Universidade Federal de Juiz de Fora

Engenharia Elétrica

Programa de Pós-Graduação em Engenharia Elétrica

Eduardo Pestana de Aguiar

Fuzzy Logic System Applied to Classification Problems in Railways

Juiz de Fora

2016

Ficha catalográfica elaborada através do Modelo Latex do CDC da UFJF com os dados fornecidos pelo(a) autor(a)

Aguiar, Eduardo.
Fuzzy Logic System Applied to Classification Problems in Railways / Eduardo Pestana de Aguiar. – 2016.
99 f. : il.

Orientador: Moisés Vidal Ribeiro
Coorientadora: Marley Maria Bernardes Rebuzzi Vellasco
Tese de Doutorado – Universidade Federal de Juiz de Fora, Engenharia Elétrica. Programa de Pós-Graduação em Engenharia Elétrica, 2016.

1. Type-2 fuzzy logic systems. 2. Adaptive algorithms. 3. Classification. 4. Railway. I. Vidal Ribeiro, Moisés, orient. II. Bernardes Rebuzzi Vellasco, Marley Maria, coorient. III. Título.

Eduardo Pestana de Aguiar

Fuzzy Logic System Applied to Classification Problems in Railways

Tese de doutorado apresentada ao Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal de Juiz de Fora, na área de concentração em sistemas eletrônicos, como requisito parcial para obtenção do título de Doutor em Engenharia Elétrica.

Orientador: Moisés Vidal Ribeiro

Coorientadora: Marley Maria Bernardes Rebuzzi Vellasco

Juiz de Fora

2016

Eduardo Pestana de Aguiar

Fuzzy Logic System Applied to Classification Problems in Railways

Tese de doutorado apresentada ao Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal de Juiz de Fora, na área de concentração em sistemas eletrônicos, como requisito parcial para obtenção do título de Doutor em Engenharia Elétrica.

Aprovada em: 26 de setembro de 2016

BANCA EXAMINADORA

Prof. Dr. Moisés Vidal Ribeiro - Orientador
Universidade Federal de Juiz de Fora

Prof. Dr. Marley Maria Bernardes Rebuzzi Vellasco - Coorientadora
Pontifícia Universidade Católica do Rio de Janeiro

Prof. Dr. Jorge Luís Machado do Amaral
Universidade do Estado do Rio de Janeiro

Prof. Dr. Walmir Matos Caminhas
Universidade Federal de Minas Gerais

Prof. Dr. André Luís Marques Marcato
Universidade Federal de Juiz de Fora

Prof. Dr. Leonardo Willer de Oliveira
Universidade Federal de Juiz de Fora

To my parents, Gilberto and Sônia

ACKNOWLEDGMENTS

To God, for giving me health, willpower to overcome obstacles, and for brings the right people into my life.

To Prof. Moisés for the guidance, encouragement, and friendship during this gratifying learning process, which started in 2008. I am also thankful for the excellent example he has provided as a successful researcher and professor.

To Prof. Marley for the attention, support and, especially, for the valuable discussions about fuzzy logic and artificial neural networks during the period when I was at PUC-Rio.

To the doctoral examination board, who with wisdom analyzed, rated and contributed to the writing of this thesis.

To my parents Gilberto and Sônia, examples of perseverance and dedication, to whom I owe much of what I am; I believe that even all thanks would not be enough to reward them. Thanks for the love, encouragement, and support at every stage of my life.

To my relatives and to my friends, who were able to understand my absences due to the development of this work and yet always sustained me with words of support and affection.

To Rafaella, for all the fondness, love and for supporting me during the completion of this thesis.

To my cousin João Paulo, who I consider as my brother, for the friendship and for welcoming me in Rio during the period when I was at PUC-Rio.

To Guilherme Colen for the valuable guidance in formatting this thesis.

To my students, who always motivate me to improve.

To the Industrial and Mechanical Engineering Department from Federal University of Juiz de Fora for the availability granted to accomplish this thesis. Special thanks to my friend, Prof. Fernando, for sharing his knowledge and for introducing me to the world of the railway.

To MRS Logística S.A. for giving me a great opportunity to continue my research activities and to be involved with the professional environment of the company.

To Support Program for Qualification (Proquali) of the Federal University of Juiz de Fora for the financial support given to this work.

To the Brazilian people who indirectly or directly funded this work.

"Divide each difficulty into as many parts as is feasible and necessary to resolve it."
(RENÉ DESCARTES)

"Then integrate them in their entirety. Only in this way you will have a full understanding of the process."
(AUTHOR UNKNOWN)

ABSTRACT

This thesis presents new fuzzy models applied to classification problems. With this regards, we introduce the use of set-membership concept, derived from the adaptive filter theory, into the training procedure of type-1 and singleton/non-singleton fuzzy logic systems, in order to reduce computational complexity and to increase convergence speed. Also, we present different criteria for using together with set-membership. Furthermore, we discuss the usefulness of delta rule delta, local Lipschitz estimation, variable step size and variable step size adaptive algorithms to yield additional improvement in terms of computational complexity reduction and convergence speed. Another important contribution of this thesis is to address the height type-reduction and to propose a modified version of interval singleton type-2 fuzzy logic system, so−called upper and lower singleton type-2 fuzzy logic system. The obtained results are compared with other models reported in the literature, demonstrating the effectiveness of the proposed classifiers and revealing that the proposals are able to properly handle with uncertainties associated with the measurements and with the data that are used to tune the parameters of the model. Based on data set provided by a Brazilian railway company, the models outlined above are applied in the classification of three possible faults and the normal condition of the switch machine, which is an equipment used for handling railroad switches. Finally, this thesis discusses the use of set-membership concept into the training procedure of an interval and singleton type-2 fuzzy logic system and of an upper and lower singleton type-2 fuzzy logic system, aiming to reduce computational complexity and to increase the convergence speed and the classification ratio. Also, we discuss the adoption of different criteria together with set-membership based-techniques. The performance is based on the data set composed of images provided by the same Brazilian railway company, which covers the four possible rail head defects and the normal condition of the rail head. The reported results show that the proposed models result in improved convergence speed, slightly higher classification ratio and remarkable computation complexity reduction when we limit the number of epochs for training, which may be required due to real time constraint or low computational resource availability.

Key-words: type-2 fuzzy logic systems, adaptive algorithms, classification, railway

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

AWGN additive white Gaussian noise

CGT1-FLS conjugate gradient method

DRD delta rule delta

FBF fuzzy basis function

FLS fuzzy logic system

GLCM gray level co-occurrence matrix

IST2-FLS interval singleton type-2 fuzzy logic system

LLE local Lipschitz estimation

LMF lower membership function

MLP multilayer perceptron neural network

MTBF mean time between failures

MSE mean squared error

PCA principal component analysis

PCM pulse-code modulation

PHM prognosis and health management

PSNR peak signal to noise ratio

SM set-membership

SMFLS set-membership fuzzy logic system

SVM support vector machine

TST time stopped train

VS variable step size

VSA variable step size adaptive

ULST2-FLS upper and lower type-2 fuzzy logic system

UMF upper membership function

CONTENTS

1 INTRODUCTION

Recently, in part due to environmental issues and also because of congestion on the roads, the utilization of railways has been increasing all around the world. The safe and reliable operation of trains is, therefore, becoming ever more important. To achieve this objective, most of the actuators in railways are designed to be redundant; when one of the actuators fails, the railway can still maintain its function using another actuator. Although many actuators in railways are designed in this way, there are some that cannot be designed to be redundant because of their inherent structural and mechanical nature. One of them is the switch machine.

The switch machine is the actuator that drives the switch blade from one position to the opposite position in order to offer different routes to trains. Failure in the actuator has a significant effect on train operations. If this failure occurs, it leads to a less reliable service and causes discredit to the railway company. It can also lead to more disastrous consequences. In 2002 a train derailment accident caused by poor maintenance of a switch machine occurred near Potters Bar railway station in the UK, killing seven people, as shown in Figure 1. As a result, the railway infrastructure company paid several million GBP as compensation to victims and their relatives. It is, therefore, important for all infrastructure companies to minimize the occurrence of failures in switch machine [1–3].

(a)

(b)

Figure 1: (a) An example of switch machine; (b) Train derailment near Potters Bar railway station.

Among all possible faults that can occur in an electromechanical switch machine, the three main ones are the lack of lubrication, lack of adjustment and malfunction of a component. The model developed in this work is able to identify those faults by monitoring the current of the switch machine motor. The proposal makes it possible to reduce the impact on trains operation and on preventive maintenance since interventions shall be carried out only when deviations are observed in the equipment [4].

The reliability of rail tracks needs is another important task to be paid more attention than before [5]. Rail inspection is an important task in railway maintenance. The speed and loads of trains have been increasing greatly in recent years, and these factors inevitably raise the risk of producing rail defects, as shown in Figure 2. For the safe operation of railway systems, the quality of rails must be closely and frequently monitored [6].

Figure 2: A typical rail head defect.

Railways are one of the prime modes of transportation and rail failures may cause a train to derail. Detecting defects of rail track before they cause an accident will improve the safety and reliability of rail transportation. On the other hand, with the rapid growth of high-speed railway, visual check which is done by human operators is no longer appropriate to perform rail defect detection task because of its low efficiency and check errors. Efforts on developing automatic and effective machine based intelligent defect classification techniques should be extensively investigated.

Rail inspection is the practice of examining rail tracks for defects that could lead to catastrophic accidents. According to the United States Federal Railroad Administration Office of Safety Analysis (`http://safetydata.fra.dot.gov/`), rail defects are the second leading cause of accidents on railways in the United States. Every year, North American railroads spend millions of dollars to inspect the rails. The model developed in this work is able to classify the main types of rail head defects that are commonly studied due to its severity and occurrence: cracking, flaking, head-check and spalling.

The proposed model responsible for classifying the type of fault in a switch machine or classify rail head defects shall assist to:

- reduce the number of unproductive hours in maintenance, due to the knowledge of failures or defect before moving the maintenance team to the field;
- reduce the number of recurrent preventive interventions;
- increase productivity of rail operations, given the reduction in frequency and time of operational maintenance;
- reduce the time stopped train (TST) index through the increase of the mean time between failures (MTBF) [7, 8].

Considering everything discussed so far, the Section 1.1 presents the objectives of this work and Section 1.2 addresses how this work is organized.

1.1 OBJECTIVES

The objectives of this work are listed as follows:

- To investigate the concept of set-membership (SM) [9–13], derived from adaptive filter theory [14], and combine it with delta rule delta (DRD) [10, 15–18], local Lipschitz estimation (LLE) [10,17–21], Variable Step Size (VS) [22–24] and Variable Step Size Adaptive algorithms (VSA) [25], in order to reduce the computational complexity and to increase the convergence speed during the training phase of a type-1 singleton/non-singleton fuzzy logic system (FLS) trained by Steepest Descent method.
- To discuss two FLS-based classifiers. The first one is based on a new approach, called upper and lower type-2 fuzzy logic system (ULST2-FLS). The second one it is based on interval singleton type-2 fuzzy logic system (IST2-FLS), discussed in [26–29].
- To apply the concept of SM with IST2-FLS and ULST2-FLS, together with DRD, LLE, VS and VSA to come up with additional computational complexity savings during the training phase.

1.2 WORK ORGANIZATION

This work is organized as follows:

- Chapter 2 introduces the use of SM concept into the training procedure of type-1 and singleton/non-singleton fuzzy logic system (FLS), in order to reduce computational complexity and to increase convergence speed during the classification of faults in a switch machine.

- Chapter 3 address two FLS-based classifiers for fault classification in a switch machine. The first one is based on a new approach, called ULST2-FLS, reducing the complexity of the training phase. The second one is based on IST2-FLS.

- Chapter 4 focuses on the classification of rail head defects, through images acquired by a rail inspection vehicle. With this regards, we discuss the use of SM concept into the training procedure of an IST2-FLS and ULST2-FLS, aiming to reduce computational complexity and to increase the convergence speed and the classification ratio.

- Chapter 5 presents the main conclusions of this thesis.

2 SET-MEMBERSHIP TYPE-1 FUZZY LOGIC SYSTEM APPLIED TO FAULT CLASSIFICATION IN A SWITCH MACHINE

A railroad switch is an electromechanical equipment that guides railway trains from one track to another, such as at a railway junction, while a switch machine is an equipment used for handling railroad switches. The increase of Brazilian railroad sector, which results in more use of switch machines, is an important problem to deal with, because real time operation, monitoring, and diagnosis of switch machine is of vital importance, especially with respect to predictive maintenance to avoid accidents and losses [4, 30].

Recently, there has been increased interest among both transportation researchers and practitioners in exploiting the feasibility of applying computational intelligence paradigms to address critical problems in order to improve the efficiency, safety, and environmental-compatibility of transportation systems [31–33].

Research on switch machines has been in evidence in recent years, since they are important equipments that must be constantly monitored to avoid accidents. For example, [34] describes a strategy and a technical architecture for prognosis and health management (PHM) of a switch machine. Feature extraction techniques and principal component analysis (PCA) have been used in the assessment of the machines' health. [35] makes use of a support vector machine (SVM) in an online switch machine condition monitoring system based on current measurements to detect faults at their earliest stage or even before they occur. In [3], wavelet transform together with SVM showed that electrical active power can be a parameter for condition monitoring of switch machines. Moreover, [36] applies expert system for fault identification in a switch machine, when the failure modes are not easily separable. In addition, [37] addresses a technique for detecting gradual failure in a switch machine, in which a Kalman filter was applied for pre-processing signal signature. Other works such as [38] and [39] present alternative techniques related to unsupervised and semi-supervised fault detection and identification in industrial plants based on the analysis of the control and error signals. Among all possible faults that can occur in a switch machine, the three main ones are lack of lubrication, lack of adjustment and malfunction of a component.

A common aspect of all previous works (see references [34]-[39]) is the focus on the detection of the existence of a fault, and not on classify different types of faults, such as those contemplated in this work. It is important to emphasize that the classification of specific types of faults (lack of lubrication, lack of adjustment and malfunction of a component) and the normal condition by monitoring the motor current of switch machine in the railroad is novel and it was first time introduced in the literature by the authors in [30] and [4].

Combinations of feature extraction technique based on higher-order statistics [40],

[41], feature selection technique based on Fisher's discriminant ratio [40], and three classifiers (Bayes theory [42], multilayer perceptron neural network (MLP) neural network [43], and type-1 and singleton fuzzy logic system (FLS) [28, 30]) showed performance improvement when applied to a data set composed of measured current signals. Later, [4], using the same data set, showed that the type-1 and singleton FLS trained by the conjugate gradient method can offer higher convergence speed and classification ratio than the steepest descent method.

This paper addresses the computational complexity reduction during the training phase of type-1 and singleton/non-singleton FLS as well as their performances to handle the classification problem. The proposal is capable of classifying those faults by monitoring the motor current of the switch machine.

The main contributions of this Chapter are summarized as follows:

- We introduce the concept of set-membership (SM) [9–13] derived from adaptive filter theory [14], and combine it with type-1 FLS trained by steepest descent method, in order to reduce the computational complexity and to increase the convergence speed during the training phase. Due to the decrease in computational complexity, we emphasize that the adoption of SM combined with FLS demands low-cost (low-speed) processors, which is something very important to come up with green technologies for railroad sector.

- We combine SM and FLS together with delta rule delta (DRD) [10, 15–18], local Lipschitz estimation (LLE) [10, 17–21] variable step size (VS) [22–24] and variable step size adaptive (VSA) [25] to come up with additional increase in the convergence speed during the training.

- We present performance analyzes in terms of classification ratio, convergence speed, and computational complexity reduction by using a data set constituted by a measured signal from switch machine. After that, we show a comparative analysis of the proposed models based on previous techniques (Bayes theory, MLP neural network, and type-1 and singleton FLS trained by steepest descent method, type-1 and singleton FLS trained by the conjugate gradient method).

Our major conclusions are as follows:

- Numerical results show that the type-1 and singleton/non-singleton FLS trained by the proposals converges faster than those previously discussed in [30] and [4].

- The SM concept combined with DRD, LLE, VS and VSA strongly contributed to reducing computational complexity during the training phase of type-1 and

singleton/non-singleton, under a limited number of epochs. They reached percentages of computational complexity reduction of 96.3% and 93.6% for type-1 and singleton and non-singleton FLS, respectively;

- Classification ratio yielded by the proposals are higher than those obtained with the previous techniques (Bayes, multilayer perceptron neural network, type-1 and singleton FLS trained by steepest descent method and type-1 and singleton FLS trained by conjugate gradient method).

The rest of the Chapter is organized as follows: Section 2.1 deals with the problem formulation. Section 2.2 aims to discuss the concept of SM combined with FLS. Section 2.3 discusses the results of computer simulations. Section 2.4 states the main conclusions regarding this Chapter.

2.1 PROBLEM FORMULATION

The switch machine is the actuator that drives the switch blade from one position to the opposite position in order to offer different routes to trains. Failure in the actuator has a significant effect on train operations. Britain's railway infrastructure operator, Network Rail (`https://www.networkrail.co.uk/`), was responsible for approximately 14 million minutes of train delay in 2002-2003, costing approximately 560 million GBP [44, 45]. Switch machine is the main component of railway infrastructure that affects the availability of the system [46]. For example in England, 3.4 million GBP is spent every year for the maintenance of switch machines for 1000 km of railways [44, 46]. Consequently, the classification of failures in these types of equipment is critical for managing and monitoring railway operational conditions. Therefore, the investigation of pattern recognition-based technique, which is capable of automatically identifying failures in switch machines, is of great importance to introduce new generations of solutions to increase system reliability.

The classification of faults in switch machines is one key purpose to assist the railroad company in verifying the conditions of this equipment. It is a part of a multi-years project that has the objective of introducing this functionality in 100 of 624 switch machines belonging to the company in the Brazilian railway sector, named MRS Logística S.A. (https://www.mrs.com.br/). The proposed model makes it possible to reduce the impact on trains operation and on preventive maintenance since interventions shall be carried out only when deviations are observed in the equipment. According to [46], the classification of faults, together with a solution that includes the application of advanced electronics, sensors and communications technology can be used to develop a new approach to switch machine monitoring that can increase reliability while, at the same time, reducing recurrent costs.

Let $\mathbf{r} \in \mathbb{R}^{N\times 1}$ a vector constituted by a sample signal of the motor current of the switch machine. Figure 3 shows the paradigm used for the classification of events. In the block "Feature Extraction", $\mathbf{p}_l$, $\mathbf{p}_a$, $\mathbf{p}_c$ and $\mathbf{p}_n$ refer respectively to lack of lubrication, lack of adjustment, malfunction of a component and normal operation feature vectors. The block "Feature Selection" provides selected features $\mathbf{K}_{p_l}$, $\mathbf{K}_{p_a}$, $\mathbf{K}_{p_c}$ and $\mathbf{K}_{p_n}$ vectors (with remarkable dimensionality reduction) from $\mathbf{p}_l$, $\mathbf{p}_a$, $\mathbf{p}_c$ and $\mathbf{p}_n$, respectively. The block "Classification" applies one of the classification techniques discussed in this work to obtain the output vector $\mathbf{s}$, thereby identifying the type of fault. We use four independent classifiers, one for each possible event.

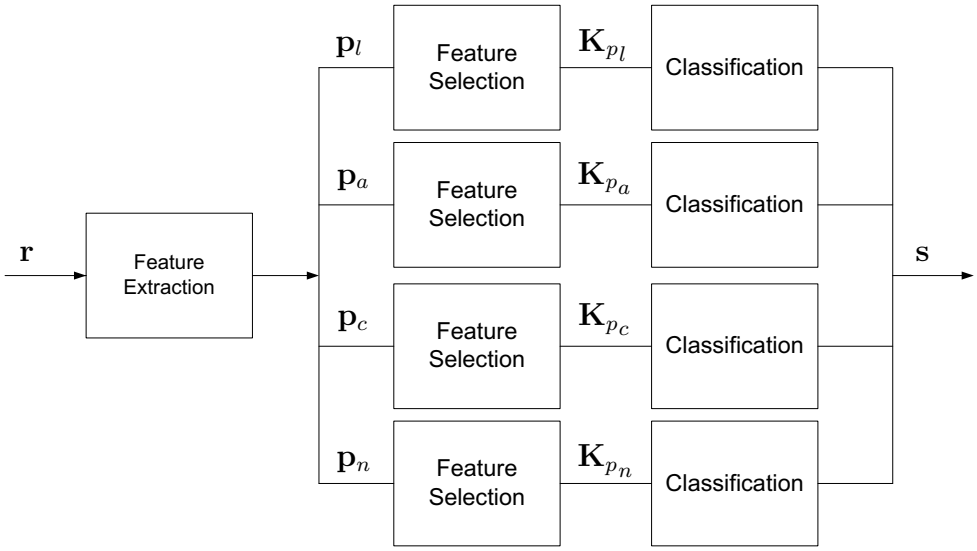


Figure 3: Block diagram of the scheme for classification of events.

The classification of events in the vector $\mathbf{r}$ can be formulated as a simple decision between hypotheses related to the occurrence of the events covered in this work, as shown below:

$$\begin{aligned} \mathcal{H}_{r,0} &: \mathbf{r} = \mathbf{r}_{lub}, \\ \mathcal{H}_{r,1} &: \mathbf{r} = \mathbf{r}_{adj}, \\ \mathcal{H}_{r,2} &: \mathbf{r} = \mathbf{r}_{comp}, \\ \mathcal{H}_{r,3} &: \mathbf{r} = \mathbf{r}_{norm}. \end{aligned} \tag{2.1}$$

in which $\mathbf{r}_{lub}$, $\mathbf{r}_{adj}$, $\mathbf{r}_{comp}$ and $\mathbf{r}_{norm}$ denotes the lubrication, adjustment, component and normal operation events, respectively.

Regarding monitoring and diagnosis, several works discussed and analyzed the performance of techniques for classifying typical faults that can occur in a switch machine. The drawback associated with former techniques is the high computational complexity for training. Considering a limited number of epochs, the training with reduced computational complexity, high accuracy, and high convergence speed consists of a hard task. The hardness is associated with the limited hardware resources is worse and real-time constraint.

Four independent classifiers are developed, one for each possible event, since multiple events can occur simultaneously. The characteristics of each block in Fig. 3 will be presented in the following sections.

2.1.1 Feature extraction based on higher-order statistics

Some contributions as [40], related to problems of detection, classification and identification of disturbances in electrical systems, have reported significant results obtained through the use of higher-order statistics (HOS). This is because HOS-based techniques are better suited to non-Gaussian processes and nonlinear systems when compared to those that employ second-order statistics.

Given a stationary random process $\{z[n]\}$, such that $E\{z[n]\} = 0$, the cumulants of the second, third and fourth order can be computed as [41]:

$$c_{2,z}[i] = E\{z[n]z[n+i]\}, \tag{2.2}$$

$$c_{3,z}[i] = E\left\{z[n]z^2[n+i]\right\}, \tag{2.3}$$

and

$$c_{4,z}[i] = E\left\{z[n]z^3[n+i]\right\} - 3c_{2,z}[i]c_{2,z}[0], \tag{2.4}$$

where $E\{\cdot\}$ denotes the expected value operator and i is the ith lag.

Assuming that $\{z[n]\}$ is an L-length random process, then equations (2.2) - (2.4) can be stochastically approximated by the following expressions

$$\hat{c}_{2,z}[i] \cong \frac{2}{L}\sum_{n=0}^{L/2-1} z[n]z[n+i], \tag{2.5}$$

$$\hat{c}_{3,z}[i] \cong \frac{2}{L}\sum_{n=0}^{L/2-1} z[n]z^2[n+i], \tag{2.6}$$

and

$$\begin{aligned}\hat{c}_{4,z}[i] \cong\ & \frac{2}{L}\sum_{n=0}^{L/2-1} z[n]z^3[n+i] - \\ & -\frac{12}{L^2}\sum_{n=0}^{L/2-1} z[n]z[n+i]\sum_{n=0}^{L/2-1} z^2[n],\end{aligned} \tag{2.7}$$

where $i = 0, 1, ..., L/2-1$.

An alternative approach would be [47]:

$$\tilde{c}_{2,z}[i] \cong \frac{1}{L}\sum_{n=0}^{L-1} z[n]z\left[\mathrm{mod}(n+i, L)\right] \tag{2.8}$$

$$\tilde{c}_{3,z}[i] \cong \frac{1}{L}\sum_{n=0}^{L-1} z[n]z^2\,[\mathrm{mod}(n+i,L)] \tag{2.9}$$

and

$$\begin{aligned}\tilde{c}_{4,z}[i] \cong\ & \frac{1}{L}\sum_{n=0}^{L-1} z[n]z^3\,[\mathrm{mod}(n+i,L)] \\ & -\frac{3}{L^2}\sum_{n=0}^{L-1} z[n]z\,[\mathrm{mod}(n+i,L)]\sum_{n=0}^{L-1} z^2[n],\end{aligned} \tag{2.10}$$

where $i = 0, 1, ..., L-1$ and $\mathrm{mod}(\cdot)$ is the modulus operator.

Thus, for each $\mathbf{s}$ it is possible to obtain a feature vector given by

$$\mathbf{p}_{\mathcal{H}} = [\hat{\mathbf{c}}_{2,z}^{\,T}\ \tilde{\mathbf{c}}_{2,z}^{\,T}\ \hat{\mathbf{c}}_{3,z}^{\,T}\ \tilde{\mathbf{c}}_{3,z}^{\,T}\ \hat{\mathbf{c}}_{4,z}^{\,T}\ \tilde{\mathbf{c}}_{4,z}^{\,T}]^T, \quad \mathcal{H} = 0, 1, \tag{2.11}$$

where $\mathcal{H} = 0$ is the class without disturbance, whereas $\mathcal{H} = 1$ is a class with disturbance, $\hat{\mathbf{c}}_{2,z} = [\hat{c}_{2,z}(0), \cdots,$
$\hat{c}_{2,z}(L/2-1)]^T$, $\tilde{\mathbf{c}}_{2,z} = [\tilde{c}_{2,z}(0), \cdots, \tilde{c}_{2,z}(L-1)]^T$, $\hat{\mathbf{c}}_{3,z} = [\hat{c}_{3,z}(0), \cdots, \hat{c}_{3,z}(L/2-1)]^T$, $\tilde{\mathbf{c}}_{3,z} = [\tilde{c}_{3,z}(0), \cdots, \tilde{c}_{3,z}(L-1)]^T$, $\hat{\mathbf{c}}_{4,z} = [\hat{c}_{4,z}(0), \cdots, \hat{c}_{4,z}(L/2-1)]^T$ and $\tilde{\mathbf{c}}_{4,z} = [\tilde{c}_{4,z}(0), \cdots, \tilde{c}_{4,z}(L-1)]^T$.

2.1.2 Feature selection based on Fisher's discriminant ratio

Recent works, such as [40], have used Fisher's discriminant ratio (FDR) for feature selection, since it is simple and provides satisfactory results. In this methodology, K_p features are selected in order to build the vector $\mathbf{p}_{\mathcal{H}}$ with length $\frac{9L}{2}$, due to $\hat{\mathbf{c}}_{2,z}$, $\hat{\mathbf{c}}_{3,z}$ and $\hat{\mathbf{c}}_{4,z}$ having length $L/2$ and $\tilde{\mathbf{c}}_{2,z}$, $\tilde{\mathbf{c}}_{3,z}$, $\tilde{\mathbf{c}}_{4,z}$ having length L. This makes the classifier complexity quite low. This pre-processing step is always executed before the classification stage. For a problem involving only two distinct classes associated with vectors with elements normally distributed and uncorrelated, it is defined as [42]:

$$\mathbf{F}_{FDR} = \Lambda_{\mu_0,\mu_1}\Lambda_{\sigma}^{-1}, \tag{2.12}$$

where $\Lambda_\sigma = diag\{\sigma_{0,0}^2 + \sigma_{1,0}^2, \sigma_{0,1}^2 + \sigma_{1,1}^2, \ldots, \sigma_{0,\frac{9L}{2}-1}^2 + \sigma_{1,\frac{9L}{2}-1}^2\}$ is a diagonal matrix formed by the vector covariance associated with each class, and $\Lambda_{\mu_0,\mu_1} = diag\{(\mu_{0,0}-\mu_{1,0})^2, (\mu_{0,1}-\mu_{1,1})^2, \ldots, (\mu_{0,\frac{9L}{2}-1} - \mu_{1,\frac{9L}{2}-1})^2\}$ is the diagonal matrix composed by their average vectors.

Note that $\mathbf{v}_{FDR} \in R^{\frac{9L}{2}\times 1}$ is a vector of elements from the main diagonal of $\mathbf{F}_{FDR}$, such that $v_{FDR}(0) \geq v_{FDR}(1) \geq \ldots \geq v_{FDR}(\frac{9L}{2}-1)$. The K_p selected features, corresponding to the K_p first elements of the vector $\mathbf{v}_{FDR}$, will form vectors $\mathbf{K}_{p_l}$, $\mathbf{K}_{p_a}$, $\mathbf{K}_{p_c}$ or $\mathbf{K}_{p_n}$.

2.1.3 Classifiers

The classifiers used in this work are based on type-1 and singleton FLS and type-1 and non-singleton FLS. We followed [28][30] and [4] to implement the former FLS. Next section develops the latter FLS.

2.2 ADAPTING THE CONCEPT OF SET-MEMBERSHIP FOR FLS

Considering non-singleton fuzzification, max-product composition, product implication, height defuzzifier and gaussian membership functions, the output of a type-1 and non-singleton FLS is [28]

$$f_{ns}(\mathbf{x}) = \sum_{l=1}^{M} \theta_l \phi_l(\mathbf{x}), \tag{2.13}$$

where $\phi_l(\mathbf{x})$ is called fuzzy basis function (FBF) [28] and is given by

$$\phi_l(\mathbf{x}) = \frac{\prod_{k=1}^{p} \exp\left(-\frac{1}{2} \frac{\left(x_k^{(q)} - m_{F_k^l}\right)^2}{\sigma_{F_k^l}^2 + \sigma_X^2} \right)}{\sum_{l=1}^{M} \left[\prod_{k=1}^{p} \exp\left(-\frac{1}{2} \frac{\left(x_k^{(q)} - m_{F_k^l}\right)^2}{\sigma_{F_k^l}^2 + \sigma_X^2} \right) \right]}, \tag{2.14}$$

where $\mathbf{x} \in \mathbb{R}^{K_p}$ is the vector constituted by K_p elements, $\prod$ denotes the product operator, $m_{F_k^l}$ and $\sigma_{F_k^l}$ are the mean and standard deviation associated to the k-th input feature of the l-th rule. θ_l is the weight associated with the l-th rule, $l = 1, \ldots, M$. σ_X is the standard deviation associated to each input membership function. The subscript "ns" in $f_{ns}(\mathbf{x})$ informs that this is a non-singleton FLS.

Consider a set of input-output pairs $(\mathbf{x}^{(q)} : y^{(q)})$, where q denotes the qth iteration. To obtain the suitable parameters of $f_{ns}(\mathbf{x})$ for our classification problem, the task is to minimize the following cost function [28]:

$$J(\mathbf{w}^{(q)}) = \frac{1}{2} \left[f_{ns}(\mathbf{x}^{(q)}) - y^{(q)} \right]^2. \tag{2.15}$$

By applying steepest descent method, we obtain

$$\begin{aligned} m_{F_k^l}(q+1) = m_{F_k^l}(q) - \alpha \left[f_{ns}(\mathbf{x}^{(q)}) - y^{(i)} \right] \times \\ \left[\theta_l(q) - f_{ns}(\mathbf{x}^{(q)}) \right] \left[\frac{x_k^{(q)} - m_{F_k^l}(q)}{\sigma_X^2(q) + \sigma_{F_k^l}^2(q)} \right] \phi_l(\mathbf{x}^{(q)}), \end{aligned} \tag{2.16}$$

$$\theta_l(q+1) = \theta_l(q) - \alpha \left[f_{ns}(\mathbf{x}^{(q)}) - y^{(q)} \right] \phi_l(\mathbf{x}^{(q)}), \tag{2.17}$$

$$\begin{aligned} \sigma_{F_k^l}(q+1) = \sigma_{F_k^l}(q) - \alpha \left[f_{ns}(\mathbf{x}^{(q)}) - y^{(q)} \right] \times \\ \left[\theta_l(q) - f_{ns}(\mathbf{x}^{(q)}) \right] \sigma_{F_i^l}(q) \left[\frac{x_k^{(q)} - m_{F_k^l}(q)}{\sigma_X^2(q) + \sigma_{F_k^l}^2(i)} \right]^2 \phi_l(\mathbf{x}^{(q)}), \end{aligned} \tag{2.18}$$

and

$$\begin{aligned} \sigma_X(q+1) = \sigma_X(q) - \alpha \left[f_{ns}(\mathbf{x}^{(q)}) - y^{(q)} \right] \times \\ \left[\theta_l(q) - f_{ns}(\mathbf{x}^{(q)}) \right] \sigma_X(q) \left[\frac{x_k^{(q)} - m_{F_k^l}(q)}{\sigma_X^2(q) + \sigma_{F_k^l}^2(q)} \right]^2 \phi_l(\mathbf{x}^{(q)}). \end{aligned} \tag{2.19}$$

Note that $\alpha \in \mathbb{R}|0 < \alpha < 1$ is the step size value used to update parameters of the type-1 and non-singleton FLS.

From (2.15), we know that

$$\begin{aligned}\mathbf{w}^{(q)} = \; & \Big[m_{F_1^1}(q), \ldots, m_{F_p^1}(q), \ldots, m_{F_1^M}(q), \ldots, m_{F_p^M}(q), \\ & \sigma_{F_1^1}(q), \ldots, \sigma_{F_p^1}(q), \ldots, \sigma_{F_1^M}(q), \ldots, \sigma_{F_p^M}(q), \\ & \theta_1(q), \ldots, \theta_M(q), \sigma_X(q)\Big]^T\end{aligned} \tag{2.20}$$

and

$$\begin{aligned}& \nabla J(\mathbf{w}^{(q)}) = \\ & \Big[\nabla_{m_{F_1^1}(q)} J(\mathbf{w}^{(q)}), \ldots, \nabla_{m_{F_P^1}(q)} J(\mathbf{w}^{(q)}), \ldots, \\ & \nabla_{m_{F_1^M}(q)} J(\mathbf{w}^{(q)}), \ldots, \nabla_{m_{F_P^M}(q)} J(\mathbf{w}^{(q)}), \\ & \nabla_{\sigma_{F_1^1}(q)} J(\mathbf{w}^{(q)}), \ldots, \nabla_{\sigma_{F_P^1}(q)} J(\mathbf{w}^{(q)}), \ldots, \\ & \nabla_{\sigma_{F_1^M}(q)} J(\mathbf{w}^{(q)}), \ldots, \nabla_{\sigma_{F_P^M}(q)} J(\mathbf{w}^{(q)}), \\ & \nabla_{\theta_1(q)} J(\mathbf{w}^{(q)}), \ldots, \nabla_{\theta_M(q)} J(\mathbf{w}^{(q)}), \nabla_{\sigma_X(q)} J(\mathbf{w}^{(q)}\Big]^T,\end{aligned} \tag{2.21}$$

denote, respectively, the parameters vector and the gradient vector of a type-1 and non-singleton FLS.

The first-order derivative of $J(\mathbf{w}^{(q)})$ with regard to parameters $m_{F_k^l}(q)$, $\sigma_{F_k^l}(q)$, $\theta_l(q)$ and $\sigma_X(q)$ are

$$\begin{aligned}\nabla_{m_{F_k^l}(q)} J(\mathbf{w}^{(q)}) = & \Big[f_{ns}(\mathbf{x}^{(q)}) - y^{(q)}\Big] \times \\ & \Big[\theta_l(q) - f_{ns}(\mathbf{x}^{(q)})\Big] \phi_l(\mathbf{x}^{(q)}) a_{F_k^l}(q),\end{aligned} \tag{2.22}$$

$$\begin{aligned}\nabla_{\sigma_{F_k^l}(q)} J(\mathbf{w}^{(q)}) = & \Big[f_{ns}(\mathbf{x}^{(q)}) - y^{(q)}\Big] \times \\ & \Big[\theta_l(q) - f_{ns}(\mathbf{x}^{(q)})\Big] \phi_l(\mathbf{x}^{(q)}) b_{F_k^l}(q),\end{aligned} \tag{2.23}$$

$$\nabla_{\theta_l(q)} J(\mathbf{w}^{(q)}) = \Big[f_{ns}(\mathbf{x}^{(q)}) - y^{(q)}\Big] \phi_l(\mathbf{x}^{(q)}), \tag{2.24}$$

and

$$\begin{aligned}\nabla_{\sigma_X(q)} J(\mathbf{w}^{(q)}) = & \Big[f_{ns}(\mathbf{x}^{(q)}) - y^{(q)}\Big] \times \\ & \Big[\theta_l(q) - f_{ns}(\mathbf{x}^{(q)})\Big] \phi_l(\mathbf{x}^{(q)}) c_{F_k^l}(q).\end{aligned} \tag{2.25}$$

Note that

$$a_{F_k^l}(q) = \frac{x_k^{(q)} - m_{F_k^l}(q)}{\sigma^2_{F_k^l}(q) + \sigma^2_X(q)}, \tag{2.26}$$

$$b_{F_k^l}(q) = \frac{\Big(x_k^{(q)} - m_{F_k^l}(q)\Big)^2}{\Big(\sigma^2_{F_k^l}(q) + \sigma^2_X(q)\Big)^2} \sigma_{F_k^l}(q) \tag{2.27}$$

and

$$c_{F_k^l}(q) = \frac{\Big(x_k^{(q)} - m_{F_k^l}(q)\Big)^2}{\Big(\sigma^2_{F_k^l}(q) + \sigma^2_X(q)\Big)^2} \sigma_X(q). \tag{2.28}$$

The equations for training type-1 and non-singleton differ from those applied to type-1 and singleton due to the presence of σ_X, which model the uncertainty associated with input.

2.2.1 The Concept of Set-Membership

The SM concept is a framework applicable to adaptive-filtering problems, as described in [12]. By introducing this concept in the FLS, we state that the objective of the set-membership fuzzy logic system (SMFLS) is to design $\mathbf{w}^{(q)}$ such that the magnitude of $\left|e^{(q)}\right| = \left|f_{ns}(\mathbf{x}^{(q)}) - y^{(q)}\right|$ is upper bounded by a prescribed constraint $\bar{\gamma} \in \mathbb{R}$, where $|\ |$ denotes the modulus operator.

Assuming that $\overline{S}$ denotes the set of all possible input-desired data pairs $(\mathbf{x}^{(1)} : y^{(1)}), (\mathbf{x}^{(2)} : y^{(2)}), ..., (\mathbf{x}^{(N)} : y^{(N)})$ of interest, it is possible to define ψ as the set of all possible vectors $\mathbf{w}^{(q)}$ leading to $e^{(q)}$, whose magnitudes are bounded by $\bar{\gamma}$ whenever $(\mathbf{x}^{(1)} : y^{(1)}), (\mathbf{x}^{(2)} : y^{(2)}), ..., (\mathbf{x}^{(N)} : y^{(N)}) \in \bar{S}$.

Let's define $\mathcal{H}(q)$ as the set containing all vectors $\mathbf{w}^{(q)}$ such that $e^{(q)}$ is upper bounded in magnitude by $\bar{\gamma}$. That is,

$$\mathcal{H}(q) = \left\{\mathbf{w}^{(q)} \in \mathbb{R}^{N+1} : \left|f_{ns}(\mathbf{x}^{(q)}) - y^{(q)}\right| \leq \bar{\gamma}\right\}, \tag{2.29}$$

. The set $\mathcal{H}(q)$ is usually referred to as the constraint set. The boundaries of $\mathcal{H}(q)$ are hypersurfaces. For the two-dimensional case, where the coefficient vector has two elements, $\mathcal{H}(q)$ comprises the region between the curves where $f_{ns}(\mathbf{x}^{(q)}) - y^{(q)} = \pm\bar{\gamma}$ as depicted in Figure 4.

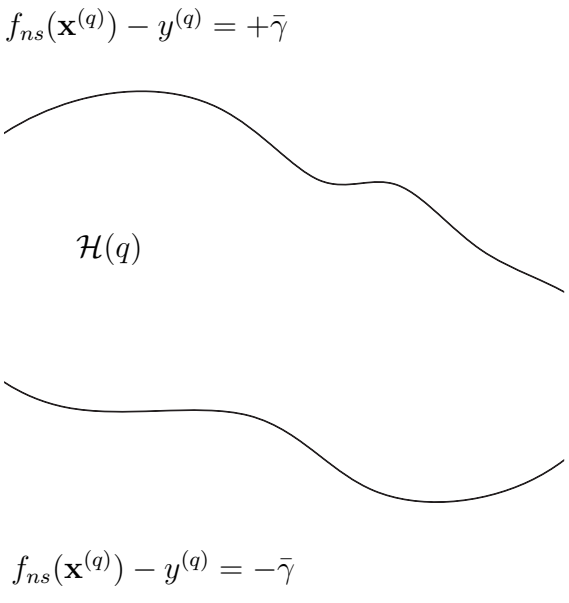


Figure 4: Constraint set in $\mathbf{w}^{(q)}$ in a two-dimension.

Since for each data pair, there is an associated constraint set, the intersection of the constraint sets over all the available iteration is called the exact membership set $\psi(q)$,

formally expressed by

$$\psi(q) = \bigcap_{q=1}^{N} \mathcal{H}(q), \tag{2.30}$$

in which $\cap$ denotes intersection operator.

The set $\psi(q)$ represents a specific area in the parameter space whose location is the main purpose to be accomplished.

For a set of data pairs including substantial innovation, the specific region in $\mathbf{w}^{(q)}$, $\psi(q)$, should become small. This property usually occurs after a large number of iterations, when most likely $\psi(q) = \psi(q-1)$, since $\psi(q) = \psi(q-1)$ is entirely contained in the constraint set $\mathcal{H}(q)$ as depicted in Figure 5. In this case, the parameters of SMFLS do not need updating because the current membership set is totally inside the constraint set $\mathcal{H}(q)$.

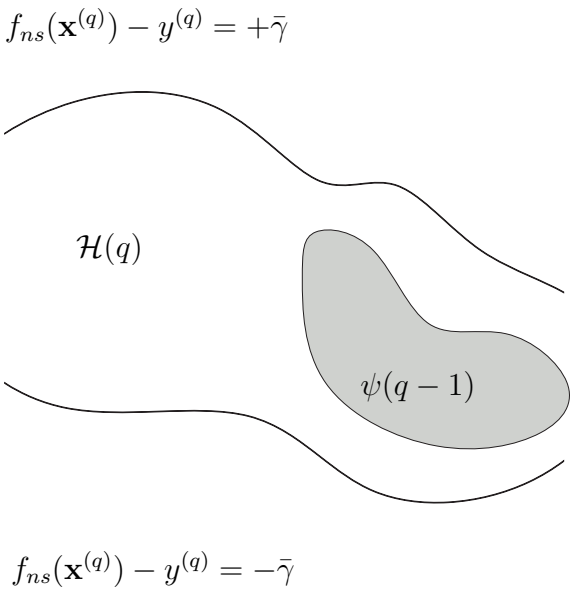


Figure 5: Exact membership set, $\psi(q-1)$, contained in the constraint set, $\psi(q-1) \subset \mathcal{H}(q)$.

The selective updating of the SMFLS brings about opportunities for computational complexity reductions, so crucial in engineering applications such as Big Data and periodic training phase to include new signatures. It is worth stating that in the early iterations it is highly possible that the constraint set reduces the size of the membership-set hypersurface as illustrates in Figure 6. In this case, the exact membership set $\psi(q-1)$ is not part of $\mathcal{H}(q)$.

The key idea of the SMFLS is to perform a test to verify if $\mathbf{w}^{(q)}$ lies outside the constraint set $\mathcal{H}(q)$, i.e., $\left|f_{ns}(\mathbf{x}^{(q)}) - y^{(q)}\right| > \bar{\gamma}$, if we consider the type-1 and singleton FLS. If the modulus of the error signal is greater than the specified bound, the vector $\mathbf{w}^{(q+1)}$ will be updated to the closest boundary of $\mathcal{H}(q)$. The update of $\mathbf{w}^{(q)}$ is based on an orthogonal projection onto the closest boundary of $\mathcal{H}(q)$. Figure 7 illustrates the updating procedure of the SMFLS.

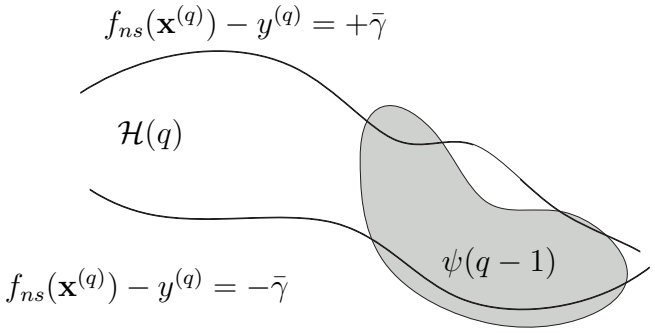


Figure 6: Exact membership set $\psi(q-1)$ not contained in the constraint set $\psi(q-1) \not\subseteq \mathcal{H}(q)$.

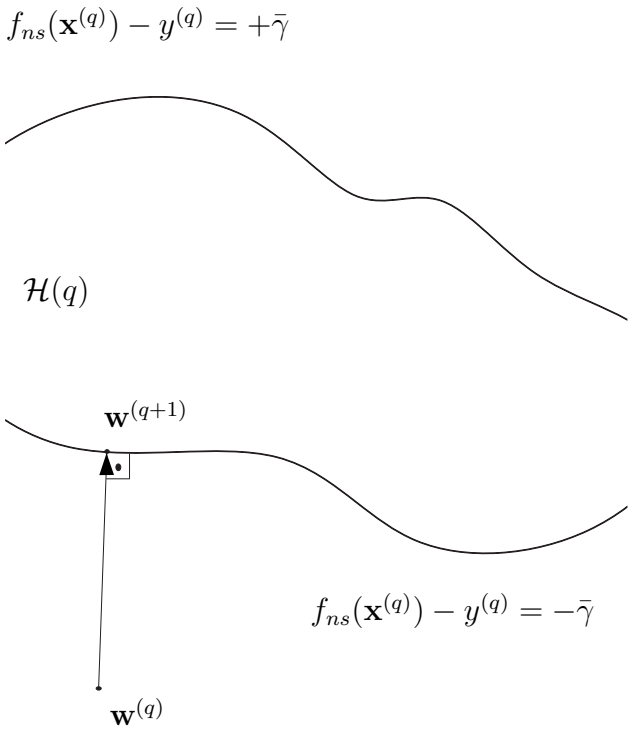


Figure 7: Parameters vector updating for the SMFLS algorithm.

Following [12], the SMFLS consists in the adoption of variable step size $\mu_e(q)$, given by

$$\mu_e(q) = \begin{cases} 1 - \frac{\bar{\gamma}}{\left|e^{(q)}\right|}, & \text{if } \left|e^{(q)}\right| > \bar{\gamma} \\ 0, & \text{otherwise.} \end{cases} \tag{2.31}$$

Even through $\mu_e(q)$ has been successfully employed in adaptive filtering, we point out that there are other criteria to substitute (2.31). Few of them are suggested as follows:

$$\mu_{MSE}(q+1) = \begin{cases} 1 - \frac{\bar{\gamma}}{\sum_{q=1}^{N} \frac{1}{N}\left[f_{ns}(\mathbf{x}^{(q)}) - y^{(q)}\right]^2}, & \text{if } \sum_{q=1}^{N} \frac{1}{N}\left[f_{ns}(\mathbf{x}^{(q)}) - y^{(q)}\right]^2 > \bar{\gamma} \\ 0, & \text{otherwise} \end{cases}, \tag{2.32}$$

$$\mu_{l2}(q+1) = \begin{cases} 1 - \frac{\bar{\gamma}}{\left\|\nabla J(\mathbf{w}^{(q)})\right\|^2}, & \text{if } \left\|\nabla J(\mathbf{w}^{(q)})\right\|^2 > \bar{\gamma} \\ 0, & \text{otherwise} \end{cases}, \tag{2.33}$$

$$\mu_{max\nabla}(q+1) = \begin{cases} 1 - \frac{\bar{\gamma}}{\max_i \left\|(\nabla J(\mathbf{w}^{(q)})\right\|}, & \text{if } \max_i \left\|(\nabla J(\mathbf{w}^{(q)})\right\| > \bar{\gamma} \\ 0, & \text{otherwise} \end{cases} \tag{2.34}$$

We adopt the changes proposed in (2.32)-(2.34) in an attempt to achieve higher computational complexity reductions. The mean squared error (MSE) gives more importance on large values of errors than small ones, because the final result of $\left[f_{ns}(\mathbf{x}^{(q)}) - y^{(q)}\right]$ is squared. Additionally, (2.32) emphasizes outliers that are inconsistent with the mean of data set. In (2.33), the use of $\left\|\nabla J(\mathbf{w}^{(q)})\right\|^2$ implies that the energy of the gradient vector controls the step size, which is an interesting approach to be taken into account. Finally, the criterion based on $\max_i \left\|(\nabla J(\mathbf{w}^{(q)})\right\| > \bar{\gamma}$, where i denotes the i-th element of vector $\nabla J(\mathbf{w}^{(q)})$, in (2.34), establish that the choice by the highest absolute value of a element of $\nabla J(\mathbf{w}^{(q)})$ can control the update rule. It is a more sensitive approach than that one based on the energy, see (2.33).

To facilitate the understanding, acronyms are adopted. Thus, we have the following situations: SMFLS refers to (2.31); MSE SMFLS refers to (2.32); $\left\|\nabla J(\mathbf{w}^{(q)})\right\|^2$ SMFLS indicates (2.33) and $\max(\nabla J(\mathbf{w}^{(q)})$ SMFLS specifies (2.34).

2.2.2 Variable Step Size Adaptive Techniques with Set-Membership

The variable step size techniques considered here use a distinct step size for updating each parameter of the SMFLS. The first technique, discussed in [22] and referred to as the variable step size (VS), increases the step size by multiplying the current value of step size by a factor α_{VS}, or decreases the current value of step size by dividing it by α_{VS} as expressed by

$$\mu_{VS}(q+1) = \begin{cases} \mu_{VS}(q) \times \alpha_{VS}, & \text{if } \left|e^{(q)}\right| > \bar{\gamma} \\ \mu_{VS}(q)/\alpha_{VS}, & \text{otherwise} \end{cases}. \tag{2.35}$$

We adopt the acronym VS SMFLS for this technique.

The second technique, a modified version of VS SMFLS, consists in not updating

the step size when $\left|e^{(q)}\right| \leq \bar{\gamma}$. In other words,

$$\mu_{MVS}(q+1) = \begin{cases} \mu_{MVS}(q) \times \alpha_{VS}, & \text{if } \left|e^{(q)}\right| > \bar{\gamma} \\ 0, & \text{otherwise} \end{cases} . \tag{2.36}$$

For (2.36), we apply the abbreviation MVS SMFLS to mean the modified VS SMFLS.

The VSA technique [25] adapts the step size by adding or subtracting a constant value. The use of SM together with VSA was named VSA SMFLS. The update of step size is given by

$$\mu_{VSA}(q+1) = \begin{cases} \mu_{VSA}(q) + \alpha_{VSA}, & \text{if } \left|e^{(q)}\right| > \bar{\gamma} \\ \mu_{VSA}(q) - \alpha_{VSA}, & \text{otherwise} \end{cases} \tag{2.37}$$

A modified version of VSA SMFLS, called MVSA SMFLS, applies the following rule:

$$\mu_{MVSA}(q+1) = \begin{cases} \mu_{MVSA}(q) + \alpha_{VSA}, & \text{if } \left|e^{(q)}\right| > \bar{\gamma} \\ 0, & \text{otherwise} \end{cases} \tag{2.38}$$

The main idea behind the MVS SMFLS and the MVSA SMFLS is to provide a decrease in the computational complexity and to maintain the same accuracy by not updating the step size when $\left|e^{(q)}\right| > \bar{\gamma}$. We adopt (2.35)-(2.38) because they are comparatively easier to implement than the previous approaches which are applied to update the parameters of SMFLS.

Finally, the algorithm in Figure 8 portrays the training of a type-1 and non-singleton FLS when the SMFLS is applied. If we want to use another SMFLS, the updating rule must be replaced. The similar algorithm applies to type-1 and singleton FLS, as discussed in [4].

Input
M: Number of rules of the type-1 and non-singleton FLS;
p: Number of parameters of the type-1 and non-singleton FLS;
μ: Step size;
nn: Number of epochs;
$nn_{\max}$: Maximum number of epochs;
N: Number of samples;
Output
$performance$: Performance of SMFLS ;
begin
  while $nn_{\max} > nn$ **do**
    for $q = 1$ ***to*** N **do**
      for $l = 1$ ***to*** M **do**
        for $k = 1$ ***to*** p **do**
          To evaluate equation (2.13)
        end
      end
      if $y(\mathbf{x}^{(q)}) > 0$ **then**
        $y^{(q)} = 1$
      else
        $y^{(q)} = -1$
      end
      To evaluate equation (2.20)
      if $type$=$\mu_e(q)$ **then**
        To evaluate equation (2.31)
        if $\left|e^{(q)}\right| > \bar{\gamma}$ **then**
          $\mu_e(q) = 1 - \frac{\bar{\gamma}}{\left|e^{(q)}\right|}$
        else
          $\mu_e(q) = 0$
        end
        $\mathbf{w}^{(q+1)} = \mathbf{w}^{(q)} - \mu_e(q)\nabla J(\mathbf{w}^{(q)})$
      end
      To update $m_{F_k^l}(q)$, $\sigma_{F_k^l}(q)$, $\theta_l(q)$ and $\sigma_X(q)$ from $\mathbf{w}^{(q+1)}$
    end
    $performance\,(\%) = (type, nn) = 100 \times (find\,(y_d - y) == 0)/N$
  end
end

Figure 8: Training algorithm for type-1 and non-singleton SMFLS.

2.3 EXPERIMENTAL RESULTS

Performance analyses discussed in this section made use of measured data set provided by MRS Logística S.A.. This data set is representative to illustrate the complexity of the problem of classification of events and provide an excellent interpretation of the trouble by the maintenance team, thus not requiring the continuous update of the proposed model.

The data set consists of current signals [A] against time sampled through four channels of an industrial data acquisition board. The sampling rate was of 100Hz in a period of 2 seconds. Once the FLS has been trained and the switch machines are functional, the acquired current signals [A] are sent to the server where the proposed

model is operating in real-time in order to classify the faults discussed in this paper. The acquired signals fall within the following four classes: lack of lubrication, lack of adjustment, malfunction of a component and normal operation. The adopted equipment is the Alstom's P80 Switch Machine. The measured data set was originated through a measurement campaign in various switch machines considering several aspects (availability of the equipment, the complexity of the operation, favorable weather conditions, etc.). Initially, there were 27 measures of the current signal for the lack of lubrication, 16 for the lack of adjustment, 74 for the malfunction of a component and 1376 for the normal operation, respectively. We added copies of instances from the under-represented classes in order to create a balanced data set. Thus, there are 1,376 measures of the current signal for each of the four classes. The specialists from MRS Logística S.A., through their prior experiences and numerical results showed that such amount of data is enough to design a classification technique for the given problem. We show samples of such signals in Figure 9.

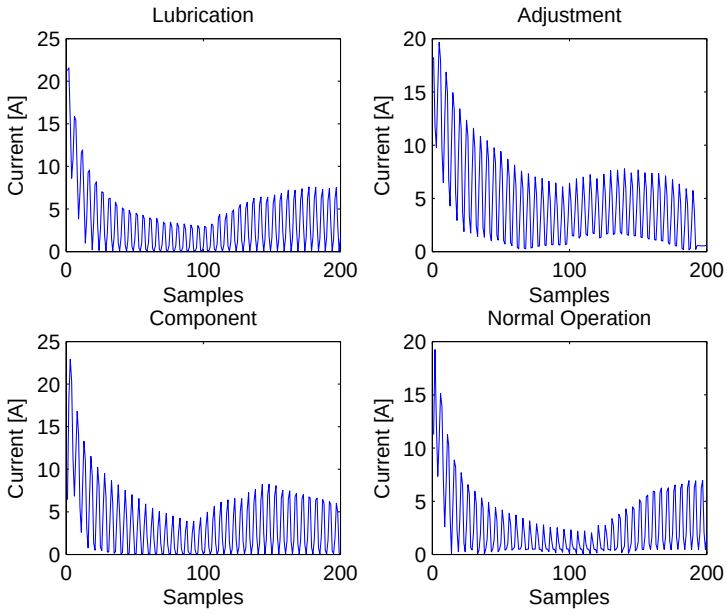


Figure 9: Typical signal waveform for the four classes.

Following the block diagram of the classification technique depicted in Figure 3, we adopted techniques for Feature Extraction and Feature Selection described in [30] and [4], since they provided promising results in this application. Basically, the feature extraction is based on higher-order statistics [40], [41] while feature selection is based on Fisher's discriminant ratio (FDR) [40]. Also, for comparison purpose, we implemented the type-1 and singleton/non-singleton FLS trained by local Lipschitz estimation (LLE) [18], the delta rule delta (DRD)[10, 15–18] and the combination of DRD with the variable step size $\mu_e(q)$ described in equation (4.16). We named these techniques as LLE FLS, DRD FLS

and DRD SMFLS, respectively, and compared its numerical results with those obtained from the proposals.

The adopted step size for type-1 and singleton/non-singleton FLS trained with the steepest descent method is $\alpha = 0.0001$. Considering the SMFLS, we heuristically had chosen $\bar{\gamma} = 0.4472$. For the implementation of the DRD FLS and the DRD SMFLS, we considered $\alpha_{DRD} = 0.00001$, $K_1 = 0.0001$, $K_2 = 0.99$ and $\rho_{\mathbf{w}} = 20$. For the VS, the MVS, the VSA and the MSVA SMFLS we used $\alpha_{VS} = \alpha_{VSA} = 0.0002$.

Each type-1 and singleton/non-singleton FLS is composed of four rules ($M = 4$), two rules for the class that has the presence of the event and another two for the class that does not have the presence of the event. After evaluating the performance with $K_p = 1, 2, 3, 4$ and 5 features selected with FDR, we chose $K_p = 3$ and use them as an input of the proposed classifiers. The performance gains in term of accuracy and convergence speed are not relevant when $K_p > 3$. We have obtained the parameters of membership functions heuristically from the calculation of means and variances of the features vector constituted by K_p elements. Observing algorithm in Figure 8, note that $y^{(q)} = 1$ means that the vector $\mathbf{r}$, constituted by a sample signal of the motor current of the switch machine, is associated with the occurrence of the event. On the other hand, $y^{(q)} = -1$ states that the vector $\mathbf{r}$ is associated with the absence of the event.

Furthermore, we equally and randomly distributed the data set in training and test sets and considered 100 epochs for the training phase.

2.3.1 Convergence Speed Analysis

Figures 10 and 11 show the convergence speed for the lubrication class and Figures 12 and 13 illustrate the convergence speed for the component class, considering the type-1 and singleton/non-singleton FLS, respectively. Although they are not presented here, the proposed models offer similar convergence speed for the other classes of faults.

For both type-1 and singleton/non-singleton FLS, the proposal based on the MVS SMFLS achieve higher convergence speed than the others. It is important to mention that techniques based on the MSE SMFLS and the SMFLS returned interesting convergence speed and we can consider them as reasonable options. All proposals presented higher convergence speed than the training phase based on the steepest descent method.

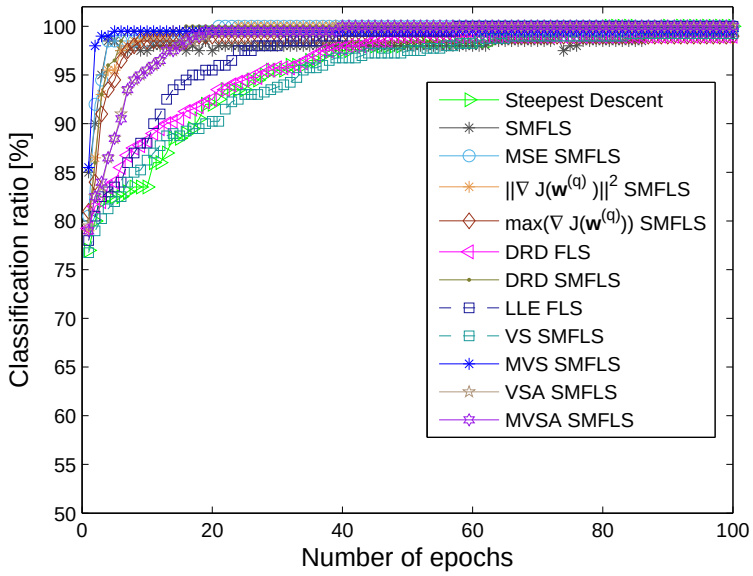


Figure 10: The convergence speed of the type-1 and singleton FLS for lubrication.

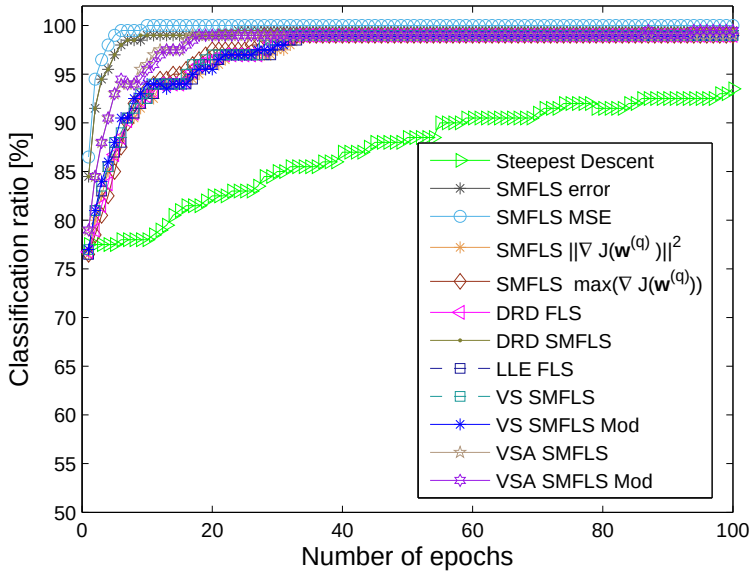


Figure 11: The convergence speed of the type-1 and non-singleton FLS for lubrication.

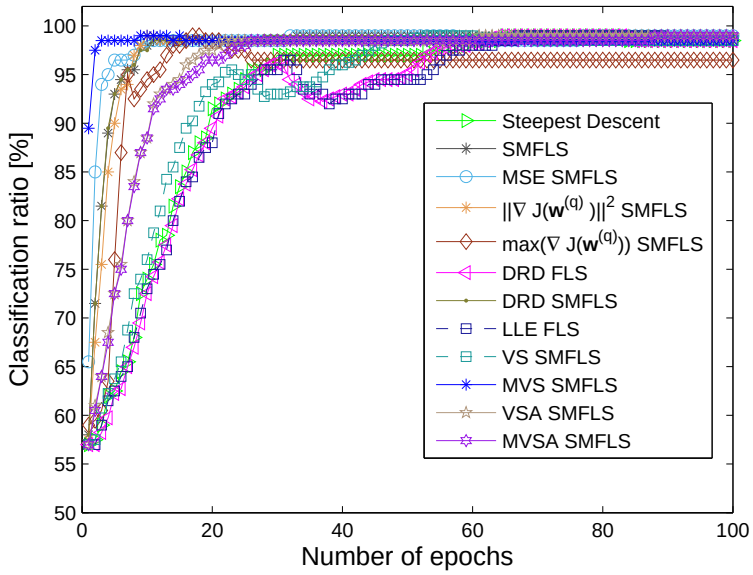


Figure 12: The convergence speed of the type-1 and singleton FLS for component.

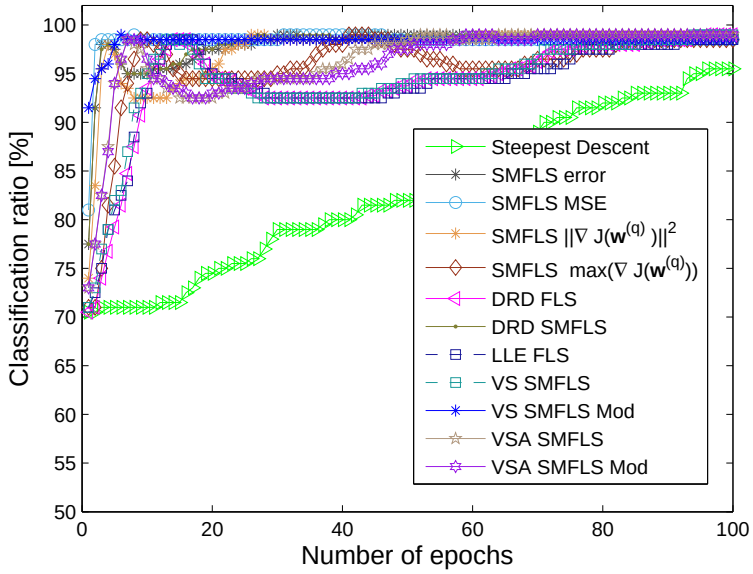


Figure 13: The convergence speed of the type-1 and non-singleton FLS for component.

2.3.2 Computational Complexity Reduction

This Section discusses the computational complexity reduction achieved during the training phase by the proposed models. To make this analysis, we adopted the percentage of occurrences when $\mu(q)$ is not updated, which is expressed by

$$\tau[\%] = 100 \times \frac{N_{nup}}{N}, \tag{2.39}$$

where N_{nup} is the number of data set that result in $\mu(q) = 0$, according to the SM concept, and N is the cardinality of the measured data set. This parameter is important because $\mu(q) = 0$ means that $\mathbf{w}^{(q+1)} = \mathbf{w}^{(q)}$ and, as a consequence, no computational effort to update $\mathbf{w}^{(q+1)}$ in the q^{th} iteration is needed.

Figures 14 and 15 show the computational complexity reduction on the hundredth epoch for lack of adjustment class and Figures 16 and 17 follow the same information for normal operation class, considering the type-1 and singleton/non-singleton FLS classifiers, respectively. Similar computational complexity reductions are obtained for the other classes of faults discussed in this work. In general, for both type-1 and singleton/non-singleton FLS, the proposal based on the MSE SMFLS had higher computational complexity reduction than the other proposals. It is worth mentioning that the training based on the MVS SMFLS and the DRD SMFLS also achieved interesting values of $\tau[\%]$.

Table 1 expresses the obtained values of $\tau[\%]$ for type-1 and singleton FLS and type-1 and non-singleton FLS, respectively. As a consequence, Table 2 shows the best computational complexity reduction achieved in [%] for the type-1 and singleton/non-singleton FLS. The efficiency $\eta[\%]$ is given by

$$\eta\,[\%] = \frac{1}{N_c} \sum_{j=1}^{N_c} \tau_{\max}\,[\%]\,, \tag{2.40}$$

where $\tau_{\max}$ is the highest value of $\tau[\%]$ for the jth class and N_c is the total number of classes.

In general, we note that SMFLS based techniques have been successful, obtained interesting results with high percentages of $\tau[\%]$. Note that type-1 and singleton FLS achieved a computational complexity reduction of 96.3% and returns better results when it is compared with type-1 and non-singleton FLS, which achieved 93.6%. The difference between the achieved values is due to the fact that type-1 and non-singleton FLS does not properly handle with uncertainties associated with the input vector.

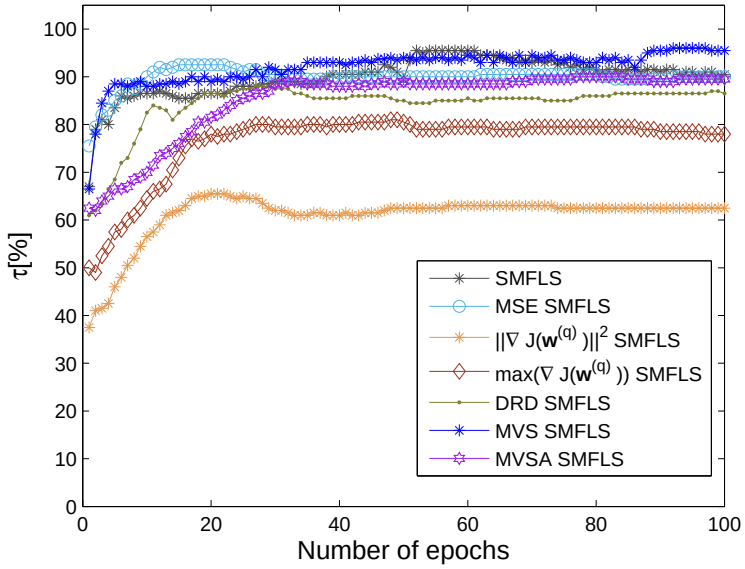


Figure 14: The computational complexity reduction for the lack of adjustment, when we use the type-1 and singleton FLS.

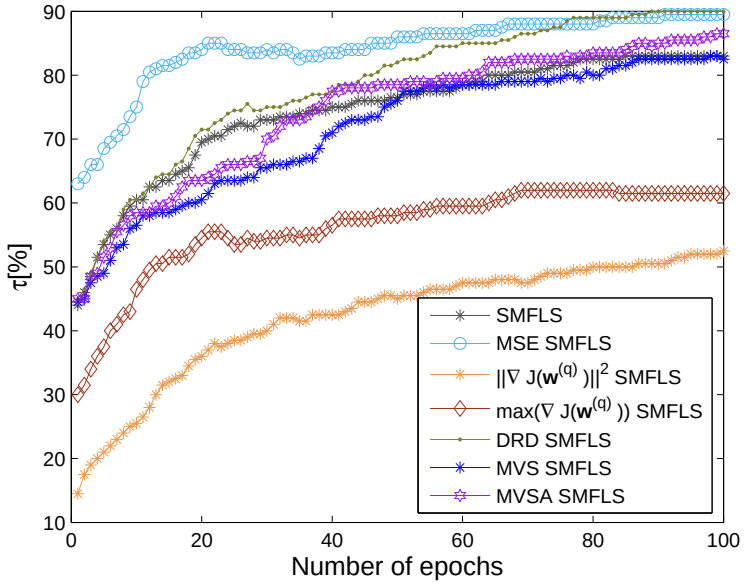


Figure 15: The computational complexity reduction for the lack of adjustment, when we use the type-1 and non-singleton FLS.

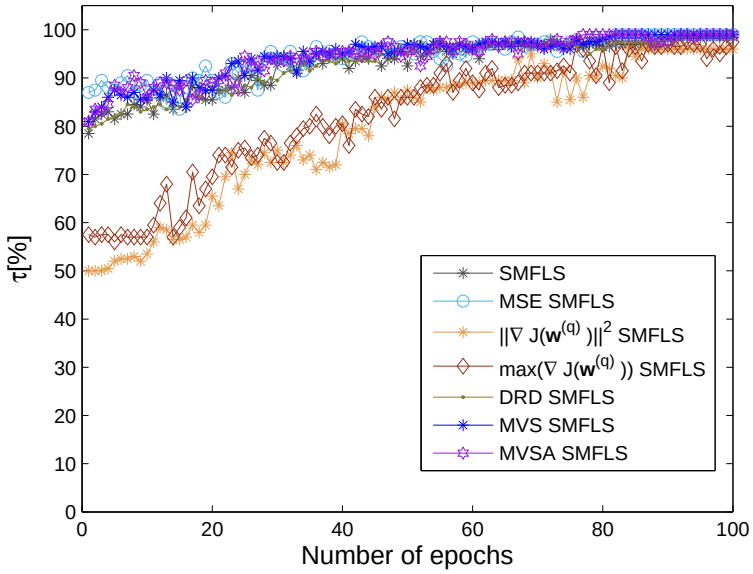


Figure 16: The computational complexity reduction for the normal operation, when we use the type-1 and singleton FLS.

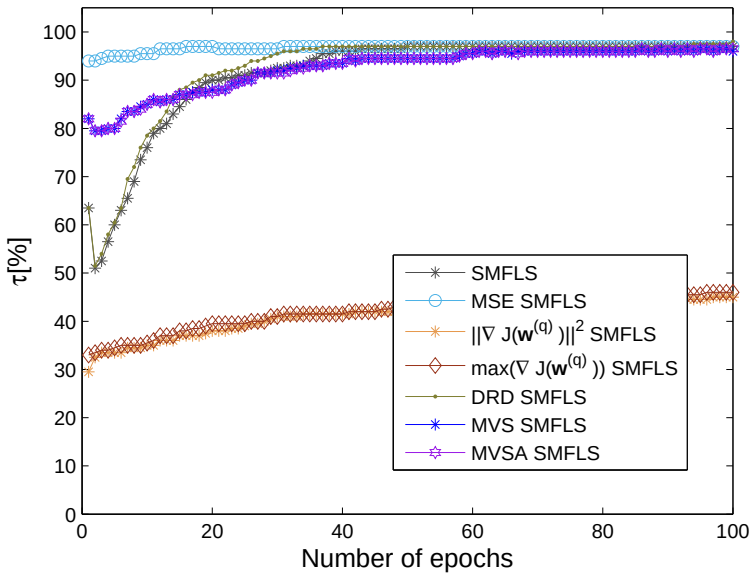


Figure 17: The computational complexity reduction for the normal operation, when we use the type-1 and non-singleton FLS.

Table 1: Computational complexity reduction of the type-1 and singleton/non-singleton FLS.

Events	Singleton FLS	Non-Singleton FLS
Lack of lubrication	τ[%]	τ[%]
SMFLS	88.5	73.5
MSE SMFLS	97.0	90.5
$\left\|\nabla J(\mathbf{w}^{(q)})\right\|^2$ SMFLS	67.0	24.5
$\max(\nabla J(\mathbf{w}^{(q)})$ SMFLS	90.5	35.0
DRD SMFLS	95.5	90.5
MVS SMFLS	94.5	75.0
MVSA SMFLS	95.5	76.5
Lack of adjustment	τ[%]	τ[%]
SMFLS	90.5	83.0
MSE SMFLS	90.0	89.5
$\left\|\nabla J(\mathbf{w}^{(q)})\right\|^2$ SMFLS	62.5	52.5
$\max(\nabla J(\mathbf{w}^{(q)})$ SMFLS	78.0	61.5
DRD SMFLS	86.5	90.0
MVS SMFLS	95.5	82.5
MVSA SMFLS	89.5	86.5
Component malfunction	τ[%]	τ[%]
SMFLS	83.0	79.0
MSE SMFLS	90.5	87.5
$\left\|\nabla J(\mathbf{w}^{(q)})\right\|^2$ SMFLS	78.2	80.0
$\max(\nabla J(\mathbf{w}^{(q)})$ SMFLS	88.0	78.0
DRD SMFLS	86.0	84.5
MVS SMFLS	93.5	96.0
MVSA SMFLS	91.5	88.0
Normal Operation	$\tau(q)$[%]	$\tau(q)$[%]
SMFLS	98.5	97.0
MSE SMFLS	99.0	97.0
$\left\|\nabla J(\mathbf{w}^{(q)})\right\|^2$ SMFLS	96.0	45.0
$\max(\nabla J(\mathbf{w}^{(q)})$ SMFLS	97.5	46.0
DRD SMFLS	98.0	98.0
MVS SMFLS	99.0	96.0
MVSA SMFLS	99.0	96.5

Table 2: Best computational complexity reduction achieved in [%] for the type-1 and singleton/non-singleton FLS.

Events	Singleton FLS	
	τ[%]	Choosen Technique
Lack of lubrication	97.0	MSE SMFLS
Lack of adjustment	95.5	MVS SMFLS
Component malfunction	93.5	MVS SMFLS
Normal Operation	99.0	MSE SMFLS, MVS SMFLS and MVSA SMFLS
Efficiency (η)	**96.3**	
Events	Non-Singleton FLS	
	τ[%]	Choosen Technique
Lack of lubrication	90.5	MSE SMFLS and DRD SMFLS
Lack of adjustment	90.0	DRD SMFLS
Component malfunction	96.0	MVS SMFLS
Normal Operation	98.0	DRD SMFLS
Efficiency (η)	**93.6**	

2.3.3 Classification Rate Analysis

Numerical results presented in Table 3 indicates the correct classification rates for the type-1 and singleton/non-singleton FLS. Based on this Table, we see that the proposed models can, in addition to the higher convergence speed, provide higher accuracy when compared with the type-1 and singleton/non-singleton FLS trained by steepest descent method. Tables 4 and 5 show, for type-1 and singleton FLS and type-1 non-singleton FLS, respectively, which proposal provides the best performance in terms of percentage of classification rate.

The final results for all techniques are shown in Table 6, in which we compare the proposals with those obtained in [30] and [4]. In this case the efficiency $\rho\,[\%]$ is expressed by

$$\rho\,[\%] = \frac{1}{N_c}\sum_{j=1}^{N_c} CR_{j,\max}\,[\%]\,, \tag{2.41}$$

where $CR_{j,\max}$ is the highest classification ratio on the hundredth epoch of training, considering the jth class. According to the results, the type-1 and singleton FLS had a similar efficiency when compared with type-1 and non-singleton FLS, with 99.6% of accuracy for both. These results show that the proposals yield improved performance if compared with previous techniques discussed in [30] and [4].

Still analyzing Table 6, we can note that the SMFLS is promising because it outperformed the techniques based on Bayes theory [30], type-1 and singleton FLS trained by steepest descent method [30], type-1 and singleton FLS trained by conjugate gradient method (CGT1-FLS) [4]. Also, it achieves a similar performance in [%] in comparison

with multilayer perceptron neural network[30]. However, even providing performances greater than or equal than other techniques discussed in the literature, the proposals based on the SMFLS offer higher convergence speed and yield computational complexity reduction during the training phase in the order of 96.3% and 93.6%, as shown in Table 2, when type-1 and singleton/non-singleton FLS are applied, respectively.

2.4 SUMMARY

In this Chapter, we have introduced the use of the set-membership concept from adaptive filter theory to reduce the computational complexity associated with the training phase of the type-1 singleton/non-singleton FLS and analyzed their effectiveness to classify typical faults in switch machine. Additionally, we outlined the set-membership combined with the delta rule delta, the variable step size and the variable step size adaptive techniques to yield additional computational complexity savings.

Considering a reduced number of epochs, the numerical results obtained through the use of measured data set showed that the type-1 and singleton/non-singleton FLS combined with the proposals result in fast convergence rates. Additionally, these proposed models can make the type-1 and singleton/non-singleton achieve classification rates higher than those obtained with the other classifier models (Bayes, multilayer perceptron neural network, type-1 singleton/non-singleton FLS tuned by the steepest descent method and by the conjugate gradient method) if a limited number of epochs applies. Moreover, the computational complexity reduction offered by the proposed models is very relevant, since a significant savings occur when the number of epochs increases. In all analyzed situations, which are representative for the addressed classification problem, the proposed models turn out to be an attractive option due to their high efficiency in reducing computational complexity, reaching percentages on the order of 96.3% and 93.6% for the type- 1 singleton/non-singleton FLS, respectively.

Table 3: Classification Rate.

Events	Singleton FLS		Non-Singleton FLS	
Lack of lubrication	Training	Test	Training	Test
Steepest Descent	100.0	100.0	99.0	99.0
SMFLS	99.5	99.5	99.5	99.5
MSE SMFLS	100.0	100.0	100.0	100.0
$\left\|\nabla J(\mathbf{w}^{(q)})\right\|^2$ SMFLS	100.0	100.0	99.0	99.0
$\max(\nabla J(\mathbf{w}^{(q)})$ SMFLS	99.5	99.5	99.0	99.0
DRD FLS	99.5	99.5	99.0	99.0
DRD SMFLS	100.0	100.0	99.5	99.5
LLE FLS	100.0	100.0	99.0	99.0
VS SMFLS	99.3	99.3	99.0	99.0
MVS SMFLS	100.0	100.0	99.0	99.0
VSA SMFLS	100.0	100.0	99.0	99.0
MVSA SMFLS	100.0	100.0	99.5	99.5
Lack of adjustment	Training	Test	Training	Test
Steepest Descent	97.5	97.5	97.0	97.0
SMFLS	99.5	99.5	99.5	99.0
MSE SMFLS	98.5	98.5	99.5	99.0
$\left\|\nabla J(\mathbf{w}^{(q)})\right\|^2$ SMFLS	98.5	98.5	99.0	98.5
$\max(\nabla J(\mathbf{w}^{(q)})$ SMFLS	98.5	98.5	99.0	96.5
DRD FLS	98.0	98.0	94.8	98.5
DRD SMFLS	98.5	98.5	99.5	98.5
LLE FLS	97.0	97.0	96.5	98.5
VS SMFLS	94.0	94.0	97.0	98.5
MVS SMFLS	99.5	99.5	98.5	98.5
VSA SMFLS	98.5	98.5	99.0	99.0
MVSA SMFLS	98.5	98.5	99.0	99.0
Component malfunction	Training	Test	Training	Test
Steepest Descent	98.5	98.5	99.0	99.0
SMFLS	99.5	99.5	99.5	99.5
MSE SMFLS	99.5	99.5	98.5	98.5
$\left\|\nabla J(\mathbf{w}^{(q)})\right\|^2$ SMFLS	98.5	98.5	99.5	99.5
$\max(\nabla J(\mathbf{w}^{(q)})$ SMFLS	96.5	96.5	98.5	98.5
DRD FLS	98.5	98.5	99.5	99.5
DRD SMFLS	98.5	98.5	98.5	98.5
LLE FLS	98.5	98.5	98.5	98.5
VS SMFLS	98.5	98.5	99.5	99.5
MVS SMFLS	98.5	98.5	98.5	98.5
VSA SMFLS	99.5	99.5	99.5	99.5
MVSA SMFLS	99.5	99.5	99.5	99.5
Normal Operation	Training	Test	Training	Test
Steepest Descent	99.5	99.5	100.0	100.0
SMFLS	99.5	99.5	100.0	100.0
MSE SMFLS	99.5	99.5	100.0	100.0
$\left\|\nabla J(\mathbf{w}^{(q)})\right\|^2$ SMFLS	99.5	99.5	100.0	100.0
$\max(\nabla J(\mathbf{w}^{(q)})$ SMFLS	98.5	98.5	100.0	100.0
DRD FLS	97.0	97.0	98.0	98.0
DRD SMFLS	99.5	99.5	100.0	100.0
LLE FLS	99.5	99.5	100.0	100.0
VS SMFLS	99.5	99.5	100.0	100.0
MVS SMFLS	99.5	99.5	100.0	100.0
VSA SMFLS	98.5	98.5	100.0	100.0
MVSA SMFLS	99.5	99.5	100.0	100.0

Table 4: The classification rate for the type-1 and singleton FLS.

Events	Training	Test	Choosen Technique
Lack of lubrication	100.0	100.0	Steepest Descent, MSE SMFLS, $\left\|\nabla J(\mathbf{w}^{(q)})\right\|^2$ SMFLS , DRD SMFLS, LLE FLS, MVS SMFLS, VSA SMFLS and MVSA SMFLS
Lack of adjustment	99.5	99.5	SMFLS and MVS SMFLS
Component malfunction	99.5	99.5	SMFLS, MSE SMFLS, VSA SMFLS and MVSA SMFLS
Normal Operation	99.5	99.5	Steepest Descent, SMFLS, MSE SMFLS, $\left\|\nabla J(\mathbf{w}^{(q)})\right\|^2$ SMFLS , $\max(\nabla J(\mathbf{w}^{(q)})$ SMFLS , DRD SMFLS, LLE FLS, VS SMFLS, MVS SMFLS, VSA SMFLS and MVSA SMFLS

Table 5: The classification rate for the type-1 and non-singleton FLS.

Events	Training	Test	Choosen Technique
Lack of lubrication	100.0	100.0	MSE SMFLS
Lack of adjustment	99.5	99.0	SMFLS, MSE SMFLS, VSA SMFLS and MVSA SMFLS
Component malfunction	99.5	99.5	SMFLS, $\left\|\nabla J(\mathbf{w}^{(q)})\right\|^2$ SMFLS , DRD SMFLS, VS SMFLS, VSA SMFLS and MVSA SMFLS
Normal Operation	100.0	100.0	Steepest Descent, SMFLS, MSE SMFLS, $\left\|\nabla J(\mathbf{w}^{(q)})\right\|^2$ SMFLS, DRD SMFLS, LLE FLS, VS SMFLS, MVS SMFLS and MVSA SMFLS

Table 6: The comparative classification rate.

Events	Bayes [30]	MLP [30]	Steepest Descent [30]	CGT1-FLS [4]	Singleton FLS	Non-Singleton FLS
Lack of lubrication	94.5	100.0	100.0	100.0	100.0	100.0
Lack of adjustment	98.0	100.0	97.5	97.5	99.5	99.0
Component malfunction	99.0	98.3	98.5	99.5	99.5	99.5
Normal Operation	98.5	100.0	99.5	100.0	99.5	100.0
Efficiency (ρ)	**97,5**	**99,6**	**98,9**	**99,3**	**99,6**	**99,6**

3 AN ENHANCED SINGLETON TYPE-2 FUZZY LOGIC SYSTEM FOR FAULT CLASSIFICATION IN A SWITCH MACHINE

A railroad switch is an electromechanical equipment that guides railway trains from one track to another, such as at a railway junction, while a switch machine is an equipment used for handling railroad switches. Among all possible faults that can occur in a switch machine, the three main ones are: lack of lubrication, lack of adjustment and malfunction of a component. The expansion of Brazilian railroad sector results in more use of switch machines. Therefore, the reliability of switch machine is a challenging issue to deal with, because real time operation, monitoring, and diagnosis of switch machines are of vital importance, especially with respect to predictive maintenance to avoid accidents and losses [4, 30].

Researches on switch machines have been in evidence in recent years since they are important types of equipment that must be constantly monitored to avoid accidents. For example, [34] describes a strategy and a technical architecture for prognosis and health management of a switch machine. Feature extraction techniques and principal component analysis have been used in the assessment of the machines' health. [35] makes use of a support vector machine (SVM) in an online switch machine condition monitoring system based on current measurements trends to detect faults at their earliest stage or even before they occur. In [3], wavelet transform together with SVM show that electrical active power can be a useful parameter for condition monitoring of switch machines. Moreover, [36] applies expert system for fault identification in a switch machine, when the failure modes are not easily separable. In addition, [37] addresses a technique for detecting of gradual failure in a switch machine, in which a Kalman filter was applied for pre-processing signal signature. Other works such as [38] and [39] present alternative techniques related to unsupervised and semi-supervised fault detection and identification in industrial plants based on the analysis of the control and error signals. The work in [44] developed a model based on SVM with Gaussian kernel to identify "drive-rod out-of-adjustment" failure mode in switch machines. A common aspect of all these works is the focus on the detection of the existence of faults.

Regarding monitoring and diagnosis, [30] discussed and analyzed the performance of techniques for classifying typical faults (lack of lubrication, lack of adjustment and malfunction of a component) that can occur in a switch machine. Combinations of feature extraction technique based on higher-order statistics [40], [41], feature selection technique based on Fisher's discriminant ratio [40], and three classifiers (Bayes theory [42], multilayer perceptron neural network [43], and type-1 and singleton fuzzy logic system (FLS) [28, 30]) showed performance improvement when applied to a data set, composed of measured current signals. Later, [4] discussed a type-1 and singleton FLS trained by the conjugate gradient method (i.e., 2nd-order information) and the reported results show

that such FLS can offer higher convergence speed and classification ratio for a limited number of epochs than that one trained by the steepest descent method.

Due to the computational complexity of FLSs, we point out that the training phase with high accuracy and convergence speed is of utmost importance because it can result in less hardware resource utilization and, as a consequence, can allow almost real-time training of the classification technique when new patterns need to be covered. Although type-1 FLSs offer improved performance, certain kind of uncertainties associated with the classification problem is something that, unfortunately, type-1 FLS may not properly handle.

In order to deal with the limitation of type-1 FLS and to reduce computational complexity during the training phase, this work proposes a modified version of interval singleton type-2 fuzzy logic system (IST2-FLS) for classifying the aforementioned types of faults in a railroad switch machine when a reduced set of features is extracted from the current signal (fault signature).

The main contributions of this Chapter are summarized as follows:

- We propose two FLS-based classifiers. The first one is based on a new approach, called upper and lower type-2 fuzzy logic system (ULST2-FLS), reducing the complexity of the training phase. The second one is based on IST2-FLS, which was introduced in [28]
- We adopt the steepest descent as the training method, the height type-reduction, max-product composition, product implication and adopt the gaussian primary membership function with uncertain mean [28, 29]. Also, we present the deductions for updating the parameters of IST2-FLS and ULST2-FLS for the first time in the literature.
- We present performance analyzes in terms of classification ratio and convergence speed by using a real data set constituted by measured signal from several switch machines. After that, we discuss comparative performances analysis among the proposals and previous techniques (Bayes theory, multilayer perceptron neural network, type-1 and singleton FLS trained by steepest descent method and type-1 and singleton FLS trained by the conjugate gradient method).

Our major conclusions are as follows:

- Numerical results show that the IST2-FLS and ULST2-FLS converge faster than those techniques previously discussed in [30] and [4].
- The ULST2-FLS is easier to implement and yield better results when compared with the classical IST2-FLS.

- Comparative analyses show that classification ratios achieved by the IST2-FLS and the ULST2-FLS are higher than those obtained with the previous techniques.

The rest of the Chapter is organized as follows: Section 3.1 deals with the problem formulation. Section 3.2 is devoted to the IST2-FLS background. Section 3.3 aims to discuss the proposal based on ULST2-FLS. Section 3.4 discusses the results of computer simulations. Section 3.5 states the main conclusions regarding this Chapter. Finally, Appendixes A and B present the deduction of equations responsible for update parameters of IST2-FLS and ULST2-FLS, respectively.

3.1 PROBLEM FORMULATION

Among all possible faults that can occur in a switch machine, the three main ones are lack of lubrication, lack of adjustment and malfunction of a component. Let $\mathbf{r} \in \mathbb{R}^{N\times 1}$ a vector constituted by a sample signal of current of the motor of switch machine. As mentioned in Chapter 2, Figure 3 shows the paradigm used for the classification of events. In the block "Feature Extraction", $\mathbf{p}_l$, $\mathbf{p}_a$, $\mathbf{p}_c$ and $\mathbf{p}_n$ refer respectively to lack of lubrication, lack of adjustment, malfunction of a component and normal operation feature vectors. The block "Feature Selection" provides selected features $\mathbf{K}_{p_l}$, $\mathbf{K}_{p_a}$, $\mathbf{K}_{p_c}$ and $\mathbf{K}_{p_n}$ vectors (with remarkable dimensionality reduction) from $\mathbf{p}_l$, $\mathbf{p}_a$, $\mathbf{p}_c$ and $\mathbf{p}_n$, respectively. The block "Classification" applies one of the classification techniques discussed in this work to obtain the output vector $\mathbf{s}$, thereby identifying the type of fault. We use four independent classifiers, one for each possible event.

Under a limited number of epochs, an effective training is a very difficult task to be accomplished. The depicted situation worsens when a specific application demands a periodic update of the technique, i.e, the classification technique needs to be retrained to cover new patterns. As a result, the increase of the convergence speed during the training phase of classification technique without decreasing its accuracy is the problem that this work pays attention when a type-2 based classification technique is applied.

3.2 THE INTERVAL SINGLETON TYPE-2 FUZZY LOGIC SYSTEM

Mendel in [29] makes the use of type-2 FLS much more accessible to FLS designers, providing mathematical formulas and flowcharts for computing the necessary derivatives to implement steepest descent methods for training. It explains why computing such derivatives is much more challenging than it is for a type-1 FLS. Classifiers based on type-2 FLS are being increasingly used different applications due to their ability to model the database uncertainties such as those present in the measurements that activate the FLS or in the data that are used to tune the parameters of a FLS [28]. The IST2-FLS have demonstrated better abilities to handle uncertainties than their type-1 FLS counterparts in

many applications. The overview and comparison discussed in [48, 49] helped researchers and practitioners on IST2-FLS to choose the most suitable structure and type-reduction algorithms, from the computational cost perspective for implementation. In [50], the IST2-FLS achieved better results when compared to type-1 FLS and an adaptive neural FLS for vehicle classification. The proposal discussed in [51] addresses a new technique which aims at classifying different states from an incoming non-stationary signal. To overcome the failure of standard classifiers at generalizing the patterns for such signals, the authors have proposed an interval type-2 fuzzy based adaptive neural FLS. Hosseini *et al.* [52] presented an automatic approach to learn and tune gaussian interval type-2 membership functions with application to multidimensional pattern classification problems. In [53], Alhaddad *et al.* discussed a genetic type-2 fuzzy logic-based classifier which is able to handle the inter- and intra-user uncertainties to produce better accuracy when compared with other competing classifiers such as bayesian linear discriminant analysis and regularized Fisher linear discriminant analysis. Recently, Juang *et al.* in [54] proposed an interval type-2 neural fuzzy classifier learned through soft margin minimization and applies it to human body posture classification.

The IST2-FLS has five main blocks: fuzzifier, rules, inference, type-reducer and defuzzifier. They are widely discussed in [28] and are interconnected as shown in Figure 18. Once the rules are defined, the IST2-FLS can be viewed like a mapping of the inputs to the outputs and this can be expressed by a function $y = f(x)$.

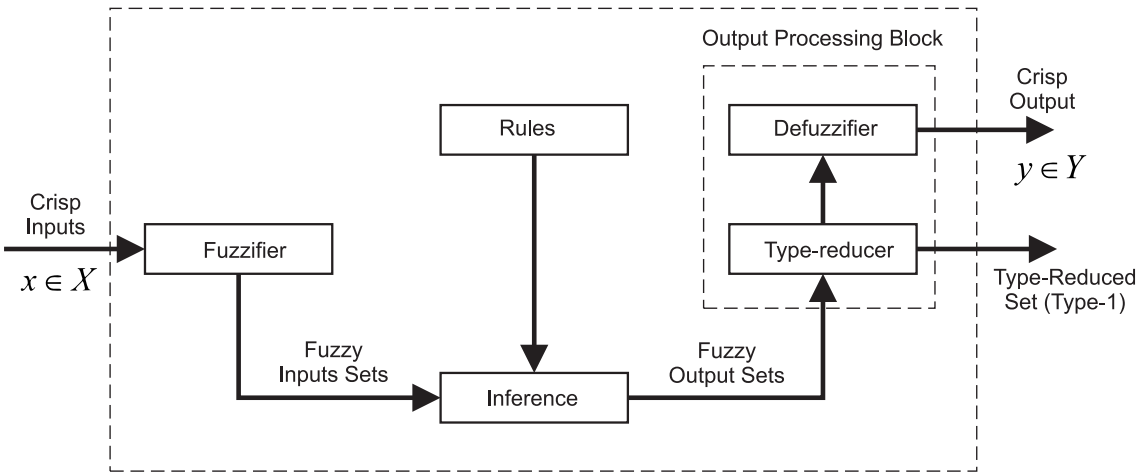


Figure 18: The type-2 FLS [28].

Considering the membership functions as gaussian functions with having a fixed standard deviation σ, and an uncertain mean situated in $m \in [m_1, m_2]$, as given by:

$$\mu_A(x) = \exp\left[-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2\right], \qquad m \in [m_1, m_2], \tag{3.1}$$

where A symbolize a type-1 fuzzy set.

The upper membership function (UMF) and lower membership function (LMF) are given, respectively, by [28, 29]:

$$\bar{\mu}_{\tilde{A}}(x) = \begin{cases} \exp\left[\frac{1}{2}\left(\frac{x-m_1}{\sigma}\right)^2\right], & x < m_1 \\ 1, & m_1 \leq x \leq m_2 \\ \exp\left[\frac{1}{2}\left(\frac{x-m_2}{\sigma}\right)^2\right], & x > m_2 \end{cases} \tag{3.2}$$

and

$$\underline{\mu}_{\tilde{A}}(x) = \begin{cases} \exp\left[\frac{1}{2}\left(\frac{x-m_2}{\sigma}\right)^2\right], & x \leq \frac{m_1+m_2}{2} \\ \exp\left[\frac{1}{2}\left(\frac{x-m_1}{\sigma}\right)^2\right], & x > \frac{m_1+m_2}{2} \end{cases}. \tag{3.3}$$

The type-reducer is a part of the output processing block which aims to reduce the type-2 output fuzzy set to type-1 output fuzzy set [55–58]. The height type-reduction has the following structure[59–62]:

$$f_{s2}(\mathbf{x}) = [Y_L(\mathbf{x}), Y_R(\mathbf{x})], \tag{3.4}$$

where $f_{s2}(\mathbf{x})$ is an interval fuzzy set determined by two final outputs related with the left and the right part of primary membership function, called $Y_L(\mathbf{x})$ and $Y_R(\mathbf{x})$ respectively. They can be expressed by

$$Y_L(\mathbf{x}) = \frac{\sum_{l=1}^{le} \theta_l \overline{\vartheta}_l(\mathbf{x})}{\sum_{l=1}^{le} \overline{\vartheta}_l(\mathbf{x})} + \frac{\sum_{l=le}^{M} \theta_l \underline{\vartheta}_l(\mathbf{x})}{\sum_{l=le}^{M} \underline{\vartheta}_l(\mathbf{x})} \tag{3.5}$$

and

$$Y_R(\mathbf{x}) = \frac{\sum_{l=1}^{ri} \theta_l \underline{\vartheta}_l(\mathbf{x})}{\sum_{l=1}^{ri} \underline{\vartheta}_l(\mathbf{x})} + \frac{\sum_{l=ri}^{M} \theta_l \overline{\vartheta}_l(\mathbf{x})}{\sum_{l=ri}^{M} \underline{\vartheta}_l(\mathbf{x})}, \tag{3.6}$$

where θ_l is the weight associated with the l-th rule, $l = 1, \ldots, M$, le is the rule that exchange the UMF to LMF, ri is the rule that exchange the LMF to UMF. $\overline{\vartheta}_l(\mathbf{x})$ and $\underline{\vartheta}_l(\mathbf{x})$ are calculated as follows [27–29]:

$$\overline{\vartheta}_l(\mathbf{x}) = \overline{\mu}_{\tilde{F}_1^l}(x_1) \star \cdots \star \overline{\mu}_{\tilde{F}_p^l}(x_p) \tag{3.7}$$

and

$$\underline{\vartheta}_l(\mathbf{x}) = \underline{\mu}_{\tilde{F}_1^l}(x_1) \star \cdots \star \underline{\mu}_{\tilde{F}_p^l}(x_p). \tag{3.8}$$

To calculate the final output $Y_L(\mathbf{x})$, see (3.5), it is necessary to determine

$$\vartheta_{lL}(\mathbf{x}) = \frac{\overline{\vartheta}_l(\mathbf{x}) + \underline{\vartheta}_l(\mathbf{x})}{2}. \tag{3.9}$$

We first consider UMF in the primary rule $l = 1$ to le. After that, we exchange UMF by LMF, considering now $l = le$ to M. To determine this specific rule $l = le$ and to calculate the final output $Y_L(\mathbf{x})$, we use the procedure described in Figure 19 [28, 59]. In this procedure, the first calculus of $Y_L(\mathbf{x})$ is given by

$$Y_L(\mathbf{x}) = \frac{\sum_{l=1}^{M} \theta_l \vartheta_{lL}(\mathbf{x})}{\sum_{l=1}^{M} \vartheta_{lL}(\mathbf{x})}. \tag{3.10}$$

The procedure for the calculus of $Y_R(\mathbf{x})$ is similar to the that adopted for $Y_L(\mathbf{x})$. However, the rule le is exchanged by ri, and (3.6) is evaluated instead of (3.5). Finally, the defuzzification of IST2-FLS is given by

$$f_{s2}(\mathbf{x}) = \frac{Y_L(\mathbf{x}) + Y_R(\mathbf{x})}{2}. \tag{3.11}$$

```
Input
M: Number of rules;
le: Number of rule that exchange the UMF to LMF;
Output
Y_L(x) output
begin
    for l = 1 to M do
        To evaluate equations (3.7) and (3.8);
        To evaluate equations (3.9);
    end
    To evaluate equation (3.10);
    Y_1 = Y_L(x);
    while η ≤ M do
        le = 1;
        while l ≤ M do
            if (θ_l ≤ Y_L(x)) and (Y_L(x) ≤ θ_{l+1}) then
                le = l;
                l = M + 1;
            else
                le = le + 1;
                l = l + 1;
            end
        end
        To evaluate equation (3.5);
        Y_2 = Y_L(x);
        if Y_2 = Y_1 then
            η = M + 1;
            Y_L(x) = Y_2;
        else
            Y_1 = Y_2;
            η = η + 1;
        end
    end
end
```

Figure 19: Procedure to determine $Y_L(\mathbf{x})$.

Given an input-output database $\left(\mathbf{x}^{(q)} : y^{(q)}\right)$, where q denotes the qth iteration. We now wish to design an IST2-FLS such that the following cost function is minimized:

$$J\left(\mathbf{w}^{(q)}\right) = \frac{1}{2}\left[f_{s2}\left(\mathbf{x}^{(q)}\right) - y^{(q)}\right]^2, \tag{3.12}$$

where $f_{s2}\left(\mathbf{x}^{(q)}\right)$ is the output of IST2-FLS. From (3.12), we know that

$$\begin{aligned}
&\mathbf{w}^{(q)} = \\
&\Big[m_{1F_1^1}(q), \ldots, m_{1F_p^1}(q), \ldots, m_{1F_1^M}(q), \ldots, m_{1F_p^M}(q), \\
&m_{2F_1^1}(q), \ldots, m_{2F_p^1}(q), \ldots, m_{2F_1^M}(q), \ldots, m_{2F_p^M}(q), \\
&\sigma_{F_1^1}(q), \ldots, \sigma_{F_p^1}(q), \ldots, \sigma_{F_1^M}(q), \ldots, \sigma_{F_p^M}(q), \\
&\theta_1(q), \ldots, \theta_M(q)]
\end{aligned} \tag{3.13}$$

and

$$\begin{aligned}
&\nabla J(\mathbf{w}^{(q)}) = \\
&\Bigg[\frac{\partial J(\mathbf{w}^{(q)})}{\partial m_{1F_1^1}(q)}, \ldots, \frac{\partial J(\mathbf{w}^{(q)})}{\partial m_{1F_p^1}(q)}, \ldots, \frac{\partial J(\mathbf{w}^{(q)})}{\partial m_{1F_1^M}(q)}, \ldots, \frac{\partial J(\mathbf{w}^{(q)})}{\partial m_{1F_p^M}(q)}, \\
&\frac{\partial J(\mathbf{w}^{(q)})}{\partial m_{2F_1^1}(q)}, \ldots, \frac{\partial J(\mathbf{w}^{(q)})}{\partial m_{2F_p^1}(q)}, \ldots, \frac{\partial J(\mathbf{w}^{(q)})}{\partial m_{2F_1^M}(q)}, \ldots, \frac{\partial J(\mathbf{w}^{(q)})}{\partial m_{2F_p^M}(q)}, \\
&\frac{\partial J(\mathbf{w}^{(q)})}{\partial \sigma_{F_1^1}(q)}, \ldots, \frac{\partial J(\mathbf{w}^{(q)})}{\partial \sigma_{F_p^1}(q)}, \ldots, \frac{\partial J(\mathbf{w}^{(q)})}{\partial \sigma_{F_1^M}(q)}, \ldots, \frac{\partial J(\mathbf{w}^{(q)})}{\partial \sigma_{F_p^M}(q)}, \\
&\frac{\partial J(\mathbf{w}^{(q)})}{\partial \theta_1(q)}, \ldots, \frac{\partial J(\mathbf{w}^{(q)})}{\partial \theta_M(q)}\Bigg]^T,
\end{aligned} \tag{3.14}$$

denote, respectively, the parameters vector and the gradient vector of a IST2-FLS.

According to [26], the type-2 FLS have the potential to provide better performance than a type-1 FLS (e.g., [63–71]). However, because of the computational complexity of a general type-2 FLS, most developed applications are based on the IST2-FLS [27]. The computations associated with IST2-FLSs are very manageable, which makes this FLS quite practical [72].

This Chapter proposes a IST2-FLS trained by steepest descent method [28], using max-product composition, product implication, height type-reduction and making use of gaussian primary membership function with uncertain mean. The algorithm is presented in Figure 20. The equations used in the algorithm to update its parameters are:

$$m_{1F_k^l}(q+1) = m_{1F_k^l}(q) - \alpha \frac{\partial J\left(\mathbf{w}^{(q)}\right)}{\partial m_{1F_k^l}(q)}, \tag{3.15}$$

$$m_{2F_k^l}(q+1) = m_{2F_k^l}(q) - \alpha \frac{\partial J\left(\mathbf{w}^{(q)}\right)}{\partial m_{2F_k^l}(q)}, \tag{3.16}$$

$$\sigma_{F_k^l}(q+1) = \sigma_{F_k^l}(q) - \alpha \frac{\partial J\left(\mathbf{w}^{(q)}\right)}{\partial \sigma_{F_k^l}(q)}, \tag{3.17}$$

and

$$\theta_l(q+1) = \theta_l(q) - \alpha \frac{\partial J\left(\mathbf{w}^{(q)}\right)}{\partial \theta_l(q)}. \tag{3.18}$$

The complete equations for updating parameters is not available in the literature, which hinders the diffusion and applicability of this kind of type-2 FLSs. Therefore, the first-order derivative of $J(\mathbf{w}^{(q)})$ with regard to parameters $m_{1F_k^l}(q)$, $m_{2F_k^l}(q)$, $\sigma_{F_k^l}(q)$, $\theta_l(q)$ are shown in Appendix A in order to facilitate the comprehension and spreading the use of the IST2-FLS in situations where there are high levels of uncertainty.

Inputs
N: Number of samples;
ep_{max}: Maximum number of epochs;
α: Learning step;
Output
$performance$: Performance of IST2-FLS;
begin
  for $ep = 1$ ***to*** ep_{max} **do**
    for $q = 1$ ***to*** N **do**
      To evaluate $Y_L\left(\mathbf{x}^{(q)}\right)$ and $Y_R\left(\mathbf{x}^{(q)}\right)$ according Equations (3.5) and (3.6);
      if $Y_L\left(\boldsymbol{x}^{(q)}\right) \leq 0$ **then**
        $Y_L\left(\mathbf{x}^{(q)}\right) = -1$;
      else
        $Y_L\left(\mathbf{x}^{(q)}\right) = 1$;
      end
      if $Y_R\left(\boldsymbol{x}^{(q)}\right) \leq 0$ **then**
        $Y_R\left(\mathbf{x}^{(q)}\right) = -1$;
      else
        $Y_R\left(\mathbf{x}^{(q)}\right) = 1$;
      end
      To calculate the equation (3.11);
      if $f_{s2}\left(\boldsymbol{x}^{(q)}\right) > 0$ **then**
        $f_{s2}\left(\mathbf{x}^{(q)}\right) = 1$;
      else
        $f_{s2}\left(\mathbf{x}^{(q)}\right) = -1$;
      end
      $\mathbf{w}^{(q+1)} = \mathbf{w}^{(q)} - \alpha \nabla J(\mathbf{w}^{(q)})$;
      To update $m_{1F_k^l}(q)$, $m_{2F_k^l}(q)$, $\sigma_{F_k^l}(q)$ and $\theta_l(q)$ from $\mathbf{w}^{(q+1)}$;
    end
    $performance\,(\%) = 100 \times \left[f_{s2}\left(\mathbf{x}^{(q)}\right) - y^{(q)}\right]/N$
  end
end

Figure 20: Training algorithm for IST2-FLS.

3.3 UPPER AND LOWER SINGLETON TYPE-2 FUZZY LOGIC SYSTEM: THE NOVELTY

Considering the IST2-FLS showed in Figure 18, we can state that the main difference in comparison to the type-1 FLS is the block called type-reducer. This block is responsible for converting the output type-2 membership function in its output type-1 membership function, which posteriorly will be defuzzified [28]. Liang *et al.* in [27] stated that IST2-FLSs are computationally intensive because the type-reducer is very intensive. As a result, the training phase is extremely computationally intensive.

To overcome this problem, we propose a modified version of IST2-FLS, so called upper and lower singleton type-2 fuzzy logic system (ULST2-FLS), which consists in adopting $Y_{Up}(\mathbf{x})$ and $Y_{Low}(\mathbf{x})$ (see (3.20) and (3.21)) instead of $Y_L(\mathbf{x})$ and $Y_R(\mathbf{x})$ (see (3.5) and (3.6)). The main advantage of ULST2-FLS is the elimination of the procedure described in Figure 19, thereby reducing the complexity of the training phase. This technique makes use of max-product composition, product implication and gaussian primary membership function with uncertain mean.

The ULST2-FLS consists in reducing the type of mentioned FLS in functions that are characterized by only the upper and lower membership functions. Therefore, the type reducer uses the following structure:

$$f_{s2}(\mathbf{x}) = [Y_{Up}(\mathbf{x}), Y_{Low}(\mathbf{x})], \tag{3.19}$$

where $f_{s2}(\mathbf{x})$ is a set determined by two final outputs $Y_{Up}(\mathbf{x})$ and $Y_{Low}(\mathbf{x})$, given by

$$Y_{Up}(\mathbf{x}) = \frac{\sum_{l=1}^{M} \theta_l \bar{\vartheta}_l(\mathbf{x})}{\sum_{l=1}^{M} \bar{\vartheta}_l(\mathbf{x})} \tag{3.20}$$

and

$$Y_{Low}(\mathbf{x}) = \frac{\sum_{l=1}^{M} \theta_l \underline{\vartheta}_l(\mathbf{x})}{\sum_{l=1}^{M} \underline{\vartheta}_l(\mathbf{x})}, \tag{3.21}$$

where $\bar{\vartheta}_l(\mathbf{x})$ and $\underline{\vartheta}_l(\mathbf{x})$ is given by (3.7) and (3.8), respectively. The ULST2-FLS is defuzzified as follows:

$$f_{s2}(\mathbf{x}) = \frac{Y_{Up}(\mathbf{x}) + Y_{Low}(\mathbf{x})}{2}. \tag{3.22}$$

For implement the algorithm for ULST2-FLS, we need to replace $Y_L(\mathbf{x})$, $Y_R(\mathbf{x})$ by $Y_{Up}(\mathbf{x})$, $Y_{Low}(\mathbf{x})$, respectively and to consider the roadmap portrayed in Figure 20. Also, we need to evaluate the Equations (3.15)-(3.18) to update the parameters from ULST2-FLS by using steepest descent method.

Considering the ULST2-FLS, the first-order derivative of $J(\mathbf{w}^{(q)})$ with regard to parameters $m_{1F_k^l}(q)$, $m_{2F_k^l}(q)$, $\sigma_{F_k^l}(q)$, $\theta_l(q)$ are shown in Appendix B.

Section 3.4 shows that the ULST2-FLS can considerably increase the convergence speed and, as a consequence, less hardware resource utilization is demanded. Also, it offers improved performance.

3.4 EXPERIMENTAL RESULTS

Performance analyses discussed in this section made use of the same measured data set used in Chapter 2 and provided by MRS Logística S.A. (`https://www.mrs.com.br/`). The data set consists of the current signal [A] against time sampled through four channels of an industrial data acquisition board. The sampling rate was of 100 Hz in a period of 2 seconds. The adopted equipment is the Alstom's P80 Switch Machine.

The acquired signals fall within the following four classes: lack of lubrication, lack of adjustment, malfunction of a component and normal operation. As mentioned in Chapter 2, there are 1,376 measures of the current signal for each of the four classes. We show samples of such signals in Figure 9.

Following the block diagram of the classification technique depicted in Figure 3, we adopted techniques for feature extraction and feature selection described in [30] and [4]. Basically, the feature extraction is based on higher-order statistics [40], [41] while feature selection is based on Fisher's discriminant ratio (FDR) [40], such as presented in Chapter 2. We have adopted them because [4, 30] demonstrated their relevance for this application.

The adopted step size for IST2-FLS and ULST2-FLS is $\alpha = 0.0001$ and both are composed of four rules ($M = 4$), two rules for the class that has the presence of the event and another two for the class that does not have the presence of the event. After evaluating the performance with $K_p = 1, 2, 3, 4$ and 5 features which were previously selected with FDR, we chose $K_p = 3$, since the performance gains in term of accuracy and convergence speed are not relevant when $K_p > 3$. We had obtained the initialization of the parameters of membership functions heuristically from the calculation of means and variances of the features vector constituted by K_p elements.

Furthermore, we equally and randomly distributed the data set in training and test sets and considered 100 epochs for the training phase.

3.4.1 Convergence Speed Analysis

Simulations were performed with the proposed algorithms and its results is compared with others obtained from [30] and [4]. The technique presented in [30] consists of a type-1 and singleton FLS trained by steepest descent method. The technique based on type-1 and singleton FLS trained by conjugate gradient (CGT1-FLS) is presented in [4] and we chose the best results to compare with the techniques presented in this work. Figures 21, 22, 23 and 24 illustrate the convergence speed for each of four different classes. The IST2-FLS and ULST2-FLS has a higher convergence than others techniques, highlighting that the ULST2-FLS achieve a faster convergence speed than the IST2-FLS, proposed by [28].

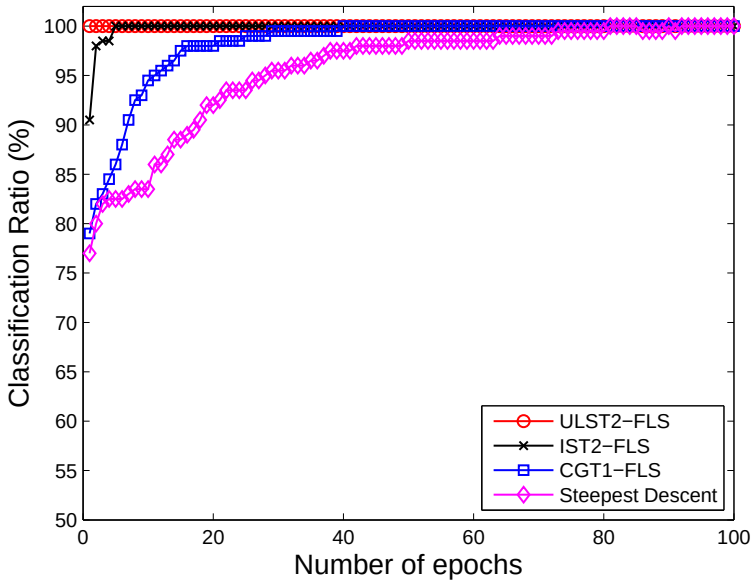


Figure 21: The convergence speed of the type-2 FLS for the lack of lubrication.

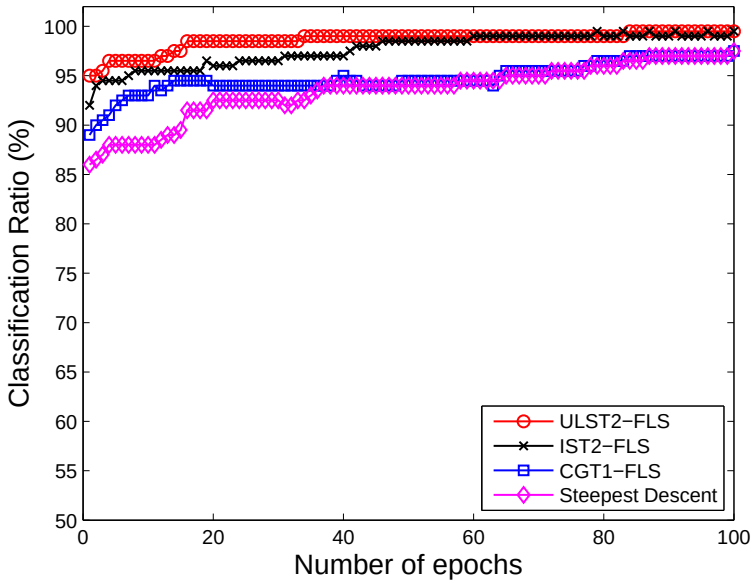


Figure 22: The convergence speed of the type-2 FLS for lack of adjustment.

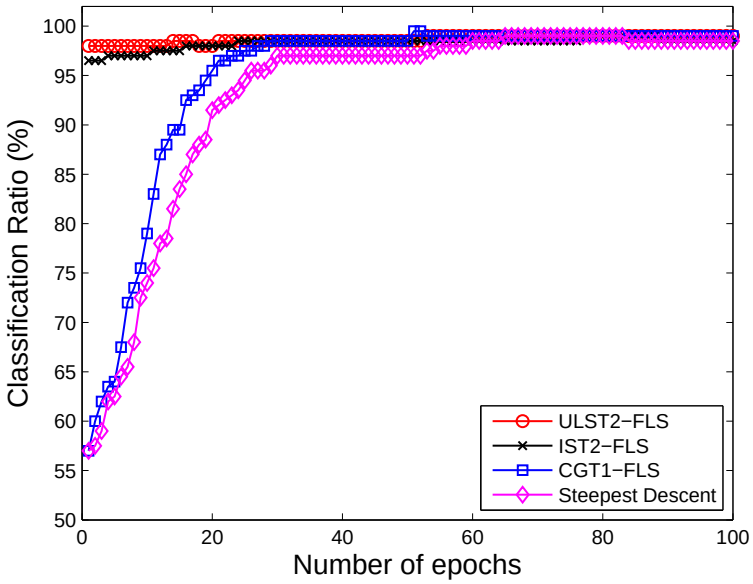


Figure 23: The convergence speed of the type-2 FLS for the component malfunction.

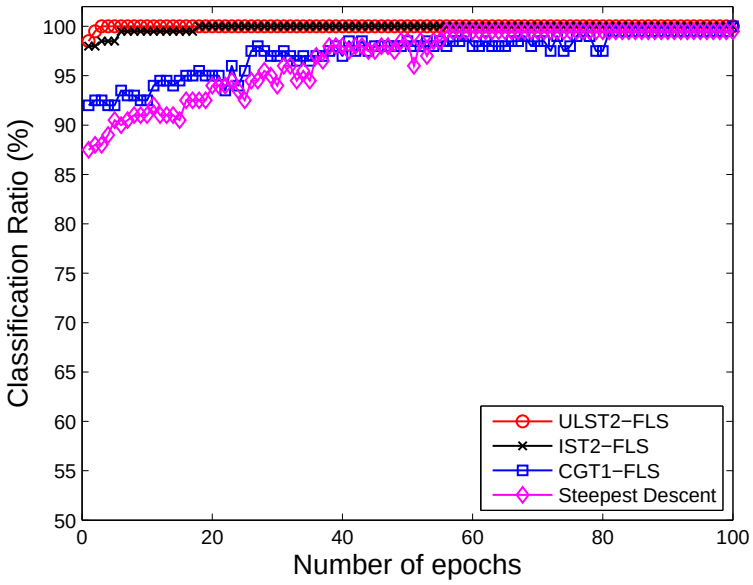


Figure 24: The convergence speed of the type-2 FLS for normal operation.

3.4.2 Classification Rate Analysis

Results presented in Table 7 shows that for the majority of classes the proposals yield a classification rate in [%] greater or equal than type-1 FLS trained by steepest descent method and CGT1-FLS during the test phase. The final results are shown in Table 8 and comparing them with those in [30] and [4]: Bayes classifier, multilayer perceptron neural network, type-1 and singleton FLS tuned by steepest descent method and type-1 and singleton FLS tuned by conjugate gradient methods. In this case the efficiency $\rho\,[\%]$ is expressed by

$$\rho\,[\%] = \frac{1}{N_c}\sum_{j=1}^{N_c} CR_{j,\max}\,[\%]\,, \tag{3.23}$$

where $CR_{j,\max}$ is the highest classification ratio on the hundredth epoch of training, considering the jth class. The IST2-FLS and the ULST2-FLS reached an efficiency of 99.6%, showing us that the proposed techniques return an improved performance if compared with previous techniques discussed in the literature and equal to multilayer perceptron neural network.

Table 7: Performance in (%) of lack of lubrication, lack of adjustment, component mulfunction and normal operation classes.

Lack of lubrication	Train	Test
IST2-FLS	100.0	100.0
ULST2-FLS	100.0	100.0
CGT1-FLS	100.0	100.0
Steepest Descent	100.0	100.0
Lack of adjustment	Train	Test
IST2-FLS	99.5	99.5
ULST2-FLS	99.5	99.5
CGT1-FLS	97.5	97.5
Steepest Descent	97.5	97.5
Component malfunction	Train	Test
IST2-FLS	99.0	99.0
ULST2-FLS	99.0	99.0
CGT1-FLS	99.0	99.0
Steepest Descent	98.5	98.5
Normal Operation	Train	Test
IST2-FLS	100.0	100.0
ULST2-FLS	100.0	100.0
CGT1-FLS	100.0	100.0
Steepest Descent	99.5	99.5

Table 8: Comparative performance in (%) of the classifiers during the test phase.

Events	Bayes [30]	MLP [30]	Steepest Descent [30]	CGT1-FLS [4]	IST2-FLS	ULST2-FLS
Lack of lubrication	94.5	100.0	100.0	100.0	100.0	100.0
Lack of adjustment	98.0	100.0	97.5	97.5	99.5	99.5
Component malfunction	99.0	98.3	98.5	99.0	99.0	99.0
Normal Operation	98.5	100.0	99.5	100.0	100.0	100.0
Efficiency (ρ)	**97.5**	**99.6**	**98.9**	**99.1**	**99.6**	**99.6**

3.5 SUMMARY

This Chapter discussed the use of signal processing techniques and computational intelligence for fault classification in switch machines.

A concept of ULST2-FLS trained by steepest descent method has been introduced in this work. Considering the height type-reduction, the deductions of equations from ULST2-FLS and IST2-FLS are presented for the first time in the literature. Results obtained through the use of real measured data revealed that both type-2 FLS discussed in this work converges faster. Classification rates are competitive when compared with those obtained with the previous approaches addressed in the literature for a reduced number of epochs (Bayes, multilayer perceptron neural network and type-1 FLS tuned by steepest descent method and conjugate gradient method). In any case, the classifiers proposed in this work turn out to be an extremely attractive option due to their high convergence speed. It was noted also that ULST2-FLS converges slightly faster than the IST2-FLS (the classical approach used in the literature).

4 SET-MEMBERSHIP TYPE-2 FUZZY LOGIC SYSTEM APPLIED TO CLASSIFICATION OF RAIL HEAD DEFECTS

Railways are a network of distributed rails which aims to connect cities, states, and countries. In most cases, they have been used for heavy loads, what impacts directly on rails, increasing its defects through material fatigue. Studies have been made to solve such problems in railways. As a consequence, there is an increasing interest in exploiting the feasibility of applying image processing and computational intelligence paradigms to address critical problems in order to improve the efficiency, safety, and environmental compatibility of transportation systems. Some problems, as detection of surface defects on rails, have been widely explored.

Researchers have tried to improve the classification of defects in rail heads using several techniques, such as geometrical approach [73], Gabor filter with texture analysis [74], image segmentation [5], and visual inspection [6]. The drawback is that all of these works concentrate on identifying the existence of a defect and not on classifying the defect type.

The type-1 fuzzy logic system (FLS) has been widely applied in classification problems due to its treatment of uncertainty, thus generating interesting results in several areas of the railway industry, as mentioned in [4,30]. Although type-1 FLSs offer improved performance, there is some kind of uncertainties associated with the classification problem that type-1 FLS may not properly handle. Then, the use of type-2 FLSs appears as a suitable option for dealing with it. However, [27] stated that type-2 FLSs are computationally intensive because their type-reducer is very computationally intensive and the same is associated with their training phase. Considering a limited number of epochs, the training with reduced computational complexity, high accuracy, and high convergence speed consists of a hard task to be accomplished with type-2 FLSs. The hardness is associated with the limited hardware resources availability and real-time constraint.

To deal with this challenging problem, this Chapter addresses how to reduce the computational complexity during the training phase of an interval and singleton type-2 fuzzy logic system (IST2-FLS) and a modified version of IST2-FLS, called upper and lower type-2 fuzzy logic system (ULST2-FLS), initially introduced in Chapter 3. Furthermore, we apply the proposal to classify the main types of rail head defects that are commonly studied due to its severity and occurrence: cracking [75], flaking [76], head-check and spalling [77]. Due to computational complexity reduction, we emphasize that the proposals demand low-cost (low-speed) processors, which is something very important to come up with green technologies for the railroad sector.

The main contributions of this Chapter are summarized as follows:

- We approach four types of defects that can occur in a rail head surface. The com-

bined study of cracking, flaking head-check and spalling linked to rail head surface has never been addressed before.

- We use geometric correction through a two-dimensional affine transformation [78–80] to facilitate the image processing and a gray level co-occurrence matrix (GLCM) to analyze textures on images [81–85].

- We apply the concept of set-membership (SM) [9–13] derived from adaptive filter theory [14], and combine it with IST2-FLS and ULST2-FLS, both trained by steepest descent method, in order to reduce their computational complexity and to increase the convergence speed during the training phase.

- We combine SM with IST2-FLS and ULST2-FLS, together with delta rule delta (DRD) [10, 15–18], local Lipschitz estimation (LLE) [10, 17–21] variable step size (VS) [22–24] and variable step size adaptive algorithms (VSA) [25] to come up with additional computational complexity savings during the training phase.

- We present performance analysis in terms of convergence speed, computational complexity reduction and classification ratio by using a data set constituted by images acquired from a rail inspection vehicle provided by MRS Logística S.A. (`https://www.mrs.com.br/`). Additionally, we show a comparative analysis of the proposed models based on previous techniques (type-1 and singleton/non-singleton FLS trained by steepest descent method, IST2-FLS and ULST2-FLS).

Our major conclusions are as follows:

- The two-dimensional affine transformation geometric correction and feature extraction through GLCM is effective for the current application.

- The SM concept combined with DRD, LLE, VS and VSA strongly contribute to reducing computational complexity during the training phase of type-2 FLS, under a limited number of epochs. They reached percentages of reduction greater than 98.2% for IST2-FLS and ULST2-FLS.

- The classification ratio yielded by the proposed models are higher than those obtained with the classical IST2-FLS and ULST2-FLS when a higher degree of uncertainties is presented in the input data.

The rest of the Chapter is organized as follows: Section 4.1 deals with the problem formulation. Section 4.2 aims to discuss the concept of SM combined with IST2-FLS and ULST2-FLS. Section 4.3 discusses the results of computer simulations. Section 4.4 states the main conclusions regarding this Chapter.

4.1 PROBLEM FORMULATION

Among all possible defects that can occur on the rail head, the four main ones are cracking, flaking, head check and spalling. Let $\mathbf{A} \in \mathbb{R}^{n \times m}$ be a matrix constituted by elements of an image of the rail head. Figure 25 shows the paradigm used for the classification of events. In the block "Geometric Correction", $\mathbf{p}$ refers to cracking, flaking, head check, spalling and normal condition of the rail head. The block "Feature Extraction" provides extracted features $\mathbf{K}_{p_c}$, $\mathbf{K}_{p_f}$, $\mathbf{K}_{p_h}$, $\mathbf{K}_{p_s}$ and $\mathbf{K}_{p_n}$ vectors from $\mathbf{p}$. The block "Classification" applies one of the classification techniques discussed in this work to obtain the output vector $\mathbf{S}$, thereby identifying the type of fault. We use five independent classifiers, one for each possible event.

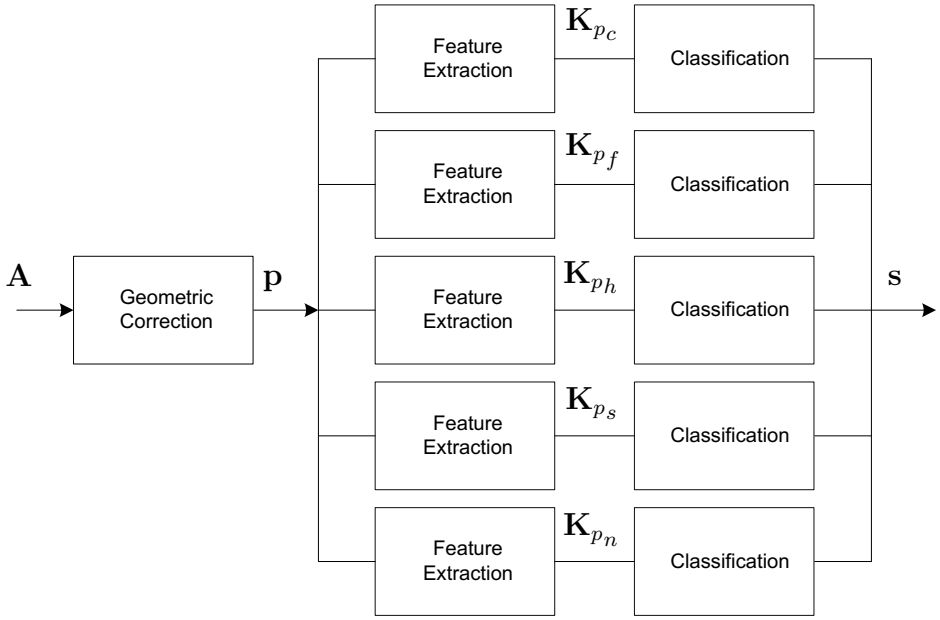


Figure 25: Block diagram of the scheme for classification of events.

The classification of events in the matrix $\mathbf{A}$ can be formulated as a simple decision between hypotheses related to the occurrence of the events covered in this work, as shown below:

$$\begin{aligned} \mathcal{H}_{A,0} &: \mathbf{A} = \mathbf{A}_{crack} \\ \mathcal{H}_{A,1} &: \mathbf{A} = \mathbf{A}_{flak} \\ \mathcal{H}_{A,2} &: \mathbf{A} = \mathbf{A}_{head} \\ \mathcal{H}_{A,3} &: \mathbf{A} = \mathbf{A}_{spall} \\ \mathcal{H}_{A,4} &: \mathbf{A} = \mathbf{A}_{norm} \end{aligned} \tag{4.1}$$

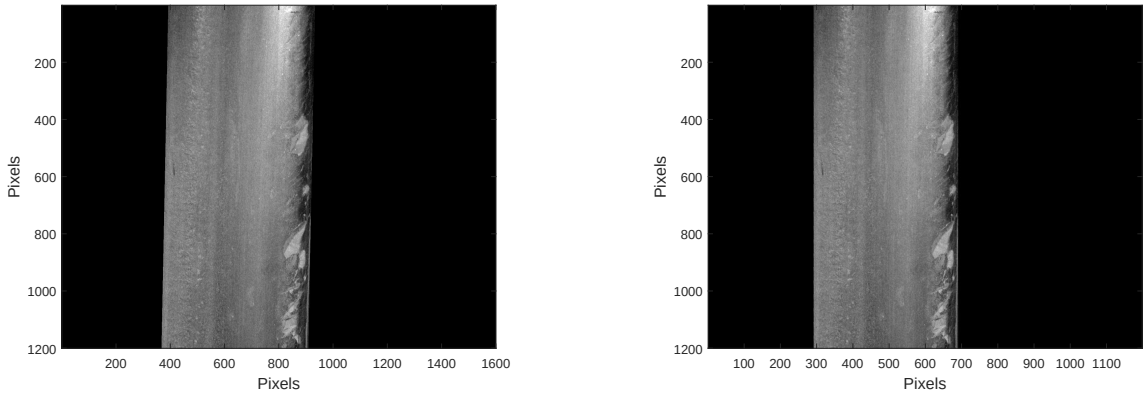
In which $\mathbf{A}_{crack}$, $\mathbf{A}_{flak}$, $\mathbf{A}_{head}$, $\mathbf{A}_{spall}$ and $\mathbf{A}_{norm}$ denotes the cracking, flaking, head check, spalling and normal condition of the rail head, respectively.

4.1.1 Geometric Correction

The geometric correction through a two-dimensional affine transformation aims to facilitate the image processing transforming the original image parallel to y axis of the cartesian coordinate system [78–80]. The affine transformation consists of a linear transformation followed by a translation. This transformation preserves the parallelism property. If two lines are parallel before the transformation, these lines are also parallel after processing. The expression that describes the affine transformation is

$$\begin{bmatrix} x^* \\ y^* \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} x_0 \\ y_0 \end{bmatrix}, \tag{4.2}$$

where a, b, c, d, x_0 and y_0 are the parameters of the affine transformation. The variables a, b, c and d comprise a rotation operation and the translation is given by x_0 and y_0. The parameters x and y are the original coordinates and x^* and y^* the transformed coordinates. Figure 26 shows an image before and after an affine transformation.



(a) (b)

Figure 26: An example of the result of the affine transformation when the cracking is considered. a) The image before the affine transformation. b) The image after the affine transformation.

4.1.2 Feature Extraction Based on Gray Level Co-Occurrence Matrix

The GLCM has been one of the most used methods to analyze textures on images and it is a two-dimensional dependence matrix that considers the spatial relationship between neighboring pixels [81–85]. The GLCM describes the textures of the images based on the frequency in which two gray levels separated by a distance d in a θ direction occur in the image. The texture information of an image can be described by GLCM

matrix, as shown below

$$\mathbf{G} = \begin{bmatrix} m(1,1) & m(1,2) & \cdots & m(1,n) \\ m(2,1) & m(2,2) & \cdots & m(2,n) \\ \vdots & \vdots & \ddots & \vdots \\ m(n,1) & m(n,2) & \cdots & m(n,n) \end{bmatrix}, \tag{4.3}$$

where $m(i,j)$ denotes the value of the matrix element having index (i,j) and n is the number of gray levels present in the image. For the construction of a GLCM, we must define the spatial relationship composed of the distance d, in pixel unit, and the adopted direction from the reference pixel, denoted by θ. Figure 27 shows us the possible angles to be adopted from the reference pixel.

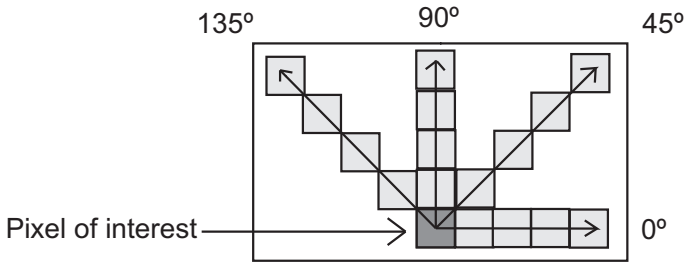


Figure 27: In addition to $\theta = 0°$, we can choose other three values to the reference angle. Thus, the GLCM provided from Figure 27 has $\theta = \{0°, 45°, 90°, 135°\}$ and $d = 4$.

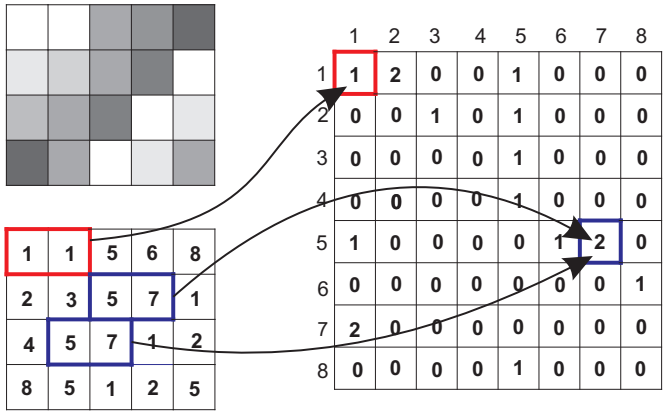


Figure 28: An example of GLCM construction from a small image, where $d = 1$, $\theta = 0°$ and the gray levels range from 1 to 8.

Figure 28 show us how GLCM is calculated for $d = 1$ and $\theta = 0°$. The element $m(1,1)$ contains the value 1 because there is only one instance in the input image where two horizontally adjacent pixels have the values 1 and 1, respectively. The element $m(5,7)$ contains the value 2 because it has two instances where two horizontally adjacent pixels have the values 5 and 7. Thus, each element $m(i,j)$ of the GLCM represents the instances where there is a transition from gray level i to j considering the distance d between the two neighboring pixels in the direction θ. The GLCM can be normalized by dividing each

entry by the number of neighboring resolution cell pairs, as follows

$$p\,(i,j) = \frac{m(i,j)}{\sum\limits_{i=0}^{n}\sum\limits_{j=0}^{n} m\,(i,j)}. \tag{4.4}$$

Each value $p(i,j)$ is the probability of a transition between gray levels under specified conditions of distance and direction[86]. Based on the co-occurrence matrices [82][83], we can use the energy, contrast, correlation and homogeneity to represent characteristic textures.

Energy [86, 87], also known as the second angular *momentum*, evaluates the uniformity texture of an image. The expression that describes the energy is

$$ASM(i,j) = \sum_{i=1}^{n}\sum_{j=1}^{n} p^2(i,j). \tag{4.5}$$

The Contrast [84, 85], also called the variance or inertia, measures the presence of unexpected transitions of gray levels in image. It can be calculated as follows

$$CON(i,j) = \sum_{i=1}^{n}\sum_{j=1}^{n} (i-j)^2 p(i,j). \tag{4.6}$$

Correlation [84, 88], measures the joint probability occurrence of the specified pixel pairs and it returns a measure of how a pixel is correlated with its neighbor along the whole image

$$COR(i,j) = \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{(i-\mu_i)(j-\mu_j)}{\sigma_i\sigma_j} p(i,j), \tag{4.7}$$

where μ_i, μ_j, σ_i and σ_j are the means and standard deviations of the pixels i and j, respectively.

Homogeneity [88, 89], returns a value that measures the proximity of the distribution of elements in GLCM with respect to its diagonal. It is given by

$$HOM(i,j) = \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{p(i,j)}{1+|i-j|}, \tag{4.8}$$

where $|\ |$ denotes the modulus operator.

It is important to state that all the abovementioned characteristic textures are used as input to the classifiers discussed in this work.

4.2 THE PROPOSAL: SET-MEMBERSHIP COMBINED WITH IST2-FLS AND ULST2-FLS

Under a limited number of epochs, an effective training is a very difficult task to be accomplished. The situation worsens when a specific application demands a periodic

update of the model, i.e., the classification model needs to be retrained to cover new patterns. Another important aspect is to attain better convergence speed during the training phase of the classification model without decreasing its accuracy.

According to [26], the type-2 FLS has the potential to provide better performance than a type-1 FLS (e.g., [63–71]). Considering the IST2-FLS and ULST2-FLS, which were presented in Chapter 3, we can state that one of the main differences between type-2 FLS and type-1 FLS is the type-reducer, which is responsible for converting the output type-2 membership function in its output type-1 membership function for posteriorly be defuzzified [28]. Liang *et al.* in [27] stated that type-2 FLSs are computationally intensive because the type-reducer is very intensive. As a result, the training phase is extremely computationally intensive.

To deal with this limitation during the training phase, this work proposes an IST2-FLS and ULT2-FLS combined with set-membership fuzzy logic system (SMFLS) for classifying events in a rail head (cracking, flaking, head check, spalling and normal condition), when a reduced set of features is extracted from the matrix $\mathbf{A}$. The proposed classifiers are trained by steepest descent method, using max-product composition, product implication, height type-reduction and making use of gaussian primary membership function with uncertain mean [28].

Let $\left(\mathbf{x}^{(q)} : y^{(q)}\right)$ be an input-output data set, where q denotes the qth iteration. We now wish to design an IST2-FLS or ULST2-FLS such that the following cost function is minimized:

$$J\left(\mathbf{w}^{(q)}\right) = \frac{1}{2}\left[f_{s2}\left(\mathbf{x}^{(q)}\right) - y^{(q)}\right]^2, \tag{4.9}$$

where $f_{s2}\left(\mathbf{x}^{(q)}\right)$ is the output of IST2-FLS or ULST2-FLS. Considering the membership functions as gaussian functions with a constant standard deviation σ, an uncertain mean situated in $m \in [m_1, m_2]$, $k = 1, \ldots, p$ inputs, and θ_l as the weight associated with the l-th rule, $l = 1, \ldots, M$. From (4.9), we know that

$$\begin{aligned}
&\mathbf{w}^{(q)} = \\
&\Big[m_{1F_1^1}(q), \ldots, m_{1F_p^1}(q), \ldots, m_{1F_1^M}(q), \ldots, m_{1F_p^M}(q), \\
&m_{2F_1^1}(q), \ldots, m_{2F_p^1}(q), \ldots, m_{2F_1^M}(q), \ldots, m_{2F_p^M}(q), \\
&\sigma_{F_1^1}(q), \ldots, \sigma_{F_p^1}(q), \ldots, \sigma_{F_1^M}(q), \ldots, \sigma_{F_p^M}(q), \\
&\theta_1(q), \ldots, \theta_M(q)\Big]^T
\end{aligned} \tag{4.10}$$

and

$$\begin{aligned}
&\nabla J(\mathbf{w}^{(q)}) = \\
&\Bigg[\frac{\partial J(\mathbf{w}^{(q)})}{\partial m_{1F_1^1}(q)}, \ldots, \frac{\partial J(\mathbf{w}^{(q)})}{\partial m_{1F_p^1}(q)}, \ldots, \frac{\partial J(\mathbf{w}^{(q)})}{\partial m_{1F_1^M}(q)}, \ldots, \frac{\partial J(\mathbf{w}^{(q)})}{\partial m_{1F_p^M}(q)}, \\
&\frac{\partial J(\mathbf{w}^{(q)})}{\partial m_{2F_1^1}(q)}, \ldots, \frac{\partial J(\mathbf{w}^{(q)})}{\partial m_{2F_p^1}(q)}, \ldots, \frac{\partial J(\mathbf{w}^{(q)})}{\partial m_{2F_1^M}(q)}, \ldots, \frac{\partial J(\mathbf{w}^{(q)})}{\partial m_{2F_p^M}(q)}, \\
&\frac{\partial J(\mathbf{w}^{(q)})}{\partial \sigma_{F_1^1}(q)}, \ldots, \frac{\partial J(\mathbf{w}^{(q)})}{\partial \sigma_{F_p^1}(q)}, \ldots, \frac{\partial J(\mathbf{w}^{(q)})}{\partial \sigma_{F_1^M}(q)}, \ldots, \frac{\partial J(\mathbf{w}^{(q)})}{\partial \sigma_{F_p^M}(q)}, \\
&\frac{\partial J(\mathbf{w}^{(q)})}{\partial \theta_1(q)}, \ldots, \frac{\partial J(\mathbf{w}^{(q)})}{\partial \theta_M(q)}\Bigg]^T,
\end{aligned} \tag{4.11}$$

denote, respectively, the parameters vector and the gradient vector of a IST2-FLS or ULST2-FLS. The algorithm for updating the parameters vector is presented in Chapter 3. The equations used in the algorithm to update its parameters are

$$m_{1F_k^l}(q+1) = m_{1F_k^l}(q) - \alpha \frac{\partial J\left(\mathbf{w}^{(q)}\right)}{\partial m_{1F_k^l}(q)}, \tag{4.12}$$

$$m_{2F_k^l}(q+1) = m_{2F_k^l}(q) - \alpha \frac{\partial J\left(\mathbf{w}^{(q)}\right)}{\partial m_{2F_k^l}(q)}, \tag{4.13}$$

$$\sigma_{F_k^l}(q+1) = \sigma_{F_k^l}(q) - \alpha \frac{\partial J\left(\mathbf{w}^{(q)}\right)}{\partial \sigma_{F_k^l}(q)}, \tag{4.14}$$

and

$$\theta_l(q+1) = \theta_l(q) - \alpha \frac{\partial J\left(\mathbf{w}^{(q)}\right)}{\partial \theta_l(q)}. \tag{4.15}$$

where $\alpha \in \mathbb{R} \,|0 < \alpha \ll 1$ is the adopted step size.

The first-order derivative of $J(\mathbf{w}^{(q)})$ with regard to parameters $m_{1F_k^l}(q)$, $m_{2F_k^l}(q)$, $\sigma_{F_k^l}(q)$, $\theta_l(q)$ were shown for the first time in Chapter 3, in order to facilitate the comprehension and spreading the use of the IST2-FLS or ULST2-FLS in situations in which there are high levels of uncertainty.

In Chapter 2, the authors discussed the concept of SM [9–13] derived from adaptive filter theory [14], and combine it with type-1 and singleton/non-singleton FLS trained by steepest descent method, in order to reduce the computational complexity and to increase the convergence speed during the training phase. Following Chapter 2, we state that the SMFLS consists in the adoption of variable step size $\mu_e(q)$, instead of α, given by

$$\mu_e(q) = \begin{cases} 1 - \frac{\bar{\gamma}}{\left|e^{(q)}\right|}, & \text{if } \left|e^{(q)}\right| > \bar{\gamma} \\ 0, & \text{otherwise.} \end{cases}, \tag{4.16}$$

where $\left|e^{(q)}\right| = \left|f_{s2}(\mathbf{x}^{(q)}) - y^{(q)}\right|$ is upper bounded by a prescribed constraint $\bar{\gamma}$.

Even though $\mu_e(q)$ has been successfully employed in adaptive filtering, we point out that there are other criteria to substitute (4.16). Our proposal consists in to apply

the SMFLS and combines it with the classifiers based on IST2-FLS and ULST2-FLS. Additionally, we combine SM with IST2-FLS and ULST2-FLS, together with other concepts discussed in Chapter 2: delta rule delta (DRD) [10,15–18], local Lipschitz estimation (LLE) [10,17–21], variable step size (VS) [22–24] and variable step size adaptive algorithms (VSA) [25] to come up with additional savings of computational complexity during the training phase.

4.3 EXPERIMENTAL RESULTS

Performance analyses discussed in this section made use of measured data set provided by MRS Logística (https://www.mrs.com.br/). The data set is composed of images of rail head surface with a resolution of 1600×1200 pixels, captured by a camera attached to a special car service that runs through the railway. Three images of the rail head surface are taken per meter.

The acquired images fall within the following five classes: the normal condition of the rail head, cracking, flaking, head check and spalling. There are 88 images for each of the five classes. We show samples of such images in Figure 29.

The peak signal to noise ratio (PSNR) [90] is the most used parameter to measure a corrupted image quality when compared to the original one. Such parameter is the ratio between the maximum possible power of a signal and the power of corrupting noise that affects the fidelity of its representation. It is computationally lightweight and is usually expressed in terms of the logarithmic decibel scale. There is an inverse relationship between PSNR and MSE. So, the higher the value of PSNR indicates the best image quality. Considering a noise-free $m \times n$ monochrome image I and its noisy approximation K, we can calculate the PSNR from a corrupted image as follows

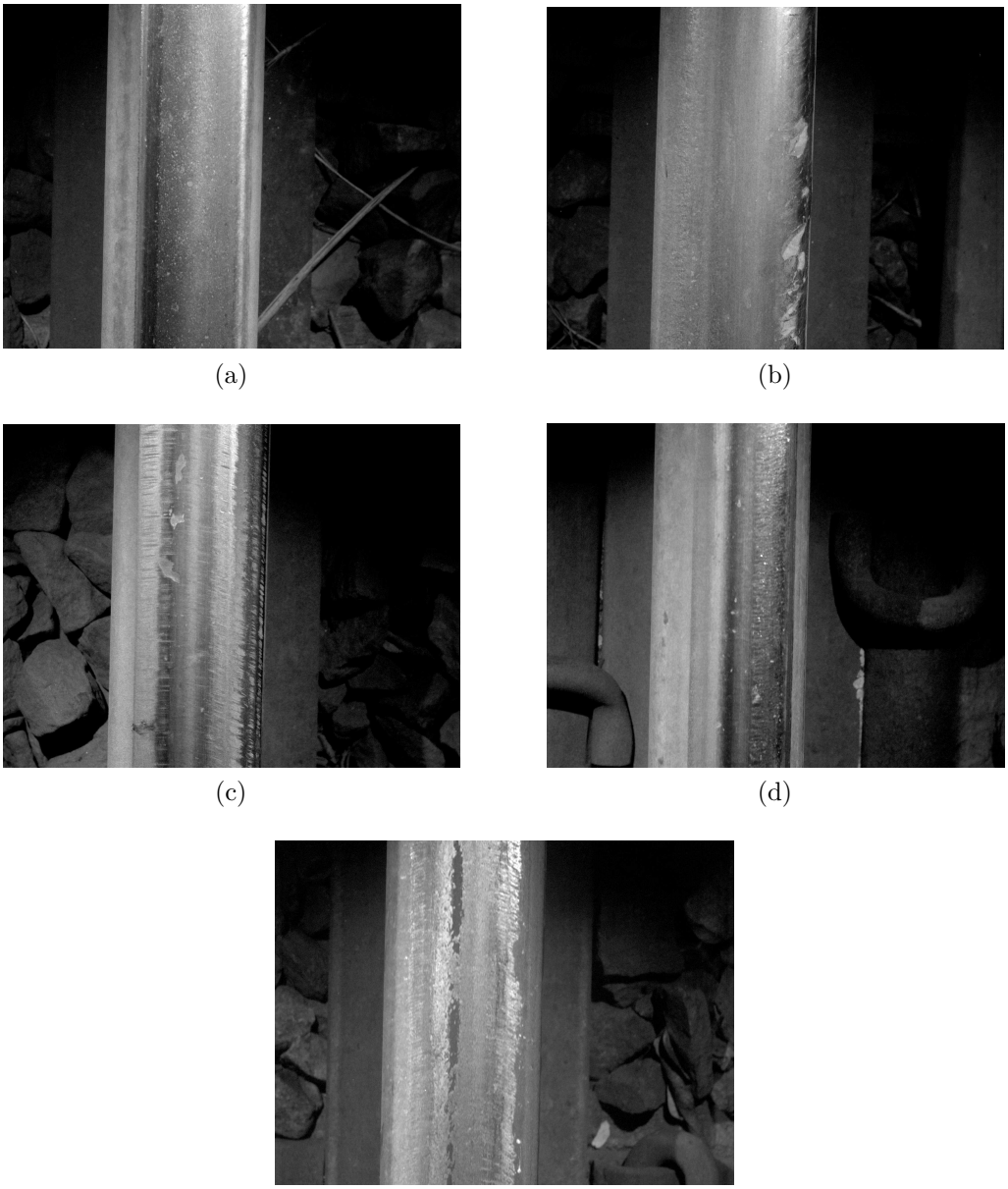
$$PSNR = 10\log_{10}\left(\frac{255^2}{MSE}\right), \tag{4.17}$$

where

$$MSE = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[I(i,j) - K(i,j)]^2. \tag{4.18}$$

and $I(i,j)$ and $K(i,j)$ denote the value of the element located in the (i,j) position in the I and K matrices, respectively.

In (4.17), the elements of the matrix are represented by using linear pulse-code modulation (PCM) [91] with B bits per sample, where the maximum is $2^B - 1$. Therefore, the value 255^2 denotes the maximum possible pixel value of the image, due to the fact of the pixels in this work are represented using 8 bits per element.



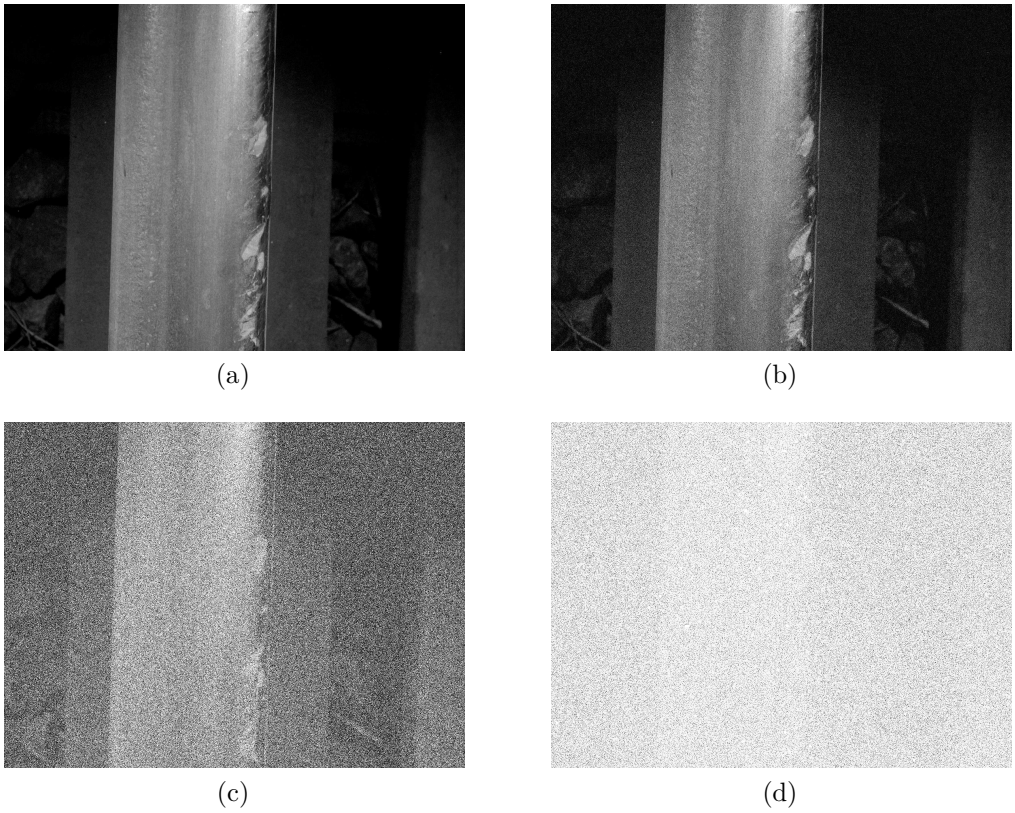
(a) (b) (c) (d) (e)

Figure 29: Typical samples for the five classes. a) Normal condition. b) Cracking. c) Flaking. d) Head check. e) Spalling.

For the representation of possible appearances by defects in the acquired images and in order to challenge the proposed models, we added three different intensities of additive white Gaussian noise (AWGN) in all data set, generating thus, images with PSNR of 20.65 dB, 6.34 dB and 2.43 dB, as shown in Figure 30. The adopted values for the PSNR are enough to report problems that can occur due to various factors, such as dirt in the lens of the equipment responsible for acquiring the images.

Also, in order to verify the effectiveness of the chosen AWGN, we added three different intensities of Salt and Pepper noise [92–94] in all data set. We adopted noise densities of 0.05, 0.2 and 0.08, generating thus, images with PSNR of 17.17 dB, 11.31 dB

and 5.36 dB, respectively. It is important to worth that the obtained results are aligned with those that will be discussed in Subsections 4.3.1, 4.3.2 and 4.3.3. This fact show us that the proposed models are able to classify events regardless of the added noise type.



(a) (b)

(c) (d)

Figure 30: Demonstrative image considering a cracking defect. a) Original image. b) Image corrupted with PSNR = 20.65 dB. c) Image corrupted with PSNR = 6.34 dB. d) Image corrupted with PSNR = 2.43 dB.

Following the block diagram of the classification technique depicted in Figure 25, we adopted techniques for geometric correction through affine transformation and feature extraction using GLCM. The GLCM adopted in this work was calculated using $\theta = \{0°, 45°, 90°, 135°\}$ and $d = 100$. For comparison purpose, we implemented the type-1 and singleton/non-singleton FLS [28] in order to classify defects in a rail head. Observing the block diagram in Figure 25, that $y^{(q)} = 1$ means that the matrix $\mathbf{A}$, constituted by pixels of taken image of the rail head is associated with the occurrence of the defect. On the other hand, $y^{(q)} = -1$ states that the matrix $\mathbf{A}$ is associated with the absence of defect.

To facilitate the understanding, acronyms are adopted in accordance with what was widely discussed in Chapter 2. Thus, we have the following techniques combined with IST2-FLS and ULST2-FLS: SMFLS, MSE SMFLS, $\left\|\nabla J(\mathbf{w}^{(q)})\right\|^2$ SMFLS, $\max(\nabla J(\mathbf{w}^{(q)})$ SMFLS, DRD FLS, DRD SMFLS, LLE FLS, VS SMFLS, MVS SMFLS, VSA SMFLS and MVSA SMFLS.

The adopted step size for the IST2-FLS and ULST2-FLS trained with the steepest descent method is $\alpha = 10^{-2}$. Considering the SMFLS, we heuristically have chosen $\bar{\gamma} = 0.4472$. For the implementation of the DRD FLS and the DRD SMFLS, we considered $\alpha_{DRD} = 10^{-8}$, $K_1 = 10^{-4}$, $K_2 = 0.99$ and $\rho_w = 20$. For the VS, the MVS, the VSA and the MVSA SMFLS we used $\alpha_{VS} = \alpha_{VSA} = 2 \times 10^{-4}$. Each IST2-FLS and ULST2-FLS is composed of four rules with uncertain mean ($M = 4$), two rules for the class that has the presence of the event and other two for the class that does not have the presence of the event. Furthermore, we equally and randomly distributed the data set in training and test ones and considered 100 epochs for the training phase.

4.3.1 Convergence Speed Analysis

Figures 31 and 32 show the convergence speed in terms of classification ratio for the cracking defect with PSNR of 20.65 dB and Figures 34 and 33 illustrate the convergence speed for the normal condition with PSNR of 2.43 dB, considering IST2-FLS and ULST2-FLS, respectively. For comparison, these figures also show the result for type-1 and singleton/non-singleton FLS in the same defect class. Note that singleton T1-FLS and non-singleton T1-FLS mean type-1 and singleton/non-singleton FLS, respectively. Although they are not presented here, the proposed models offer similar convergence speed for the other classes of defects.

Most of the proposals returned higher convergence speed than those based on type-1 and singleton/non-singleton FLS. Figures 32 and 34 also show that ULST2-FLS achieves superior results than IST2-FLS when working with the uncertainty associated with the input data.

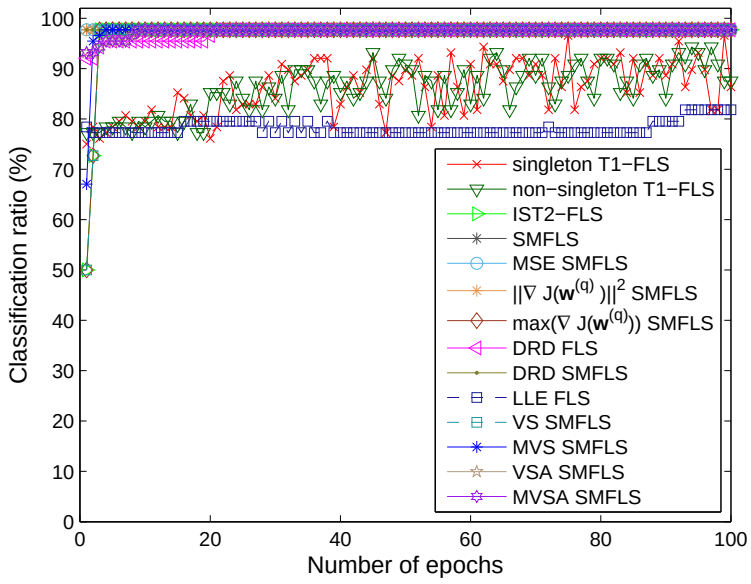


Figure 31: The convergence speed of the IST2-FLS for cracking with PSNR of 20.65 dB.

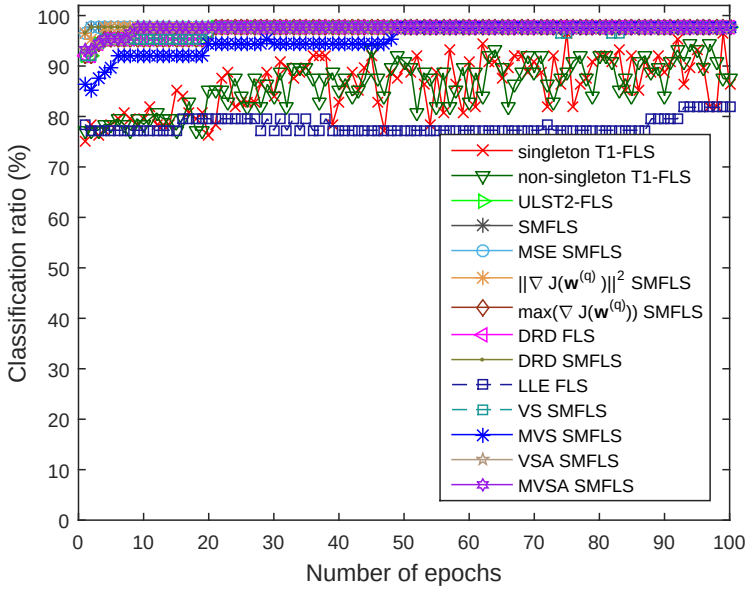


Figure 32: The convergence speed of the ULST2-FLS for cracking with PSNR of 20.65 dB.

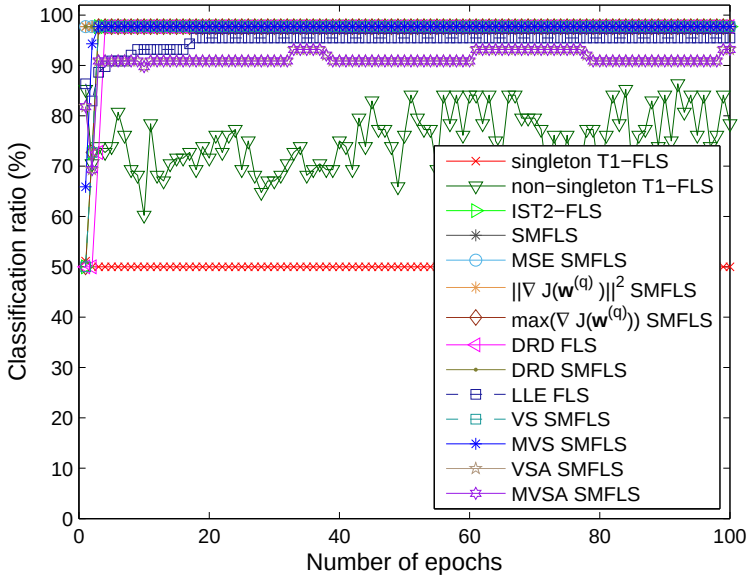


Figure 33: The convergence speed of the IST2-FLS for normal condition with PSNR of 2.43 dB.

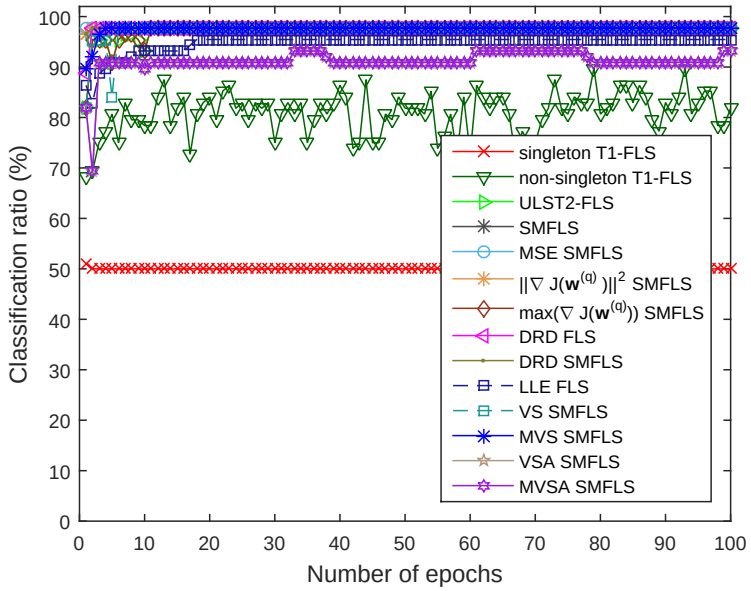


Figure 34: The convergence speed of the ULST2-FLS for normal condition with PSNR of 2.43 dB.

4.3.2 Computational Complexity Reduction

This section discusses the computational complexity reduction achieved by the proposed models. According to Chapter 2, to make this analysis we adopted the percentage of times when $\mu(q)$ is not updated, as shown below:

$$\tau[\%] = 100 \times \frac{N_{nup}}{N}, \tag{4.19}$$

where N_{nup} is the number of data set that results in $\mu(q) = 0$, according to the SM concept and N is the cardinality of the measured data set. This parameter is important because $\mu(q) = 0$ means that the parameters vector $\mathbf{w}^{(q+1)} = \mathbf{w}^{(q)}$ and, as a consequence, no computational effort to update $\mathbf{w}^{(q+1)}$ in the qth iteration is needed.

Figures 35 and 36 show the computational complexity reduction from zero to the hundredth epoch for normal condition class considering the original data set. Figures 37 and 38 follow the same information for spalling class with PSNR of 2.43 dB, considering the classifiers based on IST2-FLS and ULST2-FLS, respectively. Similar computational complexity reductions are obtained for the other classes of defects.

Table 9 shows the mean computational complexity reduction achieved in % for IST2-FLS and ULST2-FLS. This value is expressed by:

$$\tau_m\,[\%] = \frac{1}{N_e}\sum_{i=1}^{N_e}\tau\,(i)\,[\%]\,, \tag{4.20}$$

where $\tau\,(i)\,[\%]$ is the value of $\tau[\%]$ in the ith epoch of training and N_e is total number of epochs. Table 10 shows the best mean computational complexity reduction achieved in

% for IST2-FLS and ULST2-FLS. The efficiency $\eta[\%]$ is given by

$$\eta\,[\%] = \frac{1}{N_c}\sum_{j=1}^{N_c}\tau_{m,\max}\,[\%]\,, \tag{4.21}$$

where $\tau_{m,\max}$ is the highest value $\tau_m[\%]$ for the jth class and N_c is the total number of classes. In general, we note that $\max(\nabla J(\mathbf{w}^{(q)})$ SMFLS based techniques obtained the highest results in terms of $\tau_m\,[\%]$. Note that IST2-FLS achieved a mean computational complexity reduction of 98.2% for the original data set, 98.8% for PSNR= 20.65 dB, 98.9% for PSNR = 6.34 dB and 98.8% for PSNR = 2.43 dB and returned slightly better results when it is compared with ULST2-FLS, which achieved 98.5%, 98.8%, 98.7%, 98.8% for the original data set and all aforementioned PSNRs.

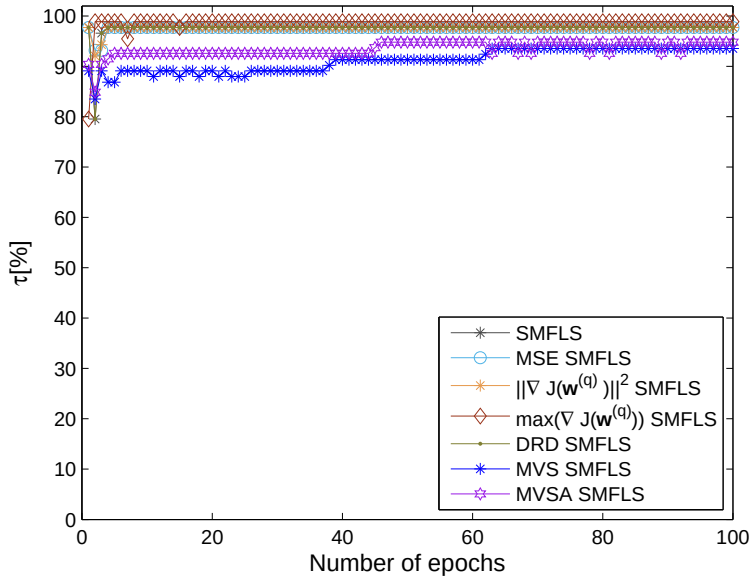


Figure 35: The computational complexity reduction of the IST2-FLS for normal condition from the original data set.

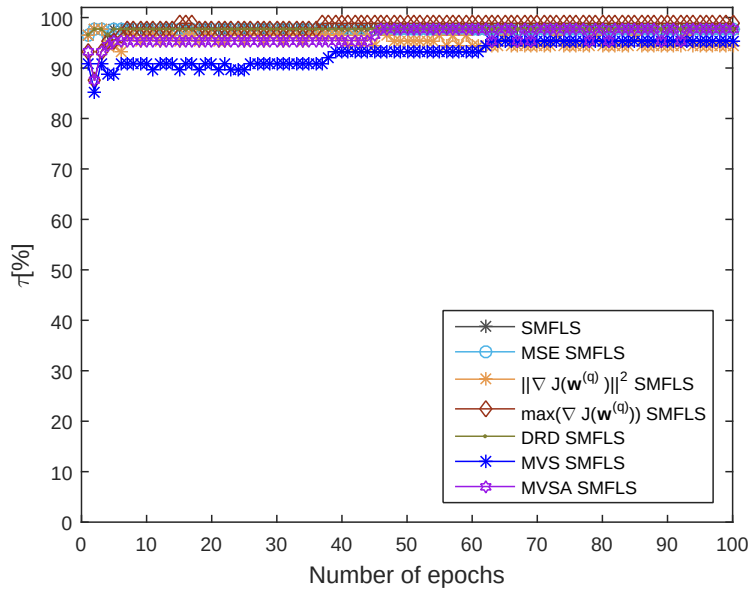


Figure 36: The computational complexity reduction of the ULST2-FLS for normal condition from the original data set.

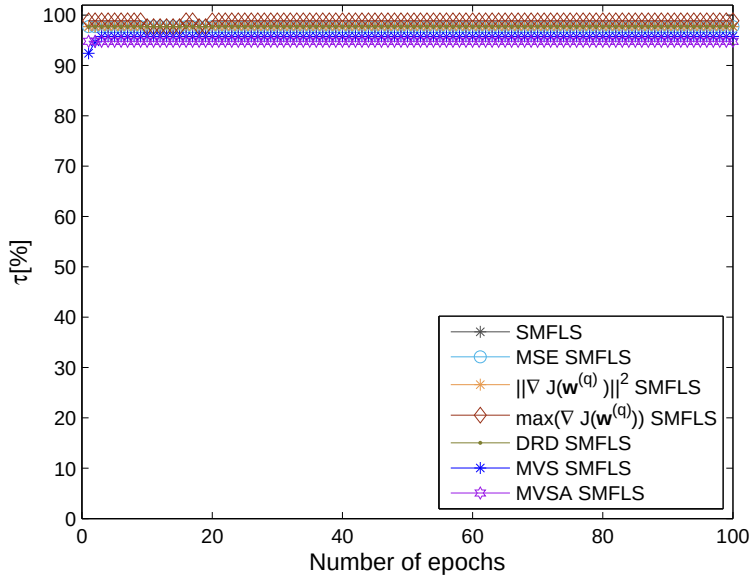


Figure 37: The computational complexity reduction of the IST2-FLS for spalling with PSNR of 6.34 dB.

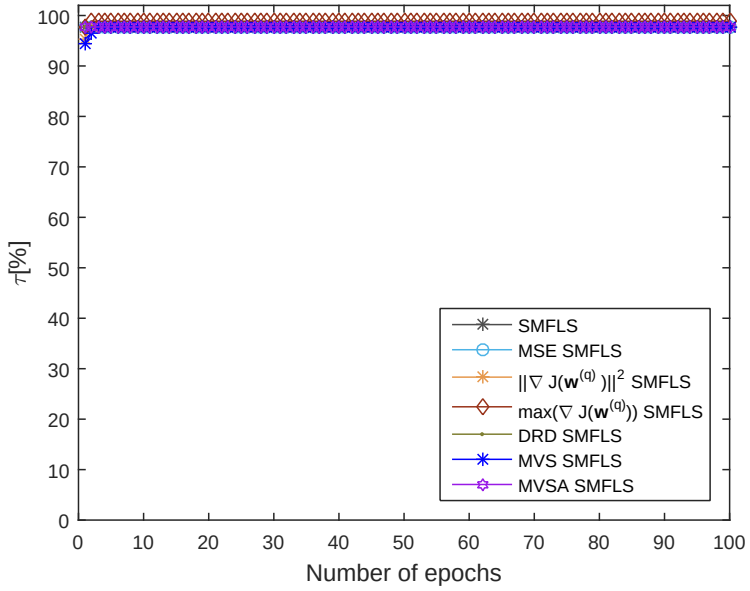


Figure 38: The computational complexity reduction of the ULST2-FLS for spalling with PSNR of 6.34 dB.

Table 9: Mean Computational complexity reduction of the IST2-FLS and ULST2-FLS for the normal condition, cracking, flaking, head check, spalling.

	Original data set		PSNR = 20.65 dB		PSNR = 6.34 dB		PSNR = 2.43 dB	
Events	IST2-FLS	ULST2-FLS	IST2-FLS	ULST2-FLS	IST2-FLS	ULST2-FLS	IST2-FLS	ULST2-FLS
Normal condition	τ_m [%]	τ_m [%]	τ_m [%]	τ_m [%]	τ_m [%]	τ_m [%]	τ_m [%]	τ_m [%]
SMFLS	97.5	97.7	97.7	97.7	97.7	97.7	97.7	97.7
MSE SMFLS	97.6	97.7	97.7	97.7	97.7	97.7	97.7	97.7
$\left\|\nabla J(\mathbf{w}^{(q)})\right\|^2$ SMFLS	97.6	95.4	97.7	97.7	97.7	97.7	97.7	97.7
$\max(\nabla J(\mathbf{w}^{(q)})$ SMFLS	98.6	98.2	98.7	98.8	98.9	98.3	98.9	98.7
DRD SMFLS	97.5	97.7	97.7	97.7	97.7	97.7	97.7	97.7
MVS SMFLS	90.2	93.0	97.0	97.6	87.8	97.7	97.4	97.6
MVSA SMFLS	93.5	96.4	60.9	97.7	61.6	97.7	61.6	91.2
Cracking	τ_m [%]	τ_m [%]	τ_m [%]	τ_m [%]	τ_m [%]	τ_m [%]	τ_m [%]	τ_m [%]
SMFLS	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7
MSE SMFLS	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7
$\left\|\nabla J(\mathbf{w}^{(q)})\right\|^2$ SMFLS	97.7	96.4	97.7	97.7	97.7	97.7	97.7	97.6
$\max(\nabla J(\mathbf{w}^{(q)})$ SMFLS	98.8	98.7	98.8	98.6	98.9	98.8	98.9	98.9
DRD SMFLS	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7
MVS SMFLS	97.4	97.5	97.4	95.5	76.4	95.5	76.3	97.6
MVSA SMFLS	60.8	97.7	61.3	97.5	61.7	97.7	61.6	97.7
Flaking	τ_m [%]	τ_m [%]	τ_m [%]	τ_m [%]	τ_m [%]	τ_m [%]	τ_m [%]	τ_m [%]
SMFLS	54.3	97.7	97.7	97.7	97.7	97.7	97.3	97.7
MSE SMFLS	55.3	97.7	97.7	97.7	97.7	97.7	97.7	97.7
$\left\|\nabla J(\mathbf{w}^{(q)})\right\|^2$ SMFLS	55.3	97.7	97.7	97.7	97.7	97.7	97.7	97.7
$\max(\nabla J(\mathbf{w}^{(q)})$ SMFLS	96.1	98.3	98.6	98.8	98.9	98.8	98.9	98.8
DRD SMFLS	51.4	97.7	97.7	97.7	97.7	97.7	97.7	97.7
MVS SMFLS	49.9	93.4	97.3	97.1	75.1	95.4	75.0	97.7
MVSA SMFLS	59.7	96.9	61.4	97.7	61.6	97.7	61.6	97.7
Head Check	τ_m [%]	τ_m [%]	τ_m [%]	τ_m [%]	τ_m [%]	τ_m [%]	τ_m [%]	τ_m [%]
SMFLS	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7
MSE SMFLS	97.7	97.7	96.6	97.7	97.7	97.7	97.7	97.7
$\left\|\nabla J(\mathbf{w}^{(q)})\right\|^2$ SMFLS	97.7	97.7	66.0	97.7	97.7	97.7	97.7	97.7
$\max(\nabla J(\mathbf{w}^{(q)})$ SMFLS	98.8	98.5	98.9	98.8	98.9	98.8	98.9	98.8
DRD SMFLS	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7
MVS SMFLS	97.4	94.3	97.4	97.6	76.4	97.2	76.2	95.5
MVSA SMFLS	59.7	97.6	62.0	97.7	61.7	96.7	61.6	97.7
Spalling	τ_m [%]	τ_m [%]	τ_m [%]	τ_m [%]	τ_m [%]	τ_m [%]	τ_m [%]	τ_m [%]
SMFLS	97.7	97.7	77.2	97.7	97.7	97.7	97.7	97.7
MSE SMFLS	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7
$\left\|\nabla J(\mathbf{w}^{(q)})\right\|^2$ SMFLS	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7
$\max(\nabla J(\mathbf{w}^{(q)})$ SMFLS	98.9	98.7	100.0	98.8	98.8	98.9	98.5	98.9
DRD SMFLS	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7
MVS SMFLS	50.2	95.1	93.8	95.4	81.6	97.7	76.3	97.7
MVSA SMFLS	70,6	97.5	95.7	97.7	70.6	97.7	61.6	97.7

Table 10: Best mean computational complexity reduction achieved in [%] for IST2-FLS and ULST2-FLS.

IST2-FLS								
	Original data set		PSNR = 20.65 dB		PSNR = 6.34 dB		PSNR = 2.43 dB	
Events	τ_m [%]	Choosen Technique	τ_m [%]	Choosen Technique	τ_m [%]	Choosen Technique	τ_m [%]	Choosen Technique
Normal condition	98.6	$\max(\nabla J(\mathbf{w}^{(q)})$ SMFLS	97.7	$\max(\nabla J(\mathbf{w}^{(q)})$ SMFLS	98.9	$\max(\nabla J(\mathbf{w}^{(q)})$ SMFLS	98.9	$\max(\nabla J(\mathbf{w}^{(q)})$ SMFLS
Cracking	98.8	$\max(\nabla J(\mathbf{w}^{(q)})$ SMFLS	98.8	$\max(\nabla J(\mathbf{w}^{(q)})$ SMFLS	98.9	$\max(\nabla J(\mathbf{w}^{(q)})$ SMFLS	98.9	$\max(\nabla J(\mathbf{w}^{(q)})$ SMFLS
Flaking	96.1	$\max(\nabla J(\mathbf{w}^{(q)})$ SMFLS	98.6	$\max(\nabla J(\mathbf{w}^{(q)})$ SMFLS	98.9	$\max(\nabla J(\mathbf{w}^{(q)})$ SMFLS	98.9	$\max(\nabla J(\mathbf{w}^{(q)})$ SMFLS
Head check	98.8	$\max(\nabla J(\mathbf{w}^{(q)})$ SMFLS	98.9	$\max(\nabla J(\mathbf{w}^{(q)})$ SMFLS	98.9	$\max(\nabla J(\mathbf{w}^{(q)})$ SMFLS	98.9	$\max(\nabla J(\mathbf{w}^{(q)})$ SMFLS
Spalling	98.9	$\max(\nabla J(\mathbf{w}^{(q)})$ SMFLS	100.0	$\max(\nabla J(\mathbf{w}^{(q)})$ SMFLS	98.8	$\max(\nabla J(\mathbf{w}^{(q)})$ SMFLS	98.5	$\max(\nabla J(\mathbf{w}^{(q)})$ SMFLS
Efficiency (η)	**98.2**		**98.8**		**98.9**		**98.8**	

ULST2-FLS								
	Original data set		PSNR = 20.65 dB		PSNR = 6.34 dB		PSNR = 2.43 dB	
Events	τ_m [%]	Choosen Technique	τ_m [%]	Choosen Technique	τ_m [%]	Choosen Technique	τ_m [%]	Choosen Technique
Normal condition	98.2	$\max(\nabla J(\mathbf{w}^{(q)})$ SMFLS	98.8	$\max(\nabla J(\mathbf{w}^{(q)})$ SMFLS	98.3	$\max(\nabla J(\mathbf{w}^{(q)})$ SMFLS	98.7	$\max(\nabla J(\mathbf{w}^{(q)})$ SMFLS
Cracking	98.7	$\max(\nabla J(\mathbf{w}^{(q)})$ SMFLS	98.6	$\max(\nabla J(\mathbf{w}^{(q)})$ SMFLS	98.8	$\max(\nabla J(\mathbf{w}^{(q)})$ SMFLS	98.9	$\max(\nabla J(\mathbf{w}^{(q)})$ SMFLS
Flaking	98.3	$\max(\nabla J(\mathbf{w}^{(q)})$ SMFLS	98.8	$\max(\nabla J(\mathbf{w}^{(q)})$ SMFLS	98.8	$\max(\nabla J(\mathbf{w}^{(q)})$ SMFLS	98.8	$\max(\nabla J(\mathbf{w}^{(q)})$ SMFLS
Head check	98.5	$\max(\nabla J(\mathbf{w}^{(q)})$ SMFLS	98.8	$\max(\nabla J(\mathbf{w}^{(q)})$ SMFLS	98.8	$\max(\nabla J(\mathbf{w}^{(q)})$ SMFLS	98.8	$\max(\nabla J(\mathbf{w}^{(q)})$ SMFLS
Spalling	98.7	$\max(\nabla J(\mathbf{w}^{(q)})$ SMFLS	98.8	$\max(\nabla J(\mathbf{w}^{(q)})$ SMFLS	98.9	$\max(\nabla J(\mathbf{w}^{(q)})$ SMFLS	98.9	$\max(\nabla J(\mathbf{w}^{(q)})$ SMFLS
Efficiency (η)	**98.5**		**98.8**		**98.7**		**98.8**	

4.3.3 Classification Rate Analysis

Numerical results presented in Table 11 show that the correct classification rate for IST2-FLS and ULST2-FLS on the hundredth epoch of training, considering the normal condition of the rail head, cracking, flaking, head check and spalling. We see that proposed models can achieve fast convergence speed and, additionally, they provide higher accuracy when compared with the classical IST2-FLS and ULST2-FLS. The results are even better when compared with type-1 and singleton/non-singleton FLS especially when the images are corrupted by AWGN.

The final results for both techniques are shown in Table 12. In this case the efficiency $\rho\,[\%]$ is expressed by

$$\rho\,[\%] = \frac{1}{N_c}\sum_{j=1}^{N_c} CR_{j,\max}\,[\%]\,, \tag{4.22}$$

where $CR_{j,\max}$ is the highest classification ratio on the hundredth epoch of training, considering the jth class. According to the results, the IST2-FLS had a similar efficiency when compared with ULST2-FLS, with 97.7% of accuracy for both, indicating that the proposed models misclassified 2 images of a total of 88, considering each of the five classes. The results are the same for the original data set and for those corrupted by AWGN, demonstrating the consistency of the proposed models.

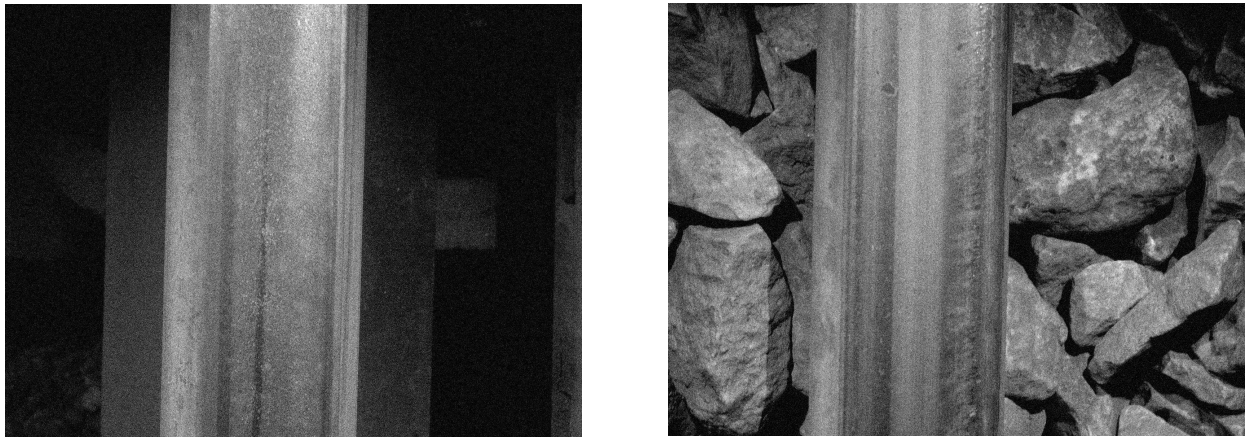
Considering the flaking defect, note that Figure 39 shows the images that were misclassified by the proposed models, stating that they are associated with the absence of a defect (in other words, $y^{(q)} = -1$). These misclassifications occurred due to the presence of the rail foot and stones in conflict with the rail head in Figure 39. This conflict affects the quality of geometric correction based on two-dimensional affine transformation and the feature extraction based on GLCM, worsening, thus, the classification of rail head defects. Therefore, further investigation to deal with this situation is needed to improve the model‘s performance.

Table 11: Classification Rate of the IST2-FLS and ULST2-FLS for the normal condition, cracking, flaking, head check, spalling.

	Original data set				PSNR = 20.65 dB				PSNR = 6.34 dB				PSNR = 2.43 dB			
Events	IST2-FLS		ULST2-FLS		IST2-FLS		ULST2-FLS		IST2-FLS		ULST2-FLS		IST2-FLS		ULST2-FLS	
Normal Condition	Training	Test	Training	Test	Training	Test	Training	Test	Training	Test	Training	Test	Training	Test	Training	Test
Steepest Descent	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7
SMFLS	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7
MSE SMFLS	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7
$\left\|\nabla J(\mathbf{w}^{(q)})\right\|^2$ SMFLS	97.7	97.7	94.3	94.3	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7
$\max(\nabla J(\mathbf{w}^{(q)})$ SMFLS	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7
DRD FLS	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7
DRD SMFLS	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7
LLE FLS	97.7	97.7	93.2	93.2	97.7	97.7	97.7	97.7	93.2	93.2	93.2	93.2	95.4	95.4	95.4	95.4
VS SMFLS	97.7	97.7	95.4	95.4	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7
MVS SMFLS	79.5	79.5	95.4	95.4	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7
VSA SMFLS	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	93.1	93.1
MVSA SMFLS	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	93.1	93.1
Cracking	Training	Test	Training	Test	Training	Test	Training	Test	Training	Test	Training	Test	Training	Test	Training	Test
Steepest Descent	95.4	95.4	95.4	95.4	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7
SMFLS	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7
MSE SMFLS	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7
$\left\|\nabla J(\mathbf{w}^{(q)})\right\|^2$ SMFLS	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7
$\max(\nabla J(\mathbf{w}^{(q)})$ SMFLS	95.4	95.4	95.4	95.4	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7
DRD FLS	95.4	95.4	95.4	95.4	95.4	95.4	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7
DRD SMFLS	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7
LLE FLS	95.4	95.4	95.4	95.4	96.6	96.6	81.8	81.8	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7
VS SMFLS	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7
MVS SMFLS	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	95.4	95.4	97.7	97.7	97.7	97.7
VSA SMFLS	97.7	97.7	97.7	97.7	96.6	96.6	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7
MVSA SMFLS	97.7	97.7	97.7	97.7	96.6	96.6	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7
Flaking	Training	Test	Training	Test	Training	Test	Training	Test	Training	Test	Training	Test	Training	Test	Training	Test
Steepest Descent	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7
SMFLS	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7
MSE SMFLS	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7
$\left\|\nabla J(\mathbf{w}^{(q)})\right\|^2$ SMFLS	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7
$\max(\nabla J(\mathbf{w}^{(q)})$ SMFLS	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7
DRD FLS	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7
DRD SMFLS	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7
LLE FLS	79.5	79.5	79.5	79.5	95.4	95.4	97.7	97.7	95.4	95.4	97.7	97.7	95.4	95.4	95.4	95.4
VS SMFLS	95.4	95.4	95.4	95.4	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	95.4	95.4	97.7	97.7
MVS SMFLS	95.4	95.4	97.7	97.7	97.7	97.7	97.7	97.7	95.4	95.4	95.4	95.4	95.4	95.4	97.7	97.7
VSA SMFLS	97.7	97.7	97.7	97.7	95.4	95.4	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7
MVSA SMFLS	97.7	97.7	97.7	97.7	95.4	95.4	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7
Head Check	Training	Test	Training	Test	Training	Test	Training	Test	Training	Test	Training	Test	Training	Test	Training	Test
Steepest Descent	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7
SMFLS	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7
MSE SMFLS	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7
$\left\|\nabla J(\mathbf{w}^{(q)})\right\|^2$ SMFLS	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7
$\max(\nabla J(\mathbf{w}^{(q)})$ SMFLS	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7
DRD FLS	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7
DRD SMFLS	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7
LLE FLS	86.4	86.4	86.4	86.4	97.7	97.7	97.7	97.7	96.6	96.6	96.6	96.6	97.7	97.7	97.7	97.7
VS SMFLS	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7
MVS SMFLS	97.7	97.7	95.4	95.4	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	95.4	95.4
VSA SMFLS	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7
MVSA SMFLS	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7
Spalling	Training	Test	Training	Test	Training	Test	Training	Test	Training	Test	Training	Test	Training	Test	Training	Test
Steepest Descent	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7
SMFLS	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7
MSE SMFLS	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7
$\left\|\nabla J(\mathbf{w}^{(q)})\right\|^2$ SMFLS	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7
$\max(\nabla J(\mathbf{w}^{(q)})$ SMFLS	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7
DRD FLS	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7
DRD SMFLS	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7
LLE FLS	90.9	90.9	90.9	90.9	95.4	95.4	95.4	95.4	95.4	95.4	95.4	95.4	97.7	97.7	97.7	97.7
VS SMFLS	95.4	95.4	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	95.4	95.4	97.7	97.7	97.7	97.7
MVS SMFLS	95.4	95.4	95.4	95.4	95.4	95.4	95.4	95.4	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7
VSA SMFLS	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7
MVSA SMFLS	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7

Table 12: The comparative classification rate.

	IST2-FLS				ULST2-FLS			
Events	Original data set	PSNR = 20.65 dB	PSNR = 6.34 dB	PSNR = 2.43 dB	Original data set	PSNR = 20.65 dB	PSNR = 6.34 dB	PSNR = 2.43 dB
Normal condition	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7
Cracking	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7
Flaking	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7
Head Check	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7
Spalling	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7
Efficiency (ρ)	**97.7**	**97.7**	**97.7**	**97.7**	**97.7**	**97.7**	**97.7**	**97.7**



(a) (b)

Figure 39: Images that were misclassified by the proposed models, considering the flaking defect. a) Presence of the rail foot in conflict with the rail head. b) Presence of stones in conflict with the rail head.

4.4 SUMMARY

In this Chapter, we have combined the set-membership concept from adaptive filter theory to reduce the computational complexity associated with the training phase of the IST2-FLS and ULST2-FLS and analyzed their effectiveness to classify typical rail head defects. Additionally, we outlined the set-membership combined with the delta rule delta, the local lipschitz estimation, the variable step size and the variable step size adaptive techniques to yield additional computational complexity savings.

Considering a reduced number of epochs, the numerical results obtained through the use of measured data set showed that the IST2-FLS and ULST2-FLS combined with the proposals result in fast convergence rates. Additionally, these proposals can make the IST2-FLS and ULST2-FLS achieve higher classification rates than those obtained with type-1 and singleton/non-singleton FLS especially when there are uncertainties in the input data and the number of epochs in the training phase is limited.

In all analyzed situations, which are representative for the addressed classification problem, the proposed models turn out to be an attractive option due to their high efficiency in reducing computational complexity, reaching percentages higher than 98.2% for both IST2-FLS and ULST2-FLS.

Additionally, it is notorious that ULST2-FLS handled satisfactorily the presence of uncertainties in the measured data showing better classification rates than IST2-FLS and type-1 and singleton/non-singleton FLS.

5 CONCLUSION

This Chapter summarizes the conclusions of this thesis by addressing the proposal for fault classification in a switch machine and the classification of rail head defects.

Chapter 2 focuses on the classification of faults in a switch machine. With this regards, we introduced the use of SM concept, derived from the adaptive filter theory, into the training procedure of type-1 and singleton/non-singleton FLS, in order to reduce computational complexity and to increase convergence speed. Also, we present different criteria for using together with set-membership. Furthermore, we discuss the usefulness of DRD, LLE, VS and VSA to yield additional improvement in terms of computational complexity reduction and convergence speed. Based on data set provided by a Brazilian railway company, which covers the four possible faults in a switch machine, we presented performance analysis in terms of classification ratio, convergence speed and computational complexity reduction. The reported results show that the proposed models result in improved convergence speed, slightly higher classification ratio and remarkable computation complexity reduction when we limit the number of epochs for training, which may be required due to real time constraint or low computational resource availability.

Chapter 3 proposes a ULST2-FLS trained by steepest descent method for fault classification in a switch machine. Considering the height type-reduction, the deductions of equations from ULST2-FLS and IST2-FLS are presented for the first time in the literature. Results obtained through the use of real measured data revealed that both type-2 FLS discussed in this work converges faster. Classification rates are competitive when compared with those obtained with the previous approaches addressed in the literature, when a reduced number of epochs is considered (Bayes, multilayer perceptron neural network and type-1 FLS tuned by steepest descent method and conjugate gradient method). In any case, the classifiers proposed in that Chapter turn out to be an extremely attractive option due to their high convergence speed. Explained this fact by modeling of the database uncertainties through type-2 FLS. It was noted also that ULST2-FLS converges slightly faster than the IST2-FLS (the classical approach used in the literature).

Chapter 4 addresses how to reduce the computational complexity during training phase of a IST2-FLS and ULST2-FLS. We apply the concept of set-membership, and combine it with IST2-FLS and ULST2-FLS, both trained by steepest descent method, in order to reduce their computational complexity and to increase the convergence speed during the training phase. Furthermore, we apply the proposal to classify the main types of rail head defects that are commonly studied due to its severity and occurrence: cracking, flaking, head-check and spalling. The classification ratio yielded by the proposed models are higher than those obtained with the classical IST2-FLS and ULST2-FLS when a higher

degree of uncertainties is presented in the input data. Also, The set-membership concept combined with DRD, LLE, VS and VSA strongly contribute to reduce computational complexity during the training phase of type-2 FLS, under a limited number of epochs. Due to computational complexity reduction, we emphasize that the proposals demand low-cost (low-speed) processors, which is something very important to come up with green technologies for the railroad sector.

5.1 FUTURE WORKS

Future efforts can be addressed in order to:

- Prototype an equipment to be integrated into the existing switch machine supervision system in the company MRS Logística S.A..
- Prototype an equipment to be integrated into the existing rail head supervision system in the company MRS Logística S.A..
- Improve the preprocessing of images, through the research of segmentation tools, in order to get a better performance.
- Investigate the usefulness of interval non-singleton type-2 FLS and upper and lower non-singleton type-2 FLS to handle the presence of uncertainty in the measured data set.

REFERENCES

[1] HSE Potters Bar Investigation Board, "Train derailment at potters bar 10 may 2002," *Office of Rail Regulation*, May 2003.

[2] D. Milmo, "Report prompts calls for inquiry into potters bar and grayrigg rail crashes," *The Guardian*, Oct. 2008. [Online]. Available: http://www.theguardian.com/uk/2008/oct/23/transport-railtravel

[3] T. Asada and C. Roberts, "Development of an effective condition monitoring system for ac point machines," in *Proc. IET Conference on Railway Condition Monitoring and Non-Destructive Testing*, pp. 1–6, Nov. 2011.

[4] E. P. de Aguiar, F. M. A. Nogueira, R. P. Amaral, D. F. Fabri, S. C. Rossignoli, J. G. Ferreira, M. M. B. R. Vellasco, R. Tanscheit, P. C. S. Vellasco, and 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, no. 5, pp. 1175–1189, Jul. 2016.

[5] Z. Liu, W. Wang, X. Zhang, and W. Jia, "Inspection of rail surface defects based on image processing," in *Proc. International Asia Conference on Informatics in Control, Automation and Robotics*, vol. 1, pp. 472–475, Mar. 2010.

[6] Q. Li and S. Ren, "A real-time visual inspection system for discrete surface defects of rail heads," *IEEE Transactions on Instrumentation and Measurement*, vol. 61, no. 8, pp. 2189−−2199, Aug. 2012.

[7] D. J. Smith, *Reliability, Maintainability and Risk*, 7th ed. Butterworth-Heinemann, 2005.

[8] J. M. Nahman, *Dependability of Engineering Systems*, 1st ed. Springer Berlin Heidelberg, 2002.

[9] P. Clarke and R. de Lamare, "Low-complexity reduced-rank linear interference suppression based on set-membership joint iterative optimization for ds-cdma systems," *IEEE Transactions on Vehicular Technology*, vol. 60, no. 9, pp. 4324–4337, Nov. 2011.

[10] M. Salmento, E. Aguiar, and M. Ribeiro, "Reduced rank adaptive filter with variable step size for impulsive uwb plc systems," in *Proc. XXXIII Brazilian Telecommunications Symposium*, pp. 1–5, Sep. 2015.

[11] L. Wang and R. DeLamare, "Low-complexity constrained adaptive reduced-rank beamforming algorithms," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 49, no. 4, pp. 2114–2128, Oct. 2013.

[12] P. S. R. Diniz, *Adaptive Filtering: Algorithms and Practical Implementation*. Norwell, MA, USA: Kluwer Academic Publishers, 2002.

[13] M. Esfand Abadi and H. Eskandari, "Performance study of set-membership nlms algorithm over an adaptive incremental network," in *Proc. 3th International eConference on Computer and Knowledge Engineering*, pp. 257–261, Oct. 2013.

[14] S. Haykin, *Adaptive Filter Theory*, 3rd ed. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1996.

[15] R. A. Jacobs, "Increased rates of convergence through learning rate adaptation," *Neural Networks*, vol. 1, no. 4, pp. 295–307, Sep. 1988.

[16] Y. Nan, L. Jiarong, S. Shuangbao, and S. Binxuan, "Application of delta-bar-delta rules trained back-propagation neural networks in nuclear fusion pattern recognition," in *Proc. International Symposium on Intelligence Information Processing and Trusted Computing*, pp. 258–261, Oct. 2010.

[17] M. V. Ribeiro, "Signal processing techniques for power line communication and power quality applications," Ph.D. dissertation, State University of Campinas, Apr. 2005.

[18] M. V. Ribeiro, "Learning rate updating methods applied to adaptive fuzzy equalizers for broadband power line communications," *EURASIP Journal on Advances in Signal Processing*, vol. 2004, no. 16, pp. 2592–2599, Jan. 2004.

[19] G. D. Magoulas and M. N. Vrahatis, "Improving the convergence of the backpropagation algorithm using learning rate adaptation methods," *Neural Computation*, vol. 11, pp. 1769–1796, 1999.

[20] L. Armijo, "Minimization of functions having lipschitz-continuous first partial derivatives," *Pacific Journal of Mathematics*, vol. 16, pp. 1–3, Nov. 1966.

[21] W.-M. Lippe, T. Feuring, and A. Tenhagen, "A fuzzy-controlled delta-bar-delta learning rule," in *Proc. IEEE International Conference on Neural Networks*, vol. 3, pp. 1686–1690, Jun. 1994.

[22] R. Harris, D. Chabries, and F. Bishop, "A variable step (vs) adaptive filter algorithm," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 34, no. 2, pp. 309–316, Apr. 1986.

[23] S. Gelfand, Y. Wei, and J. Krogmeier, "The stability of variable step-size lms algorithms," *IEEE Transactions on Signal Processing*, vol. 47, no. 12, pp. 3277–3288, Dec. 1999.

[24] S. Nunoo, U. Chude-Okonkwo, R. Ngah, and Y. Zahedi, "Variable step-size l0-norm nlms algorithm for sparse channel estimation," in *Proc. IEEE Asia Pacific Conference on Wireless and Mobile*, pp. 88–91, Aug. 2014.

[25] J. Evans, P. Xue, and B. Liu, "Analysis and implementation of variable step size adaptive algorithms," *IEEE Transactions on Signal Processing*, vol. 41, no. 8, pp. 2517–2535, Aug. 1993.

[26] J. Mendel, R. John, and F. Liu, "Interval type-2 fuzzy logic systems made simple," *IEEE Transactions on Fuzzy Systems*, vol. 14, no. 6, pp. 808–821, Dec. 2006.

[27] Q. Liang and J. Mendel, "Interval type-2 fuzzy logic systems: theory and design," *IEEE Transactions on Fuzzy Systems*, vol. 8, no. 5, pp. 535–550, Oct. 2000.

[28] J. M. Mendel, *Uncertain Rule-Based Fuzzy Logic Systems: Introduction and New Directions*. Prentice Hall PTR, 2001.

[29] J. Mendel, "Computing derivatives in interval type-2 fuzzy logic systems," *IEEE Transactions on Fuzzy Systems*, vol. 12, no. 1, pp. 84–98, Feb. 2004.

[30] E. P. de Aguiar, F. M. A. Nogueira, R. P. Amaral, D. F. Fabri, S. C. Rossignoli, J. G. Ferreira, M. M. B. R. Vellasco, R. Tanscheit, P. C. S. Vellasco, and M. V. Ribeiro, "Classification of events in switch machines using bayes, fuzzy logic system and neural network," in *Proc. 15th International Conference on Engineering Applications of Neural Networks*, pp. 81–91, Sep. 2014.

[31] D.-Y. Li, Y.-D. Song, and W.-C. Cai, "Neuro-adaptive fault-tolerant approach for active suspension control of high-speed trains," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 5, pp. 2446–2456, Oct. 2015.

[32] K. Guan, B. Ai, Z. Zhong, C. Lopez, L. Zhang, C. Briso-Rodriguez, A. Hrovat, B. Zhang, R. He, and T. Tang, "Measurements and analysis of large-scale fading characteristics in curved subway tunnels at 920 mhz, 2400 mhz, and 5705 mhz," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 5, pp. 2393–2405, Oct. 2015.

[33] M. Soler, J. Lopez, J. Mera Sanchez de Pedro, and J. Maroto, "Methodology for multiobjective optimization of the ac railway power supply system," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 5, pp. 2531–2542, Oct. 2015.

[34] H. Ardakani, C. Lucas, D. Siegel, S. Chang, P. Dersin, B. Bonnet, and J. Lee, "Phm for railway system x2014; a case study on the health assessment of the point machines," in *Proc. IEEE Conference on Prognostics and Health Management*, pp. 1–5, Jun. 2012.

[35] M. Vileiniskis, R. Remenyte-Prescott, D. Rama, and J. Andrews, "Fault diagnostics for railway point machines," in *Proc. 20th Advances in Risk and Reliability Technology Symposium*, pp. 1–5, May 2013.

[36] V. Atamuradov, F. Camci, S. Baskan, and M. Sevkli, "Failure diagnostics for railway point machines using expert systems," in *Proc. IEEE International Symposium on Diagnostics for Electric Machines, Power Electronics and Drives*, pp. 1–5, Aug. 2009.

[37] F. Márquez, F. Schmid, and J. Collado, "A reliability centered approach to remote condition monitoring. a railway points case study," *Reliability Engineering & System Safety*, vol. 80, no. 1, pp. 33–40, 2003.

[38] B. Costa, P. Angelov, and L. Guedes, "Real-time fault detection using recursive density estimation," *Journal of Control, Automation and Electrical Systems*, vol. 25, no. 4, pp. 428–437, Aug. 2014.

[39] B. Costa, P. Angelov, and L. Guedes, "Fully unsupervised fault detection and identification based on recursive density estimation and self-evolving cloud-based classifier," *Neurocomputing*, vol. 150, no. 4, pp. 289–303, Feb. 2015.

[40] E. P. de Aguiar, C. A. G. Marques, C. A. Duque, and M. V. Ribeiro, "Signal decomposition with reduced complexity for classification of isolated and multiple disturbances in electric signals," *IEEE Transactions on Power Delivery*, vol. 24, no. 4, pp. 2459–2460, Oct. 2009.

[41] J. M. Mendel, "Tutorial on higher-order statistics (spectra) in signal processing and system theory: theoretical results and some applications," *Proceedings of the IEEE*, vol. 79, no. 3, pp. 278–305, Mar. 1991.

[42] S. Theodoridis and K. Koutroumbas, *Pattern Recognition.* San Diego: Academic Press, 1999.

[43] S. Haykin, *Neural networks. A Comprehensive Foundation.*, 2nd ed. Prentice Hall, 1999.

[44] O. F. Eker, F. Camci, A. Guclu, H. Yilboga, M. Sevkli, and S. Baskan, "A simple state-based prognostic model for railway turnout systems," *IEEE Transactions on Industrial Electronics*, vol. 58, no. 5, pp. 1718–1726, May 2011.

[45] F. P. G. Marquez, R. W. Lewis, A. M. Tobias, and C. Roberts, "Life cycle costs for railway condition monitoring," *Transportation Research Part E: Logistics and Transportation Review*, vol. 44, no. 6, pp. 1175 – 1187, Nov. 2008.

[46] F. P. G. Marquez, P. Weston, and C. Roberts, "Failure analysis and diagnostics for railway trackside equipment," *Engineering Failure Analysis*, vol. 14, no. 8, pp. 1411 – 1426, Dec. 2007.

[47] M. V. Ribeiro, C. G. Marques, A. S. Cerqueira, C. A. Duque, and J. L. R. Pereira, "Detection of Disturbances in Voltage Signals for Power Quality Analysis Using HOS," *EURASIP Journal on Advances in Signal Processing*, vol. 2007, pp. 13–25, Jan. 2007.

[48] D. Wu, "Approaches for reducing the computational cost of interval type-2 fuzzy logic systems: Overview and comparisons," *IEEE Transactions on Fuzzy Systems*, vol. 21, no. 1, pp. 80–99, Feb 2013.

[49] D. Wu, "An overview of alternative type-reduction approaches for reducing the computational cost of interval type-2 fuzzy logic controllers," in *Proc. IEEE International Conference on Fuzzy Systems*, pp. 1–8, Jun. 2012.

[50] P. Sharma and P. Bajaj, "Accuracy comparison of vehicle classification system using interval type-2 fuzzy inference system," in *Proc. International Conference on Emerging Trends in Engineering and Technology*, pp. 85–90, Nov. 2010.

[51] D. Basu, S. Bhattacharyya, D. Sardar, A. Konar, D. Tibarewala, and A. Nagar, "A differential evolution based adaptive neural type-2 fuzzy inference system for classification of motor imagery eeg signals," in *Proc. IEEE International Conference on Fuzzy Systems*, pp. 1253–1260, Jul. 2014.

[52] R. Hosseini, S. Qanadli, S. Barman, M. Mazinani, T. Ellis, and J. Dehmeshki, "An automatic approach for learning and tuning gaussian interval type-2 fuzzy membership functions applied to lung cad classification system," *IEEE Transactions on Fuzzy Systems*, vol. 20, no. 2, pp. 224–234, Apr. 2012.

[53] M. J. Alhaddad, A. Mohammed, M. Kamel, and H. Hagras, "A genetic interval type-2 fuzzy logic-based approach for generating interpretable linguistic models for the brain p300 phenomena recorded via brain–computer interfaces," *Soft Computing*, vol. 19, no. 4, pp. 1019–1035, Apr. 2015.

[54] C.-F. Juang and P.-H. Wang, "An interval type-2 neural fuzzy classifier learned through soft margin minimization and its human posture classification application," *IEEE Transactions on Fuzzy Systems*, vol. 23, no. 5, pp. 1474–1487, Oct 2015.

[55] N. Karnik and J. Mendel, "Type-2 fuzzy logic systems: type-reduction," in *Proc. IEEE International Conference on Systems, Man, and Cybernetics*, pp. 2046–2051, Oct. 1998.

[56] C. Gafa and S. Coupland, "A new recursive type-reduction procedure for general type-2 fuzzy sets," in *Proc. IEEE Symposium on Advances in Type-2 Fuzzy Logic Systems*, pp. 44–49, Apr. 2011.

[57] J. Aisbett and J. Rickard, "Centroids of type-1 and type-2 fuzzy sets when membership functions have spikes," *IEEE Transactions on Fuzzy Systems*, vol. 22, no. 3, pp. 685–692, Jun. 2014.

[58] J. Figueroa Garcia, "An approximation method for type reduction of an interval type-2 fuzzy set based on α-cuts," in *Proc. Federated Conference on Computer Science and Information Systems*, pp. 49–54, Sept. 2012.

[59] N. Karnik and J. Mendel, "Introduction to type-2 fuzzy logic systems," in *Proc. IEEE International Conference on Fuzzy Systems*, pp. 915–920, May. 1998.

[60] D. Driankov, H. Hellendoorn, and M. Reinfrank, *An Introduction to Fuzzy Control*, 2nd ed. Springer-Verlag, 1996.

[61] H. Wu and J. Mendel, "Uncertainty bounds and their use in the design of interval type-2 fuzzy logic systems," *IEEE Transactions on Fuzzy Systems*, vol. 10, no. 5, pp. 622–639, Oct. 2002.

[62] J. Kacprzyk and W. Pedrycz, *Springer Handbook of Computational Intelligence*, 1st ed. Springer-Verlag Berlin Heidelberg, 2015.

[63] H. Hagras, "A hierarchical type-2 fuzzy logic control architecture for autonomous mobile robots," *IEEE Transactions on Fuzzy Systems*, vol. 12, no. 4, pp. 524–539, Aug. 2004.

[64] Q. Liang and J. Mendel, "Equalization of nonlinear time-varying channels using type-2 fuzzy adaptive filters," *IEEE Transactions on Fuzzy Systems*, vol. 8, no. 5, pp. 551–563, Oct. 2000.

[65] Q. Liang and J. Mendel, "Overcoming time-varying co-channel interference using type-2 fuzzy adaptive filters," *IEEE Transactions on Circuits and Systems*, vol. 47, no. 12, pp. 1419–1428, Dec. 2000.

[66] Q. Liang and J. Mendel, "MPEG VBR video traffic modeling and classification using fuzzy technique," *IEEE Transactions on Fuzzy Systems*, vol. 9, no. 1, pp. 183–193, 2001.

[67] Q. Liang, N. Karnik, and J. Mendel, "Connection admission control in atm networks using survey-based type-2 fuzzy logic systems," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 30, no. 3, pp. 329–339, Aug. 2000.

[68] P. Melin and O. Castillo, "A new method for adaptive control of non-linear plants using type-2 fuzzy logic and neural networks." *International Journal of General Systems*, vol. 33, no. 2/3, pp. 289 – 304, Apr. 2004.

[69] T. Ozen and J. M. Garibaldi, "Investigating adaptation in type-2 fuzzy logic systems applied to umbilical acid-base assessment," *European Symposium on Intelligent Technologies, Hybrid Systems and Their Implementation on Smart Adaptive Systems*, pp. 289–294, Jul. 2003.

[70] D. Wu and W. Tan, "A type-2 fuzzy logic controller for the liquid-level process," *IEEE International Conference on Fuzzy Systems*, vol. 2, pp. 953–958, Jul. 2004.

[71] H. Wu and J. Mendel, "Classifier designs for binary classifications of ground vehicles," *Unattended Ground Sensor Technologies and Applications V*, vol. 5090, pp. 122–133, Sep. 2003.

[72] J. Mendel, "On the importance of interval sets in type-2 fuzzy logic systems," in *Proc. International Fuzzy Systems Association World Congress and North American Fuzzy Information Processing Society International Conference*, vol. 3, pp. 1647–1652, Jul. 2001.

[73] L. Jie, L. Siwei, L. Qingyong, Z. Hanqing, and R. Shengwei, "Real-time rail head surface defect detection: A geometrical approach," in *Proc. IEEE International Symposium on Industrial Electronics*, pp. 769–774, Jul. 2009.

[74] V. Vijaykumar and S. Sangamithirai, "Rail defect detection using gabor filters with texture analysis," in *Proc. International Conference on Signal Processing, Communication and Networking*, pp. 1–6, Mar. 2015.

[75] H.-Y. Zhang, D. Feng, J.-C. Yao, and Y. ling Yu, "Guided wave propagation characteristics in the rail with a crack defect," in *Proc. Symposium on Piezoelectricity, Acoustic Waves and Device Applications*, pp. 1–4, Oct. 2013.

[76] H. Chang, C. Wilson, and J. Jackson, "Eliminating polymer flake defects using an oxygen free chemistry," in *Proc. IEEE International Symposium on Semiconductor Manufacturing Conference Proceedings*, pp. P91–P93, Oct. 1997.

[77] D. Stone, F. Carlson, and C. Bachhuber, "Effect of brake-system components on wheel spalling," in *Proc. IEEE Joint Railroad Conference*, pp. 177–183, Apr. 1999.

[78] Y. Matsuo, R. Takada, S. Iwasaki, and J. Katto, "Image super-resolution using registration of wavelet multi-scale components with affine transformation," in *Proc. IEEE International Symposium on Multimedia*, pp. 279–282, Dec. 2013.

[79] H. Lin, P. Du, W. Zhao, L. Zhang, and H. Sun, "Image registration based on corner detection and affine transformation," in *Proc. 3rd International Congress on Image and Signal Processing*, pp. 2184–2188, Oct. 2010.

[80] X. Li, Y. Xu, Q. Lv, and Y. Dou, "Affine-transformation parameters regression for face alignment," *IEEE Signal Processing Letters*, vol. 23, no. 1, pp. 55–59, Jan. 2016.

[81] R. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 3, no. 6, pp. 610–621, Nov. 1973.

[82] M. Cui, S. Prasad, M. Mahrooghy, J. Aanstoos, M. Lee, and L. Bruce, "Decision fusion of textural features derived from polarimetric data for levee assessment," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 5, no. 3, pp. 970–976, Jun. 2012.

[83] V. Thakare and N. Patil, "Classification of texture using gray level co-occurrence matrix and self-organizing map," in *Proc. International Conference on Electronic Systems, Signal Processing and Computing Technologies*, pp. 350–355, Jan. 2014.

[84] H. Nikoo, H. Talebi, and A. Mirzaei, "A supervised method for determining displacement of gray level co-occurrence matrix," in *Proc. 7th Iranian Machine Vision and Image Processing*, pp. 1–5, Nov. 2011.

[85] W. Kun and K. Songtao, "Identification method of waste based on gray level co-occurrence matrix and neural network," in *Proc. International Conference on Materials for Renewable Energy Environment*, pp. 929–931, May. 2011.

[86] Y. Hu, C. xia Zhao, and H. nan Wang, "Directional analysis of texture images using gray level co-occurrence matrix," in *Proc. Pacific-Asia Workshop on Computational Intelligence and Industrial Application*, pp. 277–281, Dec. 2008.

[87] J. Yang and J. Guo, "Image texture feature extraction method based on regional average binary gray level difference co-occurrence matrix," in *Proc. International Conference on Virtual Reality and Visualization*, pp. 239–242, Nov. 2011.

[88] R. Suganya and S. Rajaram, "Feature extraction and classification of ultrasound liver images using haralick texture-primitive features: Application of svm classifier," in *Proc. International Conference on Recent Trends in Information Technology*, pp. 596–602, Jul. 2013.

[89] G. Preethi and V. Sornagopal, "MRI image classification using GLCM texture features," in *Proc. International Conference on Green Computing Communication and Electrical Engineering*, pp. 1–6, Mar. 2014.

[90] V. Pullano, A. Vanelli-Coralli, and G. E. Corazza, "PSNR evaluation and alignment recovery for mobile satellite video broadcasting," in *Proc. Advanced Satellite Multimedia Systems Conference and 12th Signal Processing for Space Communications Workshop*, pp. 176–181, Sep. 2012.

[91] R. R. S. Tomar and K. Jain, "Lossless image compression using differential pulse code modulation and its application," in *Proc. Fifth International Conference on Communication Systems and Network Technologies*, pp. 543–545, Apr. 2015.

[92] P. Zhang and F. Li, "A new adaptive weighted mean filter for removing salt-and-pepper noise," *IEEE Signal Processing Letters*, vol. 21, no. 10, pp. 1280–1283, Oct 2014.

[93] K. K. V. Toh and N. A. M. Isa, "Noise adaptive fuzzy switching median filter for salt-and-pepper noise reduction," *IEEE Signal Processing Letters*, vol. 17, no. 3, pp. 281–284, Mar. 2010.

[94] S. Esakkirajan, T. Veerakumar, A. N. Subramanyam, and C. H. PremChand, "Removal of high density salt and pepper noise through modified decision based unsymmetric trimmed median filter," *IEEE Signal Processing Letters*, vol. 18, no. 5, pp. 287–290, May 2011.

Appendix A – Deduction of the Gradient Vector for the IST2-FLS

The following notations have been adopted for reasons of simplicity:

$$e^{(q)} = f_{s2}\left(\mathbf{x}^{(q)}\right) - y^{(q)}, \tag{A.1}$$

$$e_{\theta lL}^{(q)} = \theta_l\left(q\right) - Y_L\left(\mathbf{x}^{(q)}\right), \tag{A.2}$$

$$e_{\theta lR}^{(q)} = \theta_l\left(q\right) - Y_R\left(\mathbf{x}^{(q)}\right), \tag{A.3}$$

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

$$a_{2F_k^l}\left(q\right) = \frac{x_k^{(q)} - m_{2F_k^l}\left(q\right)}{\sigma_{F_k^l}^2\left(q\right)}, \tag{A.5}$$

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

and

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

A.1 DEDUCTION FOR $m_{1F_k^l}\left(q\right)$

If $x_k^{(q)} < m_{1F_k^l}(q)$:

$$\frac{\partial J\left(\mathbf{w}^{(q)}\right)}{\partial m_{1F_k^l}\left(q\right)} = \frac{e^{(q)}}{2}\left(e_{\theta lL}^{(q)}\frac{\overline{\vartheta}_l\left(\mathbf{x}^{(q)}\right)}{\sum_{l=1}^{M}\vartheta_{lL}\left(\mathbf{x}^{(q)}\right)}\right. \tag{A.8}$$

$$\left. + e_{\theta lR}^{(q)}\frac{\overline{\vartheta}_l\left(\mathbf{x}^{(q)}\right)}{\sum_{l=1}^{M}\vartheta_{lR}\left(\mathbf{x}^{(q)}\right)}\right) a_{1F_k^l}\left(q\right).$$

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

$$\frac{\partial J\left(\mathbf{w}^{(q)}\right)}{\partial m_{1F_k^l}\left(q\right)} = 0. \tag{A.9}$$

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

$$\frac{\partial J\left(\mathbf{w}^{(q)}\right)}{\partial m_{1F_k^l}(q)} = \frac{e^{(q)}}{2}\left(e_{\theta lL}^{(q)}\frac{\underline{\vartheta}_l\left(\mathbf{x}^{(q)}\right)}{\sum_{l=1}^{M}\vartheta_{lL}\left(\mathbf{x}^{(q)}\right)}\right. \tag{A.10}$$
$$\left. + e_{\theta lR}^{(q)}\frac{\underline{\vartheta}_l\left(\mathbf{x}^{(q)}\right)}{\sum_{l=1}^{M}\vartheta_{lR}\left(\mathbf{x}^{(q)}\right)}\right) a_{1F_k^l}(q).$$

If $x_k^{(q)} > m_{2F_k^l}(q)$:

$$\frac{\partial J\left(\mathbf{w}^{(q)}\right)}{\partial m_{1F_k^l}(q)} = \frac{e^{(q)}}{2}\left(e_{\theta lL}^{(q)}\frac{\underline{\vartheta}_l\left(\mathbf{x}^{(q)}\right)}{\sum_{l=1}^{M}\vartheta_{lL}\left(\mathbf{x}^{(q)}\right)}\right. \tag{A.11}$$
$$\left. + e_{\theta lR}^{(q)}\frac{\underline{\vartheta}_l\left(\mathbf{x}^{(q)}\right)}{\sum_{l=1}^{M}\vartheta_{lR}\left(\mathbf{x}^{(q)}\right)}\right) a_{1F_k^l}(q).$$

A.2 DEDUCTION FOR $m_{2F_k^l}(q)$

If $x_k^{(q)} < m_{1F_k^l}(q)$:

$$\frac{\partial J\left(\mathbf{w}^{(q)}\right)}{\partial m_{2F_k^l}(q)} = \frac{e^{(q)}}{2}\left(e_{\theta lL}^{(q)}\frac{\underline{\vartheta}_l\left(\mathbf{x}^{(q)}\right)}{\sum_{l=1}^{M}\vartheta_{lL}\left(\mathbf{x}^{(q)}\right)}\right. \tag{A.12}$$
$$\left. + e_{\theta lR}^{(q)}\frac{\underline{\vartheta}_l\left(\mathbf{x}^{(q)}\right)}{\sum_{l=1}^{M}\vartheta_{lR}\left(\mathbf{x}^{(q)}\right)}\right) a_{2F_k^l}(q).$$

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

$$\frac{\partial J\left(\mathbf{w}^{(q)}\right)}{\partial m_{2F_k^l}(q)} = \frac{e^{(q)}}{2}\left(e_{\theta lL}^{(q)}\frac{\underline{\vartheta}_l\left(\mathbf{x}^{(q)}\right)}{\sum_{l=1}^{M}\vartheta_{lL}\left(\mathbf{x}^{(q)}\right)}\right. \tag{A.13}$$
$$\left. + e_{\theta lR}^{(q)}\frac{\underline{\vartheta}_l\left(\mathbf{x}^{(q)}\right)}{\sum_{l=1}^{M}\vartheta_{lR}\left(\mathbf{x}^{(q)}\right)}\right) a_{2F_k^l}(q).$$

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

$$\frac{\partial J\left(\mathbf{w}^{(q)}\right)}{\partial m_{2F_k^l}(q)} = 0. \tag{A.14}$$

If $x_k^{(q)} > m_{2F_k^l}(q)$:

$$\frac{\partial J\left(\mathbf{w}^{(q)}\right)}{\partial m_{2F_k^l}(q)} = \frac{e^{(q)}}{2}\left(e_{\theta lL}^{(q)}\frac{\overline{\vartheta}_l\left(\mathbf{x}^{(q)}\right)}{\sum_{l=1}^{M}\vartheta_{lL}\left(\mathbf{x}^{(q)}\right)}\right. \tag{A.15}$$

$$\left. + e_{\theta lR}^{(q)}\frac{\overline{\vartheta}_l\left(\mathbf{x}^{(q)}\right)}{\sum_{l=1}^{M}\vartheta_{lR}\left(\mathbf{x}^{(q)}\right)}\right) a_{2F_k^l}(q)\,.$$

A.3 DEDUCTION FOR $\sigma_{F_k^l}(q)$

If $x_k^{(q)} < m_{1F_k^l}(q)$:

$$\frac{\partial J\left(\mathbf{w}^{(q)}\right)}{\partial \sigma_{F_k^l}(q)} =$$

$$\frac{e^{(q)}}{2}\left(\frac{\overline{\vartheta}_l\left(\mathbf{x}^{(q)}\right) b_{1F_k^l}(q) + \underline{\vartheta}_l\left(\mathbf{x}^{(q)}\right) b_{2F_k^l}(q)}{\sum_{l=1}^{M}\vartheta_{lL}\left(\mathbf{x}^{(q)}\right)}\, e_{\theta lL}^{(q)}\right. \tag{A.16}$$

$$\left. + e_{\theta lR}^{(q)}\frac{\underline{\vartheta}_l\left(\mathbf{x}^{(q)}\right) b_{2F_k^l}(q) + \overline{\vartheta}_l\left(\mathbf{x}^{(q)}\right) b_{1F_k^l}(q)}{\sum_{l=1}^{M}\vartheta_{lR}\left(\mathbf{x}^{(q)}\right)}\right).$$

If $m_{1F_k^l}(q) \le x_k^{(q)} \le m_{2F_k^l}(q)$ and $x_k^{(q)} \le \frac{m_{1F_k^l}(q)+m_{2F_k^l}(q)}{2}$:

$$\frac{\partial J\left(\mathbf{w}^{(q)}\right)}{\partial \sigma_{F_k^l}(q)} = \frac{e^{(q)}}{2}\left(e_{\theta lL}^{(q)}\frac{\underline{\vartheta}_l\left(\mathbf{x}^{(q)}\right)}{\sum_{l=1}^{M}\vartheta_{lL}\left(\mathbf{x}^{(q)}\right)}\right. \tag{A.17}$$

$$\left. + e_{\theta lR}^{(q)}\frac{\underline{\vartheta}_l\left(\mathbf{x}^{(q)}\right)}{\sum_{l=1}^{M}\vartheta_{lR}\left(\mathbf{x}^{(q)}\right)}\right) b_{2F_k^l}(q).$$

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

$$\frac{\partial J\left(\mathbf{w}^{(q)}\right)}{\partial \sigma_{F_k^l}(q)} = \frac{e^{(q)}}{2}\left(e_{\theta lL}^{(q)}\frac{\underline{\vartheta}_l\left(\mathbf{x}^{(q)}\right)}{\sum_{l=1}^{M}\vartheta_{lL}\left(\mathbf{x}^{(q)}\right)}\right. \tag{A.18}$$

$$\left. + e_{\theta lR}^{(q)}\frac{\underline{\vartheta}_l\left(\mathbf{x}^{(q)}\right)}{\sum_{l=1}^{M}\vartheta_{lR}\left(\mathbf{x}^{(q)}\right)}\right) b_{1F_k^l}(q)$$

If $x_k^{(q)} > m_{2F_k^l}(q)$:

$$\frac{\partial J\left(\mathbf{w}^{(q)}\right)}{\partial \sigma_{F_k^l}(q)} =$$

$$\frac{e^{(q)}}{2}\left(\frac{\overline{\vartheta}_l\left(\mathbf{x}^{(q)}\right) b_{2F_k^l}(q) + \underline{\vartheta}_l\left(\mathbf{x}^{(q)}\right) b_{1F_k^l}(q)}{\sum_{l=1}^{M} \underline{\vartheta}_{lL}\left(\mathbf{x}^{(q)}\right)} e_{\theta l L}^{(q)}\right. \tag{A.19}$$

$$\left. + e_{\theta l R}^{(q)} \frac{\underline{\vartheta}_l\left(\mathbf{x}^{(q)}\right) b_{1F_k^l}(q) + \overline{\vartheta}_l\left(\mathbf{x}^{(q)}\right) b_{2F_k^l}(q)}{\sum_{l=1}^{M} \underline{\vartheta}_{lR}\left(\mathbf{x}^{(q)}\right)}\right).$$

A.4 DEDUCTION FOR $\theta_l(q)$

$$\frac{\partial J\left(\mathbf{x}^{(q)}\right)}{\partial \theta_l(q)} = \frac{e^{(q)}}{2}\left(\frac{\overline{\vartheta}_l\left(\mathbf{x}^{(q)}\right) + \underline{\vartheta}_l\left(\mathbf{x}^{(q)}\right)}{\sum_{l=1}^{M} \vartheta_{lL}\left(\mathbf{x}^{(q)}\right)}\right. \tag{A.20}$$

$$\left. + \frac{\underline{\vartheta}_l\left(\mathbf{x}^{(q)}\right) + \overline{\vartheta}_l\left(\mathbf{x}^{(q)}\right)}{\sum_{l=1}^{M} \vartheta_{lR}\left(\mathbf{x}^{(q)}\right)}\right).$$

where $\sum_{l=1}^{M} \vartheta_{lL}\left(\mathbf{x}^{(q)}\right)$ and $\sum_{l=1}^{M} \vartheta_{lR}\left(\mathbf{x}^{(q)}\right)$ are respectively:

$$\sum_{l=1}^{M} \vartheta_{lL}\left(\mathbf{x}^{(q)}\right) = \sum_{l=1}^{le} \overline{\vartheta}_l\left(\mathbf{x}^{(q)}\right) + \sum_{l=le}^{M} \underline{\vartheta}_l\left(\mathbf{x}^{(q)}\right) \tag{A.21}$$

and

$$\sum_{l=1}^{M} \vartheta_{lR}\left(\mathbf{x}^{(q)}\right) = \sum_{l=1}^{ri} \underline{\vartheta}_l\left(\mathbf{x}^{(q)}\right) + \sum_{l=ri}^{M} \overline{\vartheta}_l\left(\mathbf{x}^{(q)}\right) \tag{A.22}$$

Appendix B – Deduction of the Gradient Vector for the ULST2-FLS

The following notations have been adopted for reasons of simplicity:

$$e_{\theta lLow}^{(q)} = \theta_l(q) - Y_{Low}\left(\mathbf{x}^{(q)}\right) \tag{B.1}$$

and

$$e_{\theta lUp}^{(q)} = \theta_l(q) - Y_{Up}\left(\mathbf{x}^{(q)}\right). \tag{B.2}$$

B.1 DEDUCTION FOR $m_{1F_k^l}(q)$

If $x_k^{(q)} < m_{1F_k^l}(q)$:

$$\frac{\partial J\left(\mathbf{w}^{(q)}\right)}{\partial m_{1F_k^l}(q)} = \frac{e^{(q)}}{2}\left(e_{\theta lUp}^{(q)} \frac{\overline{\vartheta}_l\left(\mathbf{x}^{(q)}\right)}{\sum_{l=1}^{M}\overline{\vartheta}_l\left(\mathbf{x}^{(q)}\right)} a_{1F_k^l}(q)\right) \tag{B.3}$$

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

$$\frac{\partial J\left(\mathbf{w}^{(q)}\right)}{\partial m_{1F_k^l}(q)} = 0. \tag{B.4}$$

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

$$\frac{\partial J\left(\mathbf{w}^{(q)}\right)}{\partial m_{1F_k^l}(q)} = \frac{e^{(q)}}{2}\left(e_{\theta lLow}^{(q)} \frac{\underline{\vartheta}_l\left(\mathbf{x}^{(q)}\right)}{\sum_{l=1}^{M}\underline{\vartheta}_l\left(\mathbf{x}^{(q)}\right)} a_{1F_k^l}(q)\right) \tag{B.5}$$

If $x_k^{(q)} > m_{2F_k^l}(q)$:

$$\frac{\partial J\left(\mathbf{w}^{(q)}\right)}{\partial m_{1F_k^l}(q)} = \frac{e^{(q)}}{2}\left(e_{\theta lLow}^{(q)} \frac{\underline{\vartheta}_l\left(\mathbf{x}^{(q)}\right)}{\sum_{l=1}^{M}\underline{\vartheta}_l\left(\mathbf{x}^{(q)}\right)} a_{1F_k^l}(q)\right) \tag{B.6}$$

B.2 DEDUCTION FOR $m_{2F_k^l}(q)$

If $x_k^{(q)} < m_{1F_k^l}(q)$:

$$\frac{\partial J\left(\mathbf{w}^{(q)}\right)}{\partial m_{2F_k^l}(q)} = \frac{e^{(q)}}{2}\left(e_{\theta lLow}^{(q)} \frac{\underline{\vartheta}_l\left(\mathbf{x}^{(q)}\right)}{\sum_{l=1}^{M}\underline{\vartheta}_l\left(\mathbf{x}^{(q)}\right)} a_{2F_k^l}(q)\right) \tag{B.7}$$

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

$$\frac{\partial J\left(\mathbf{w}^{(q)}\right)}{\partial m_{2F_k^l}(q)} = \frac{e^{(q)}}{2}\left(e_{\theta lLow}^{(q)} \frac{\underline{\vartheta}_l\left(\mathbf{x}^{(q)}\right)}{\sum_{l=1}^{M}\underline{\vartheta}_l\left(\mathbf{x}^{(q)}\right)} a_{2F_k^l}(q)\right) \tag{B.8}$$

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

$$\frac{\partial J\left(\mathbf{w}^{(q)}\right)}{\partial m_{2F_k^l}(q)} = 0. \tag{B.9}$$

If $x_k^{(q)} > m_{2F_k^l}(q)$:

$$\frac{\partial J\left(\mathbf{w}^{(q)}\right)}{\partial m_{2F_k^l}(q)} = \frac{e^{(q)}}{2}\left(e_{\theta lUp}^{(q)}\frac{\overline{\vartheta}_l\left(\mathbf{x}^{(q)}\right)}{\sum_{l=1}^{M}\overline{\vartheta}_l\left(\mathbf{x}^{(q)}\right)}a_{2F_k^l}(q)\right) \tag{B.10}$$

B.3 DEDUCTION FOR $\sigma_{F_k^l}(q)$

If $x_k^{(q)} < m_{1F_k^l}(q)$:

$$\begin{aligned}\frac{\partial J\left(\mathbf{w}^{(q)}\right)}{\partial \sigma_{F_k^l}(q)} = \frac{e^{(q)}}{2}\Bigg(&e_{\theta lUp}^{(q)}\frac{\overline{\vartheta}_l\left(\mathbf{x}^{(q)}\right)}{\sum_{l=1}^{M}\overline{\vartheta}_l\left(\mathbf{x}^{(q)}\right)}\,b_{1F_k^l}(q)\\ &+e_{\theta lLow}^{(q)}\frac{\underline{\vartheta}_l\left(\mathbf{x}^{(q)}\right)}{\sum_{l=1}^{M}\underline{\vartheta}_l\left(\mathbf{x}^{(q)}\right)}b_{2F_k^l}(q)\Bigg)\end{aligned} \tag{B.11}$$

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

$$\frac{\partial J\left(\mathbf{w}^{(q)}\right)}{\partial \sigma_{F_k^l}(q)} = \frac{e^{(q)}}{2}\left(e_{\theta lLow}^{(q)}\frac{\underline{\vartheta}_l\left(\mathbf{x}^{(q)}\right)}{\sum_{l=1}^{M}\underline{\vartheta}_l\left(\mathbf{x}^{(q)}\right)}b_{2F_k^l}(q)\right) \tag{B.12}$$

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

$$\frac{\partial J\left(\mathbf{w}^{(q)}\right)}{\partial \sigma_{F_k^l}(q)} = \frac{e^{(q)}}{2}\left(e_{\theta lLow}^{(q)}\frac{\underline{\vartheta}_l\left(\mathbf{x}^{(q)}\right)}{\sum_{l=1}^{M}\underline{\vartheta}_l\left(\mathbf{x}^{(q)}\right)}b_{1F_k^l}(q)\right) \tag{B.13}$$

If $x_k^{(q)} > m_{2F_k^l}(q)$:

$$\begin{aligned}\frac{\partial J\left(\mathbf{w}^{(q)}\right)}{\partial \sigma_{F_k^l}(q)} = \frac{e^{(q)}}{2}\Bigg(&e_{\theta lUp}^{(q)}\frac{\overline{\vartheta}_l\left(\mathbf{x}^{(q)}\right)}{\sum_{l=1}^{M}\overline{\vartheta}_l\left(\mathbf{x}^{(q)}\right)}\,b_{2F_k^l}(q)\\ &+e_{\theta lLow}^{(q)}\frac{\underline{\vartheta}_l\left(\mathbf{x}^{(q)}\right)}{\sum_{l=1}^{M}\underline{\vartheta}_l\left(\mathbf{x}^{(q)}\right)}b_{1F_k^l}(q)\Bigg)\end{aligned} \tag{B.14}$$

B.4 DEDUCTION FOR $\theta_l(q)$

$$\frac{\partial J\left(\mathbf{x}^{(q)}\right)}{\partial \theta_l(q)} = \frac{e^{(q)}}{2}\left(\frac{\overline{\vartheta}_l\left(\mathbf{x}^{(q)}\right)}{\sum_{l=1}^{M}\overline{\vartheta}_l\left(\mathbf{x}^{(q)}\right)} + \frac{\underline{\vartheta}_l\left(\mathbf{x}^{(q)}\right)}{\sum_{l=1}^{M}\underline{\vartheta}_l\left(\mathbf{x}^{(q)}\right)}\right). \tag{B.15}$$

Appendix C – Publications

This Appendix presents the list of publications. Items marked with * refer to contributions related to the thesis.

The list of papers published in journals during the doctoral period are as follows:

* E. P. de Aguiar, F. M. A. Nogueira, R. P. Amaral, D. F. Fabri, S. C. Rossignoli, J. G. Ferreira, M. M. B. R. Vellasco, R. Tanscheit, P. C. S. Vellasco, and 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, no.5, pp. 1175−1189, Jul. 2016.

- R. B. de Santis, B. Milanez, E. B. Pereira De Castro , L. Goliatt and E. P. de Aguiar, "Fuzzy AHP for Supplier Selection in a Railway Operator", *Production Journal*, submittted on Jul. 2016.

* E. P. de Aguiar, R. P. F. Amaral, M. M. B. R. Vellasco, and M. V. Ribeiro, "An Enhanced Singleton Type-2 Fuzzy Logic System For Fault Classification in a Switch Machine", *Information Sciences*, submittted on Jul. 2016.

* E. P. de Aguiar, F. M. de A. Nogueira, M. M. B. R. Vellasco, and M. V. Ribeiro "Set-Membership Type-1 Fuzzy Logic System Applied to Fault Classification in a Switch Machine", *IEEE Transactions on Intelligent Transportation Systems*, Accepeted for publication - submittted on Dec. 15 2015.

* E. P. de Aguiar, F. M. de A. Nogueira, M. M. B. R. Vellasco, and M. V. Ribeiro, "Set-Membership Type-2 Fuzzy Logic System Applied to Classification of Rail Head Defects", *IEEE Transactions on Fuzzy Systems*, to be submittted.

The list of conference papers published during the doctoral period are as follows:

* E. P. de Aguiar, L. P. V. da Silva, A. F. Moreira, L. G. da Fonseca, F. M. de A. Nogueira, M. M. B. R. Vellasco, and M. V. Ribeiro, "Type-1 Fuzzy Logic System Applied to Classification of Rail Head Defects", *Proc. Brazilian Conference on Fuzzy Systems*, pp. 1−12, Nov. 2016.

- R. B. de Santis, E. P. de Aguiar and L. Goliatt, " Artificial Neural Networks for Material Classification in Purchase Portfolio Management", *Proc. Ibero-Latin American Congress on Computational Methods in Engineering*, pp. 1−11, Nov. 2016.

- V. B. C. Almeida, K. A. P. Lagares, E. P. de Aguiar, M. L. Lagares Junior, "A technique for classify short-circuiting GMA welding using Neural Network", in *ABCM International Congress of Mechanical Engineering*, pp. 1−7, Dec. 2015

- M. L. G. Salmento, E. P. de Aguiar and M. V. Ribeiro, “Reduced rank adaptive filter with variable step size for impulsive uwb plc systems”, in *Proc. Brazilian Telecommunications Symposium*, pp. 1−5, Sep. 2015.

- D. Engelender, M. L. Filomeno, F. A. Santos, E. P. de Aguiar, T. R. Oliveira and M. V. Ribeiro, “ Análise do Uso de Bloqueadores em Sistemas Power Line Communication”, in *Proc. Brazilian Telecommunications Symposium*, pp. 1−5, Sep. 2015.

* E. P. de 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, and M. V. Ribeiro, “Classification of events in switch machines using bayes, fuzzy logic system and neural network”, in *Proc. International Conference on Engineering Applications of Neural Networks*, pp. 81−91, Sep. 2014.