(\gamma)}\right)}{\partial \theta_l^t(\gamma)} = \frac{e_t^{(q)}}{2}\left(\overline{\phi}_l^t\left(\mathbf{x}^{(q)}\right) + \underline{\phi}_l^t\left(\mathbf{x}^{(q)}\right)\right) \qquad\qquad \text{(C-16)}$$

D
Partial derivatives of $\nabla \mathbf{J}\left(\mathbf{w}^{(\gamma)}\right)$ for the IT2-FLSMO UL-GMFus

Adopting the cost function used for IT2-FLSMO UL-GMFum presented in Equation (3-10) the gradient vector $\nabla J\left(\mathbf{w}^{(\gamma)}\right)$ is formed by partial derivatives from $J\left(\mathbf{w}^{(\gamma)}\right)$ with respect to each fuzzy parameter, i.e., $m_{\tilde{F}_k^l}(\gamma)$, $\sigma_{1\tilde{F}_k^l}(\gamma)$, $\sigma_{2\tilde{F}_k^l}(\gamma)$ and $\theta_l^t(\gamma)$, and are expressed as

$$\frac{\partial J\left(\mathbf{w}^{(q)}\right)}{\partial m_{F_k^l}\left(\gamma\right)} = \sum_{t=1}^{\tau} \frac{e_t^{(q)}}{2} \left(\overline{e}_{\theta_l}^t(q)\overline{\phi}_l^t\left(\mathbf{x}^{(q)}\right)c_{2F_k^l(\gamma)}^{(q)} + \underline{e}_{\theta_l}^t(q)\underline{\phi}_l^t\left(\mathbf{x}^{(q)}\right)c_{1F_k^l(\gamma)}^{(q)}\right), \quad \text{(D-1)}$$

$$\frac{\partial J\left(\mathbf{w}^{(q)}\right)}{\partial \sigma_{1F_k^l}\left(\gamma\right)} = \sum_{t=1}^{\tau} \frac{e_t^{(q)}}{2} \left(\underline{e}_{\theta_l}^t(q)\underline{\phi}_l^t\left(\mathbf{x}^{(q)}\right)d_{1F_k^l(\gamma)}^{(q)}\right), \quad \text{(D-2)}$$

$$\frac{\partial J\left(\mathbf{w}^{(q)}\right)}{\partial \sigma_{2F_k^l}\left(\gamma\right)} = \sum_{t=1}^{\tau} \frac{e_t^{(q)}}{2} \left(\overline{e}_{\theta_l}^t(q)\overline{\phi}_l^t\left(\mathbf{x}^{(q)}\right)d_{2F_k^l(\gamma)}^{(q)}\right) \quad \text{(D-3)}$$

and

$$\frac{\partial J\left(\mathbf{x}^{(q)}\right)}{\partial \theta_l^t\left(\gamma\right)} = \frac{e_t^{(q)}}{2} \left(\overline{\phi}_l^t\left(\mathbf{x}^{(q)}\right) + \underline{\phi}_l^t\left(\mathbf{x}^{(q)}\right)\right). \quad \text{(D-4)}$$

E
## Preliminary results of FLSs trained by SD method

The first performance analyses of the proposed models were to adopt the same step size for SD method as used in [43] (i.e., $\alpha = 0.01$). Table E.1 lists the performance metrics comparisons between the proposed models with the T1-FLS trained by SCG method. Analysing the best epoch of training, it can be stated that in almost all datasets the mean of best epoch was near to the stipulated final epoch of training, with a higher standard deviation. Moreover, Figures E.1–E.4 show that the proposed IT2-FLS UL obtained a slight increase in training accuracy and MSE over the epochs. Therefore, based on the abovementioned aspects, we can conclude that $\alpha = 0.01$ is too small for IT2-FLS UL models.

Table E.1: Performance comparison in terms of the mean and standard deviation using $\alpha = 0.01$ for BCPs.

| Dataset	Method	Training accuracy [%]	Test accuracy [%]	Training MSE	Test MSE	Training Kappa	Test Kappa	Training F-score	Test F-score	Best epoch
Appendicitis	IT2-FLS UL-GMFum	87.93 (±3.25)	85.00 (±7.61)	0.38 (±0.08)	0.48 (±0.21)	0.59 (±0.10)	0.56 (±0.21)	0.79 (±0.05)	0.78 (±0.10)	162 (±63)
	IT2-FLS UL-GMFus	86.93 (±3.32)	84.32 (±7.80)	0.40 (±0.09)	0.49 (±0.21)	0.58 (±0.10)	0.55 (±0.21)	0.79 (±0.05)	0.77 (±0.10)	175 (±50)
	SD T1-FLS	91.10 (±3.31)	84.72 (±6.86)	0.31 (±0.09)	0.51 (±0.22)	0.67 (±0.11)	0.52 (±0.21)	0.84 (±0.06)	0.76 (±0.11)	134 (±7)
	SCG T1-FLS	95.57 (±2.11)	84.98 (±6.63)	0.15 (±0.07)	0.51 (±0.21)	0.84 (±0.10)	0.49 (±0.22)	0.92 (±0.05)	0.74 (±0.12)	39 (±58)
Haberman	IT2-FLS UL-GMFum	75.18 (±2.03)	74.20 (±2.67)	0.72 (±0.05)	0.77 (±0.06)	0.12 (±0.09)	0.09 (±0.11)	0.51 (±0.07)	0.49 (±0.07)	179 (±49)
	IT2-FLS UL-GMFus	75.44 (±2.08)	74.93 (±2.58)	0.71 (±0.05)	0.75 (±0.06)	0.15 (±0.09)	0.13 (±0.11)	0.54 (±0.06)	0.52 (±0.07)	181 (±40)
	SD T1-FLS	77.24 (±1.96)	74.88 (±3.89)	0.67 (±0.04)	0.74 (±0.07)	0.28 (±0.08)	0.20 (±0.12)	0.62 (±0.05)	0.58 (±0.07)	170 (±50)
	SCG T1-FLS	81.34 (±2.38)	74.06 (±3.82)	0.57 (±0.05)	0.75 (±0.09)	0.44 (±0.09)	0.19 (±0.12)	0.71 (±0.05)	0.58 (±0.07)	25 (±42)
Ionosphere	IT2-FLS UL-GMFum	74.37 (±4.20)	72.27 (±6.09)	0.67 (±0.09)	0.73 (±0.14)	0.51 (±0.07)	0.47 (±0.10)	0.74 (±0.04)	0.72 (±0.06)	199 (±1)
	IT2-FLS UL-GMFus	72.30 (±2.20)	70.64 (±4.85)	0.73 (±0.04)	0.78 (±0.12)	0.48 (±0.04)	0.45 (±0.08)	0.72 (±0.02)	0.71 (±0.05)	200 (±0)
	SD T1-FLS	64.54 (±1.99)	64.39 (±4.92)	0.77 (±0.03)	0.78 (±0.06)	0.36 (±0.04)	0.36 (±0.07)	0.64 (±0.03)	0.64 (±0.05)	199 (±3)
	SCG T1-FLS	95.81 (±4.83)	87.46 (±6.48)	0.13 (±0.12)	0.39 (±0.16)	0.91 (±0.10)	0.73 (±0.13)	0.95 (±0.05)	0.86 (±0.07)	91 (±60)
Liver	IT2-FLS UL-GMFum	61.74 (±3.17)	59.04 (±4.83)	0.94 (±0.04)	0.98 (±0.06)	0.14 (±0.09)	0.10 (±0.11)	0.54 (±0.07)	0.51 (±0.08)	189 (±39)
	IT2-FLS UL-GMFus	60.02 (±3.41)	57.93 (±5.12)	0.98 (±0.04)	1.00 (±0.06)	0.13 (±0.09)	0.09 (±0.11)	0.54 (±0.06)	0.52 (±0.07)	183 (±50)
	SD T1-FLS	74.68 (±2.45)	67.78 (±5.17)	0.75 (±0.04)	0.87 (±0.07)	0.44 (±0.07)	0.31 (±0.11)	0.71 (±0.04)	0.65 (±0.06)	188 (±40)
	SCG T1-FLS	85.97 (±2.12)	69.87 (±4.99)	0.47 (±0.05)	0.83 (±0.09)	0.70 (±0.05)	0.37 (±0.11)	0.85 (±0.02)	0.68 (±0.06)	25 (±21)
Monk2	IT2-FLS UL-GMFum	93.04 (±1.09)	92.48 (±2.68)	0.26 (±0.05)	0.28 (±0.05)	0.86 (±0.02)	0.85 (±0.05)	0.93 (±0.01)	0.92 (±0.03)	200 (±0)
	IT2-FLS UL-GMFus	91.48 (±1.09)	90.87 (±2.77)	0.31 (±0.03)	0.32 (±0.05)	0.83 (±0.02)	0.82 (±0.05)	0.91 (±0.01)	0.91 (±0.03)	200 (±0)
	SD T1-FLS	91.34 (±1.09)	90.71 (±3.06)	0.36 (±0.03)	0.38 (±0.05)	0.83 (±0.02)	0.82 (±0.06)	0.91 (±0.01)	0.91 (±0.03)	200 (±0)
	SCG T1-FLS	99.10 (±1.22)	98.74 (±1.83)	0.02 (±0.03)	0.03 (±0.04)	0.98 (±0.03)	0.97 (±0.04)	0.99 (±0.01)	0.99 (±0.02)	189 (±24)
Parkinson	IT2-FLS UL-GMFum	79.98 (±4.20)	75.59 (±7.39)	0.52 (±0.08)	0.64 (±0.18)	0.49 (±0.09)	0.42 (±0.15)	0.74 (±0.05)	0.70 (±0.08)	186 (±38)
	IT2-FLS UL-GMFus	79.06 (±4.91)	74.37 (±8.01)	0.57 (±0.09)	0.67 (±0.18)	0.50 (±0.08)	0.42 (±0.15)	0.74 (±0.04)	0.70 (±0.08)	189 (±32)
	SD T1-FLS	88.11 (±3.24)	80.75 (±5.82)	0.39 (±0.08)	0.60 (±0.18)	0.63 (±0.10)	0.45 (±0.17)	0.81 (±0.05)	0.72 (±0.09)	162 (±61)
	SCG T1-FLS	95.27 (±2.36)	85.80 (±5.88)	0.17 (±0.07)	0.45 (±0.17)	0.86 (±0.07)	0.58 (±0.18)	0.93 (±0.04)	0.78 (±0.09)	100 (±64)
Pima	IT2-FLS UL-GMFum	73.97 (±1.56)	73.12 (±3.20)	0.69 (±0.03)	0.71 (±0.06)	0.38 (±0.05)	0.36 (±0.08)	0.68 (±0.03)	0.67 (±0.04)	200 (±0)
	IT2-FLS UL-GMFus	73.43 (±1.62)	72.62 (±3.18)	0.72 (±0.04)	0.73 (±0.06)	0.35 (±0.05)	0.33 (±0.08)	0.67 (±0.03)	0.66 (±0.04)	200 (±1)
	SD T1-FLS	79.06 (±1.28)	76.74 (±3.23)	0.57 (±0.02)	0.63 (±0.06)	0.52 (±0.03)	0.47 (±0.07)	0.76 (±0.02)	0.73 (±0.04)	175 (±45)
	SCG T1-FLS	86.43 (±1.75)	76.56 (±3.20)	0.44 (±0.03)	0.64 (±0.08)	0.69 (±0.04)	0.47 (±0.07)	0.85 (±0.02)	0.73 (±0.04)	13 (±11)
Sonar	IT2-FLS UL-GMFum	93.12 (±3.09)	77.60 (±7.09)	0.26 (±0.09)	0.68 (±0.20)	0.83 (±0.07)	0.55 (±0.14)	0.92 (±0.04)	0.77 (±0.07)	170 (±45)
	IT2-FLS UL-GMFus	91.25 (±3.19)	77.01 (±6.85)	0.32 (±0.09)	0.68 (±0.20)	0.81 (±0.07)	0.54 (±0.14)	0.90 (±0.04)	0.77 (±0.07)	181 (±39)
	SD T1-FLS	96.63 (±1.61)	79.42 (±6.08)	0.16 (±0.04)	0.61 (±0.16)	0.87 (±0.09)	0.58 (±0.12)	0.93 (±0.05)	0.79 (±0.06)	134 (±67)
	SCG T1-FLS	99.69 (±0.51)	79.72 (±6.57)	0.01 (±0.02)	0.63 (±0.21)	0.99 (±0.01)	0.59 (±0.13)	1.00 (±0.01)	0.79 (±0.07)	34 (±52)
South Africa	IT2-FLS UL-GMFum	71.99 (±2.04)	69.63 (±3.84)	0.74 (±0.03)	0.80 (±0.08)	0.35 (±0.05)	0.31 (±0.09)	0.67 (±0.03)	0.65 (±0.05)	177 (±58)
	IT2-FLS UL-GMFus	71.95 (±2.07)	69.88 (±4.06)	0.75 (±0.03)	0.80 (±0.08)	0.35 (±0.05)	0.31 (±0.10)	0.67 (±0.03)	0.65 (±0.05)	181 (±54)
	SD T1-FLS	74.86 (±1.71)	70.58 (±4.12)	0.68 (±0.03)	0.77 (±0.08)	0.42 (±0.05)	0.34 (±0.09)	0.71 (±0.03)	0.67 (±0.04)	174 (±56)
	SCG T1-FLS	88.14 (±2.15)	71.15 (±3.96)	0.42 (±0.05)	0.77 (±0.09)	0.72 (±0.05)	0.34 (±0.09)	0.86 (±0.03)	0.67 (±0.05)	12 (±15)
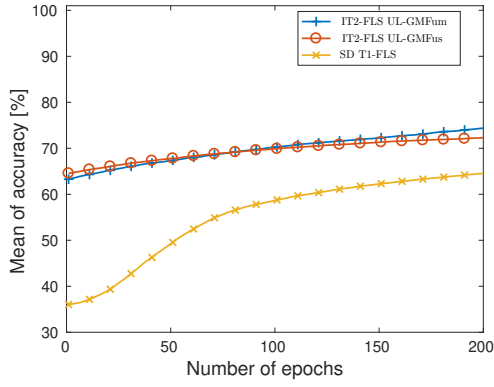
E.1(a): Accuracy.
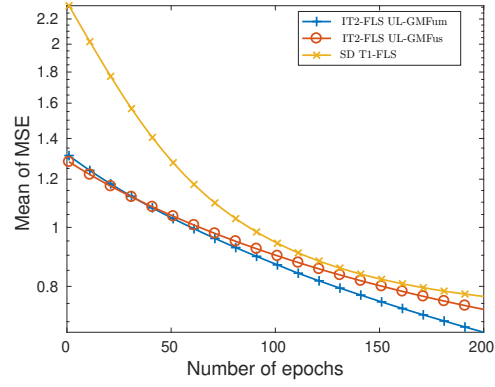
E.1(b): MSE.

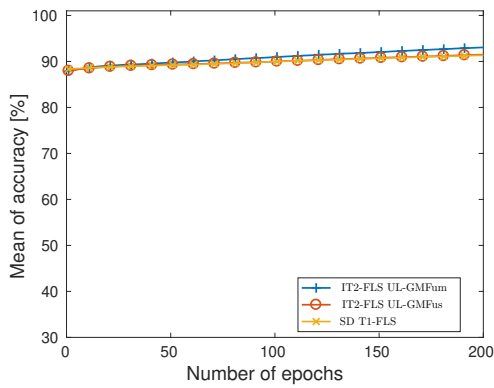Figure E.1: Ionosphere dataset: average performance of FLS models.
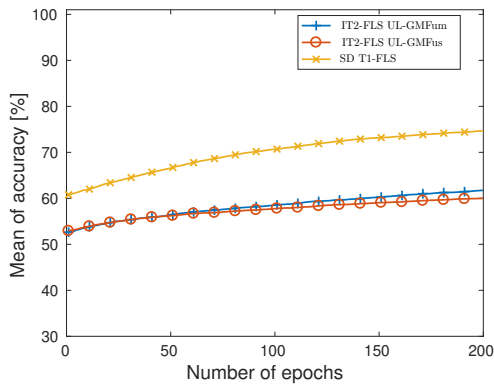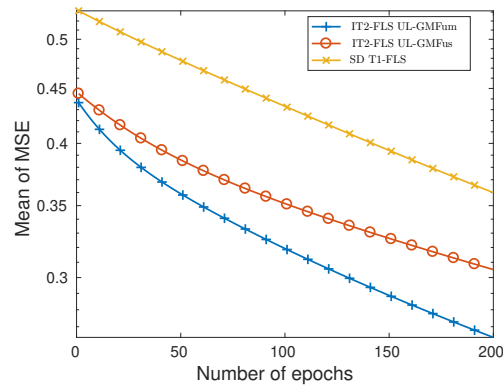

E.2(a): Accuracy.
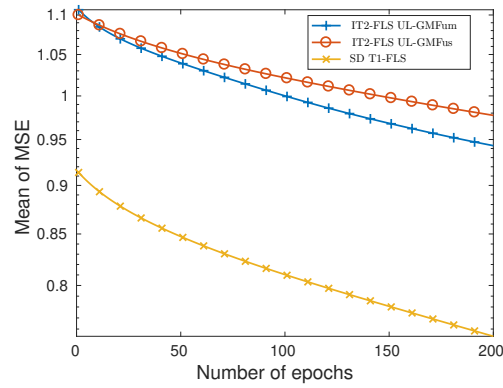
E.2(b): MSE.

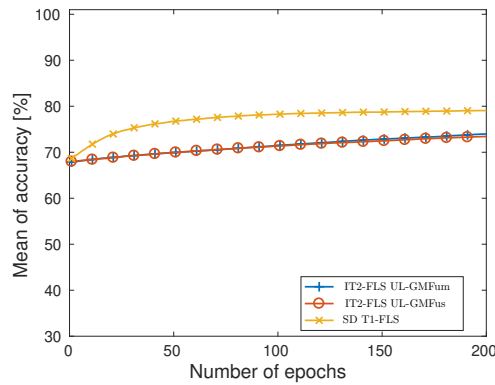Figure E.2: Liver disorders dataset: average performance of FLS models.
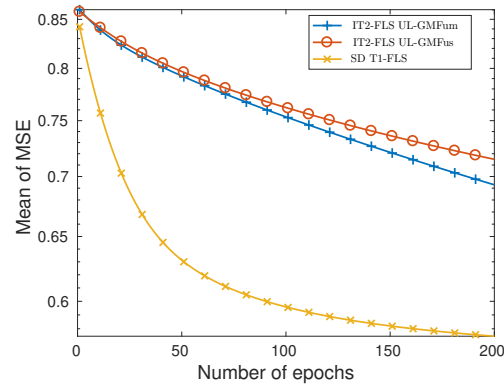

E.3(a): Accuracy.

E.3(b): MSE.

Figure E.3: Monk2 dataset: average performance of FLS models.

E.4(a): Accuracy. E.4(b): MSE.

Figure E.4: Pima dataset: average performance of FLS models.

F
Publications

During my doctoral studies, the following papers were published:

1. R. A. Campos, **R. P. F. Amaral**, I. F. M. de Menezes, L. G. da Fonseca, M. L. Lagares, E. P. de Aguiar, Classification of short circuit gma welding using type-1 and singleton fuzzy logic system, in: 2018 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), IEEE (2018), pp. 1–7.

2. M. G. de Castro Ribeiro, P. H. S. Calderano, **R. P. F. Amaral**, I. F. M. de Menezes, R. Tanscheit, M. M. B. R. Vellasco, E. P. de Aguiar, Detection and classification of faults in aeronautical gas turbine engine: a comparison between two fuzzy logic systems, in: 2018 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), IEEE (2018), pp. 1–7.

3. E. P. de Aguiar, **R. P. F. Amaral**, M. M. Vellasco, M. V. Ribeiro, An enhanced singleton type-2 fuzzy logic system for fault classification in a railroad switch machine, Electric Power Systems Research Vol. 158 (2018) pp. 195 – 206.

4. **R. P. F. Amaral**, M. V. Ribeiro, E. P. de Aguiar, Type-1 and singleton fuzzy logic system trained by a fast scaled conjugate gradient methods for dealing with binary classification problems, Neurocomputing Vol. 355 (2019) pp. 57 – 70.

5. P. H. S. Calderano, M. G. C. Ribeiro, **R. P. F. Amaral**, M. M. B. R. Vellasco, R. Tanscheit, E. P. de Aguiar, An enhanced aircraft engine gas path diagnostic method based on upper and lower singleton type-2 fuzzy

logic system, Journal of the Brazilian Society of Mechanical Sciences and Engineering Vol. 41 (2) (2019) pp. 70.

6. R. A. Campos,**R. P. F. Amaral**, N. Soares, L. G. da Fonseca, M. L. L. Júnior, and E. P. de Aguiar, A new model to distinguish welds performed by short-circuit GMAW based on FRESH algorithm and MLP ANN, Journal of the Brazilian Society of Mechanical Sciences and Engineering Vol. 41 (3) (2019) pp. 120.

7. **R. P. Finotti Amaral**, I. F. M. Menezes, M. V. Ribeiro, An extension of the type-1 and singleton fuzzy logic system trained by scaled conjugate gradient methods for multiclass classification problems, Neurocomputing Vol. 411 (2020) pp. 149 – 163