UNIVERSIDADE FEDERAL DE JUIZ DE FORA
FACULDADE DE ENGENHARIA
PROGRAMA DE PÓS GRADUAÇÃO EM MODELAGEM COMPUTACIONAL

Lara Dutra Fonseca

Swarm intelligence algorithm combined with type-2 fuzzy logic system for the classification of trends in hot boxes and hot wheels

Juiz de Fora
2024

Lara Dutra Fonseca

Swarm intelligence algorithm combined with type-2 fuzzy logic system for the classification of trends in hot boxes and hot wheels

Dissertação apresentada ao Programa de Pós Graduação em Modelagem Computacional da Universidade Federal de Juiz de Fora, como requisito parcial para a obtenção do título de Mestre em Modelagem Computacional.

Orientador: Prof. Dr. Eduardo Pestana de Aguiar

Juiz de Fora

2024

Ficha catalográfica elaborada através do programa de geração
automática da Biblioteca Universitária da UFJF,
com os dados fornecidos pelo(a) autor(a)

Fonseca, Lara Dutra.
Swarm intelligence algorithm combined with type-2 fuzzy logic system for the classification of trends in hot boxes and hot wheels / Lara Dutra Fonseca. -- 2024.
45 f.

Orientador: Eduardo Pestana de Aguiar
Dissertação (mestrado acadêmico) - Universidade Federal de Juiz de Fora, Faculdade de Engenharia. Programa de Pós-Graduação em Modelagem Computacional, 2024.

1. Type-2 Fuzzy Logic Systems. 2. Meta-heuristic Optimization. 3. Swarm Intelligence. 4. Classification. I. Aguiar, Eduardo Pestana de, orient. II. Título.

Lara Dutra Fonseca

Swarm Intelligence Algorithm combined with Type-2 Fuzzy Logic System for the classification of trends in Hot Boxes and Hot Wheels

Dissertação apresentada ao Programa de Pós-Graduação em Modelagem Computacional da Universidade Federal de Juiz de Fora como requisito parcial à obtenção do título de Mestre em Modelagem Computacional. Área de concentração: Modelagem Computacional.

Aprovada em 08 de abril de 2024.

BANCA EXAMINADORA

Prof. Dr. Eduardo Pestana de Aguiar - Orientador
Universidade Federal de Juiz de Fora

Prof. Dr. Leonardo Goliatt da Fonseca
Universidade Federal de Juiz de Fora

Prof. Dr. Fernando José Von Zuben
Universidade Estadual de Campinas

Juiz de Fora, 09/04/2024.

Documento assinado eletronicamente por **Fernando José Von Zuben**, **Usuário Externo**, em 14/04/2024, às 00:37, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Leonardo Goliatt da Fonseca**, **Professor(a)**, em 03/05/2024, às 13:45, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Eduardo Pestana de Aguiar**, **Professor(a)**, em 20/05/2024, às 10:48, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

A autenticidade deste documento pode ser conferida no Portal do SEI-Ufjf (www2.ufjf.br/SEI) através do ícone Conferência de Documentos, informando o código verificador **1776041** e o código CRC **CBFF8512**.

ACKLOWLEDGMENT

First and foremost, I thank God for granting me the strength, wisdom, and perseverance needed to complete this journey.

I am also deeply grateful to my parents, Sélbio e Cristina, for their constant love and encouragement.

My deepest gratitude goes to my advisor, Eduardo Pestana de Aguiar, for his invaluable support and guidance.

I acknowledge the Federal University of Juiz de Fora for essential support during this work and all the professors of the department of Computational Modeling.

I also thank the financial support of CAPES.

ABSTRACT

Hot box and hot wheel problems on trains are significant threats in any rail operation because they increase fatigue and wear processes, resulting in failures, costly train stoppages and even derailments. For this reason, there are many hot box and hot wheel detectors to monitor those components distributed along the railway, and when they detect overheated components, a warning is given to alert to the impending failure. However, some situations such as solar incidence, misalignment, or sensor defect can lead to improper warnings, that is, warnings that actually do not exhibit the problem. Therefore, this work develops a new model for the classification of proper and improper warnings in hot box and hot wheel problems, by the analysis of the temperatures measured in the wheels and bearings of the train when the warning was given. To this end, the discussion focuses on the use of a meta-heuristic optimization algorithm into the training phase of an upper and lower singleton type-2 fuzzy logic system, in order to increase the convergence speed and the accuracy of this procedure. Thus, the performance of the new model will be based on the real data set composed of train wheel and bearing temperatures, provided by a Brazilian railway transportation company, covering several situations of proper and improper warnings. Additionally, this work presents performance analyses based on well-known data sets provided by Knowledge Extraction based on Evolutionary Learning Repository, aiming to evaluate the proposal. For comparison purposes, the results of the proposed classifier will be compared to well-know deep-learning classifiers as well as the original fuzzy model it is based on. The performance analysis of the new model is discussed in terms of classification ratio and convergence speed. The reported results show that the proposal achieved higher classification ratio and proved to be an appropriate option for the presented problem.

Key-words: Type-2 Fuzzy Logic Systems, Meta-heuristic Optimization, Swarm Intelligence, Classification.

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ANTF	Associação Nacional dos Transportadores Ferroviários
CNN	Convolutional Neural Network
FLS	Fuzzy Logic System
FRA	United States Federal Railway Administration
GWO	Grey Wolf Optimizer
HB	Hot box
HBD	Hot box detector
HW	Hot wheel
HWD	Hot wheel detector
LSTM	Long Short-term Memory
MBF	Membership function
MSE	Mean square error
SI	Swarm Intelligence
SM	Set-membership
T1 FLS	Type-1 Fuzzy Logic System
T2 FLS	Type-2 Fuzzy Logic System
ULST2-FLS	Upper and Lower Singleton Type-2 Fuzzy Logic System

SUMMARY

1 INTRODUCTION

Among different modes of transportation, the rail transportation offers excellent energy conservation, environmental protection and reduced travel time (CHONG; SHIN, 2010). Inevitably, it plays a major role in the transportation of goods and passengers, being an effective solution for connecting urban centers and a low-cost option for industry transactions. Due to this circumstances, many train components are constantly submitted to stressing cycles with heavy loads, increasing the probability of failures and reducing the lifespan of components such as bearings and wheels.

According to (TARAWNEH et al., 2019) and (TARAWNEH et al., 2009), the increased bearing defects might cause catastrophic failure that can lead to costly train stoppages and even derailments. In addition, (HAIDARI; TEHRANI, 2015) stated the importance of the thermal effect on wheels fatigue and the impact of the fatigue on wheel failure that may cause derailments.

Hot box and hot wheel are indications of these defects, that is why the overheating of the bearings and wheels can warn of some impending failures. Therefore, monitoring these components by hot box and hot wheel detection systems is essential to prevent the aforementioned accidents.

For those reasons, the importance of maintenance and maintenance management has grown in recent years. So, the railway industries face increasing pressure from customers and owners to improve safety, capacity, and reliability, while controlling expenses and tightening the budget (PALO et al., 2014).

In the case of hot bearings or hot wheels detection, the sensor can erroneously detect overheating of these components due to some situations such as sensor defects, misalignment, or solar incidence. This incorrect detection leads to unnecessary maintenance that greatly increases the expenses of the railway industries.

In light of this, the use computational intelligence techniques to distinguish the real overheating situations from the false ones is a valid way to reduce the unnecessary maintenance. Thus, the cost and time spent on maintenance is reduced and productivity is increased without reducing safety of rail operations.

One computational intelligence method that has been extensively used in classification problems due to its treatment of uncertainty is the Type-1 fuzzy logic system (FLS) able to generate interesting results in various sectors of the railway industry, for example (AGUIAR et al., 2014; AGUIAR et al., 2015; AGUIAR et al., 2017; AGUIAR et al., 2016).

Even though type-1 FLS is capable of handling uncertainty associated with classification problems and offers improved performance, it is unfortunately unable to adequately model and minimize the effect of some kinds of uncertainties, as it is discussed in (Mendel,

2007). Those uncertainties arise from factors like randomness, imprecision, or other elements such as ambiguity, noise, distractions, and more. Such uncertainties can stem from the nature of the problem itself or from challenges in translating specialized knowledge into the model.

Therefore, a suitable option to better overcome this limitation is the type-2 FLS, which provides a higher level of imprecision modeling (JOHN; COUPLAND, 2006). The extra dimension and parameters in type-2 FLS may offer more design freedom and flexibility than type-1 FLS (ALMARAASHI et al., 2016).

However, the additional freedom provided by type-2 FLS which increases the ability for modelling uncertainty, also entails a significant increase in complexity of the fuzzy systems and their operations (WAGNER; HAGRAS, 2010). As stated in (LIANG; MENDEL, 2000b), a downside of the type-2 FLS is that they are computationally intensive because of their type-reducer, and the same is associated with their training phase.

Therefore, the limited hardware resources availability and the real-time constraint can act as restrictions for the proper functioning of type-2 FLS. Consequently, for a limited number of epochs, it is hard to achieve a training with reduced computational complexity, high accuracy, and high convergence speed with type-2 FLS.

As a result, an important asset is a training phase with high accuracy and convergence speed which may result in less utilization of the hardware resource and, accordingly, may enable almost real-time training of the classification technique when new patterns need to be covered.

In order to deal with the limitation of type-1 FLS and the computation complexity of type-2 FLS, a previous work proposed the use of the algorithm Adam, acronym for adaptive moment estimation, presented by (KINGMA; BA, 2014), for the training phase of an upper and lower singleton type-2 fuzzy logic system (ULST2-FLS), introduced by (AGUIAR et al., 2018). The combination of the stochastic optimization strategy provided by Adam with ULST2-FLS created the algorithm Adam ULST2-FLS, which was presented in (FONSECA; AGUIAR, 2022).

The concept of set-membership (SM) (AGUIAR et al., 2017; CLARKE; LAMARE, 2011; SALMENTO et al., 2017; SALMENTO; AGUIAR; RIBEIRO, 2015; WANG; DE-LAMARE, 2013; DINIZ, 2002; ABADI; ESKANDARI, 2013) derived from adaptive filter theory (HAYKIN, 1996),and the use of the mean square error (MSE) combined with SM (AGUIAR et al., 2020) were also proposed in (FONSECA; AGUIAR, 2022). Those models achieved great results with fast convergence speed during training phase and fewer use of hardware resources.

As a sequence of (FONSECA; AGUIAR, 2022), this work proposes the utilization of a meta-heuristic method for the training phase of ULST2-FLS. There are plenty of

reasons for working with meta-heuristic optimization techniques. For example, they are easy to understand, because they are generally inspired by simple concepts. And they are also very flexible and can be applied to different problems without any special changes in the structure of the algorithm. Another advantage is their hybridization property, that is they can combined with other optimization techniques to create hybrid algorithms that leverage the strengths of multiple approaches for improved performance. Moreover, the meta-heuristic methods are able to avoid stagnation in local solutions.

Therefore the chosen method was the algorithm Grey Wolf Optimizer, presented by (MIRJALILI; MIRJALILI; LEWIS, 2014). GWO is a Swarm Intelligence (SI) optimization algorithm inspired by grey wolves, which mimicks the leadership hierarchy and hunting mechanism of these animals in nature. According to (MIRJALILI; MIRJALILI; LEWIS, 2014), this algorithm is an efficient method that provides competitive results when compared to other methods such as particle swarm optimization (PSO), gravitational search algorithm (GSA), differential evolution (DE), and evolutionary programming (ES). It also showed high local optima avoidance and it is simple to implement, computationally efficient, with fewer parameters to adjust. These characteristics make it an attractive method to be applied to the training phase of a type-2 FLS for increasing its accuracy and convergence speed. Furthermore, GWO presents great results when combined with other algorithms (DAVIRAN; GHEZELBASH; MAGHSOUDI, 2023) as well as with fuzzy controllers (BOJAN-DRAGOS et al., 2021).

For the purpose of testing the proposal performance, it is applied to classify data sets of binary classification problem provided by Knowledge Extraction based on Evolutionary Learning (KEEL) Repository (ALCALá-FDEZ et al., 2010).

Additionally, the proposed method is applied to a data set consisting of temperature measurements on train wheels and bearings collected by hot box and hot wheel detection systems in order to classify the authenticity of hot box and hot wheels detection and identify false indications to reduce unneeded maintenance. This data set was provided by a Brazilian company of the railway sector, named MRS Logística S.A. (https://www.mrs.com.br/).

Therefore, the main contributions of this work can be summarized as follows:

- The introduction of GWO ULST2-FLS, a Type-2 Fuzzy Logic System trained by the SI algorithm optimization GWO (MIRJALILI; MIRJALILI; LEWIS, 2014) in order to obtain a training phase with increased accuracy and convergence speed, requiring less hardware resource utilization.

- Performance analysis of GWO ULST2-FLS in terms of convergence speed, accuracy, MSE, Cohen's kappa and F-score by using data sets of binary classification problem provided by Knowledge Extraction based on Evolutionary Learning (KEEL) Repository (ALCALá-FDEZ et al., 2010). By the means of this performance analysis,

there is a comparison between the proposed model and the previous techniques ULST2-FLS (AGUIAR et al., 2018), Adam ULST2-FLS, SM Adam ULST2-FLS and MSE ULST2-FLS (FONSECA; AGUIAR, 2022) and three deep learning methods used for binary classification: Scalable Tree Boosting System (XGBoost) (CHEN; GUESTRIN, 2016), Convolutional Neural Network (CNN) (KIM, 2017) and Long Short-term Memory (LSTM) Network (PALUSZEK; THOMAS, 2020).

- Performance analysis of GWO ULST2-FLS in terms of convergence speed, accuracy, MSE, Cohen's kappa and F-score by using a data set composed of temperature measurements on trains wheels and bearings provided by MRS Logística S.A., to discern real hot box and hot wheels problems from the incorrect indications. By the means of this performance analysis, there is a comparison between the proposed models and the prior methods ULST2-FLS (AGUIAR et al., 2018), Adam ULST2-FLS, SM Adam ULST2-FLS and MSE ULST2-FLS (FONSECA; AGUIAR, 2022) and three deep learning methods used for binary classification: Scalable Tree Boosting System (XGBoost) (CHEN; GUESTRIN, 2016), Convolutional Neural Network (CNN) (KIM, 2017) and Long Short-term Memory (LSTM) Network (PALUSZEK; THOMAS, 2020).

And the major conclusions are:

- The use of the SI optimization GWO (MIRJALILI; MIRJALILI; LEWIS, 2014) contribute to increase convergence speed and classification rate of the training phase compared to the former approaches.
- The proposal is suitable for the problem addressed in this work, due to the fact that it achieved higher classification ratios during the test phase compared to the original models and the other deep learning classifiers.

This work is organised as follows: Section 2 addresses the problem formulation. Section 3 aims to discuss the concept of GWO combined with the ULST2-FLS. Section 4 presents the results of computer simulations. Finally, Section 5 states the main conclusions.

Publications

This section presents the publications by the author during this work.

The paper published in journal is:

- **FONSECA, L. D.**; AGUIAR, E. P. Stochastic optimization combined with type-2 fuzzy logic system for the classification of trends in hot boxes and hot wheels.

International Journal of Fuzzy Systems, v. 24, 08 2022. DOI:10.1007/s40815-022-01326-8.

The conference paper is:

- LIMA, J.; MEDEIROS, A.; AGUIAR, E. P.; SILVEIRA, D.; **FONSECA, L. D.**; GUERRA, T.; JR., V.; DANTAS, Y. Sistemas Fuzzy aplicados ao Handover em Redes LTE com Falhas de Cobertura. In: XXXVII Simpósio Brasileiro de Telecomunicações e Processamento de Sinais, 2019, Petrópolis. Anais de XXXVII Simpósio Brasileiro de Telecomunicações e Processamento de Sinais, 2019.

2 PROBLEM FORMULATION

Hot wheels and hot boxes are significant threats in any railway operation. The overheating of wheels can increase the fatigue and wear processes that might result in dangerous situations and failure. At the same time, a hot box, due to bearing defects, can cause fractures in axle journals and bearing failure which might costly train stoppages and even derailments.

According to United States Federal Railroad Administration (FRA) (FRA, c), from 2010 to 2016, there were 119 train derailments due to overheated bearings in the U.S. and Canada. One example is a train derailment accident caused by the catastrophic failure of a roller bearing and subsequent burning axle journal that occurred in 2013 near Sudbury, Ontario in Canada. Six cars with 12 car bodies and 20 containers derailed and struck the railway bridge. This bridge collapsed due to the impact and seven containers, some carrying dangerous goods, fell into the river, as shown in Figure 1 (TSB, 2013).

Figure 1 – Collapsed south end of the railway bridge (TSB, 2013).

In addition, as reported by the FRA (FRA, b; FRA, a), the North American railroad industry spends more than US$800 million annually on wheel removals. Approximately 20% of this expense is due to bearing failures and more than 60% to wheel failures, resulting in US$160 million and US$480 million annually, respectively. These failures are mostly caused by the effects of their overheating which is the main factor that can alert you to the impending failure.

Between 1997 and 2019, according to the Brazilian ANTF (National Association of

Railway Carriers) (ANTF,), the volume of goods transported by railroads increased by 95%. In order to significantly expand the rolling stock fleet, about US$600 million was invested in 2019. Thus, the number of locomotives on the railroads was raised from 1,154 in 1997 to 3,405 in 2019, representing an increase of 195%. In addition, the number of wagons increased from 43,816 to 115,434, resulting in an increment of 163%. These data demonstrate the growing number of rail transport, which, as a result, tends to increase defects.

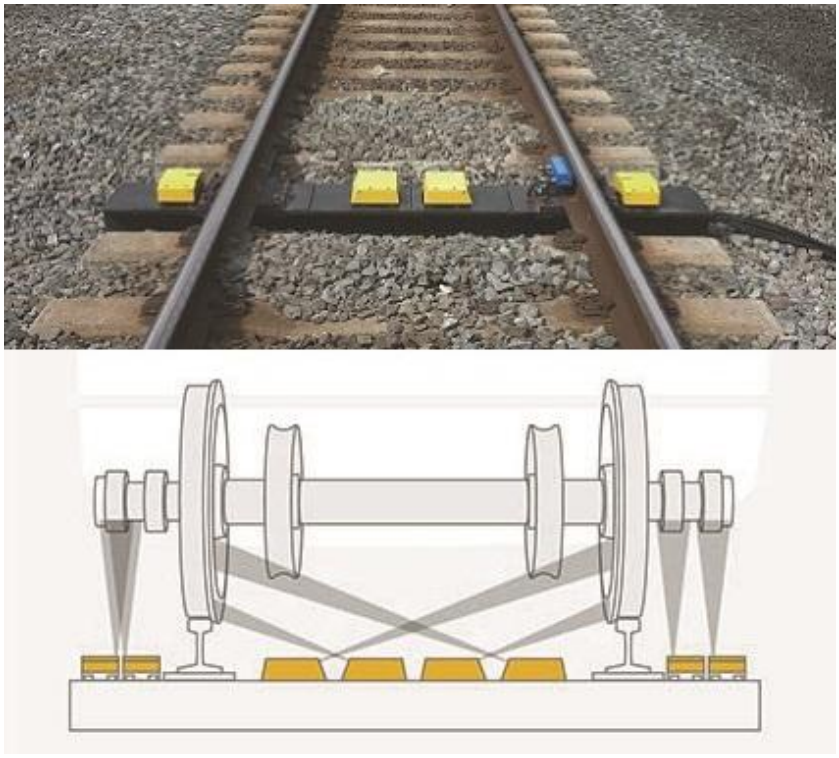

Figure 2 – Hot box detection and hot wheel detection systems (VOESTALPINE...,).

Therefore, for the purpose of predicting and preventing the failures mentioned above, bearing and wheels health monitoring systems were implemented by the railroad industry. These health monitoring systems are mainly the hot box detector (HBD) and the hot wheel detector (HWD), which are shown in Figure 2 (VOESTALPINE...,). The upper part of the figure shows a picture of HBD and HWD in a railway and the bottom part represents a schematic drawing showing how temperature measurements are taken. The main difference in the detectors on the real picture from the detectors on the drawing is the doubled sensors presented on the drawing which are showing different points of temperature measurement on the bearings and wheels of the train.

For illustrating how important is this issue, according to the Association of American Railroads (AAR) (AAR, 2015) more than 6,000 HBDs have been installed by the North American railroads. They are placed in order to reduce the risk of bearing failures due to overheating, and this results in the presence of one HBD system every approximately 25

miles on the Class I freight rail networks (BRAREN; KENNELLY; EIDE, 2009). When HBD and HWD determine that bearings and wheels temperatures are exceeding the temperature threshold, a warning is given to indicate potential issues. As a safety measure, the components are removed from service for later disassembly and inspection.

Although those detectors are widely used and have been able to prevent accidents from happening, the system is not flawless. Many variables, such as sensor defect, solar incidence or longitudinal misalignment, can affect the temperature measurements. For this reason, a HBD, for example, may greatly over-predict bearing temperature, leading to a significant number of warnings for bearings that do not have discernible defects. These bearings are then classified as "non-verified".

Non-verified bearing removals cause rail line interruptions and congestion of the rail network, due to the unnecessary train stoppages and delays, in addition to the use of unneeded maintenance time, parts and supplies. These unnecessary expenses are demonstrated by the investigation of Amsted Rail, which reveal that, from 2001 to 2007, almost 40% of bearing removals were "non-verified" bearings (TARAWNEH et al., 2018).

Under this perspective, distinguishing the components that indeed present defects from the "non-verified" ones is highly necessary in order to:

- reduce the number of unnecessary maintenance;
- reduce the number of unneeded train stoppage and delays;
- reduce the replacement of defect-free components, thereby reducing expenses with unnecessary new parts and supplies;
- increase the productivity of rail operations, once interruptions and congestion of the rail network are reduced.

In consideration of the scenario previously exposed in this section, this work aims to classify the warning signals of components that truly presents hot box or hot wheels problems (proper warnings) and the warning signs of components that do not present discernible defects (improper warnings).

Figure 3 illustrates the steps taken to accomplish this goal. The input matrix $\mathbf{I}$ consists of a list of temperature measurements on the right and left wheels and bearings of the train when a warning have been triggered. Then this data is treated by the feature extraction where the characteristics used for the classification are selected. The extracted features $\mathbf{K}$ are the mean, the median, the standard deviation, the minimum value and the maximum value of the measured temperatures and they are the input features for the classifier. At last, in the Classification block, the classification technique discussed in

this work is applied to obtain the output vector **s**, consequently identifying the proper or improper warnings.

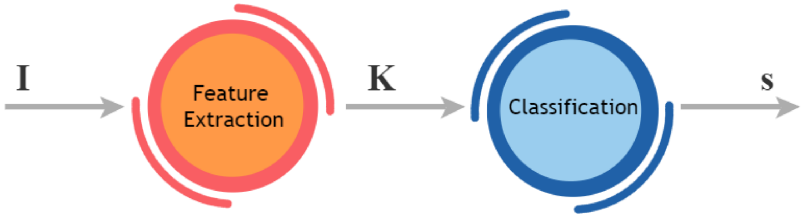

Figure 3 – Block diagram of the scheme for classification of events.

3 THE PROPOSAL: GREY WOLF OPTIMIZER COMBINED WITH ULST2-FLS

In this chapter, the background of ULST2-FLS (AGUIAR et al., 2018), Adam ULST2-FLS, SM Adam ULST2-FLS and MSE ULST2-FLS (FONSECA; AGUIAR, 2022) is first discussed. Then, the proposed method GWO ULST2-FLS is provided, including its inspiration and mathematical model.

3.1 The upper and lower singleton type-2 fuzzy logic system

Type-1 FLSs provide improved performance, however they unfortunately are unable to adequatly handle some uncertainties associated with the classification problem, in particular the uncertainties present in the measurements that activate the FLS or in the data that is used to tune the parameters of an FLS (MENDEL, 2001). So, in order to better handle the uncertainties, type-2 FLSs can be used and according to (MENDEL; JOHN; LIU, 2006), they have the potential to provide better performance than a type-1 FLS (e.g., (HAGRAS, 2004; LIANG; MENDEL, 2000a; LIANG; MENDEL, 2000c; LIANG; MENDEL, 2001; LIANG; KARNIK; MENDEL, 2000; MELIN; CASTILLO, 2004; OZEN; GARIBALDI, 2003; WU; TAN, 2004; WU; MENDEL, 2003)).

Figure 4 (AGUIAR et al., 2020) shows the components of a T1 and T2 FLSs. Both systems are composed of the same elements: the fuzzifier, the fuzzy rule base, the fuzzy inference engine and the output processor. So their rules are the same and the difference is only the models used for the membership functions. Take for example a gaussian function, type 1 membership function would be a simple gaussian, where as the type 2 membership function would present a footprint of uncertainty limited by the upper membership function and the lower membership function. The footprint of uncertainty can be seen in Figure 5 (ABIYEV, 2014).

Another major difference is the nature of the output processor, in which the T2 FLS presents a type-reducer before the defuzzifier while T1 FLS only presents the defuzzifier. The type-reducer operation converts the output type-2 membership function in its output type-1 membership function.

The proposed classifier are type-2 singleton FLS classifier based on the classifier ULST2-FLS introduced in (AGUIAR et al., 2018). They use singleton fuzzification, max-product composition, product implication, height type-reduction and make use of Gaussian primary membership function with uncertain mean.

Assuming the Gaussian membership functions with a fixed standard deviation σ, and an uncertain mean situated in $m \in [m_1, m_2]$, which is shown in Figure 5 (ABIYEV, 2014). The upper membership function and lower membership function are given, respectively, by:

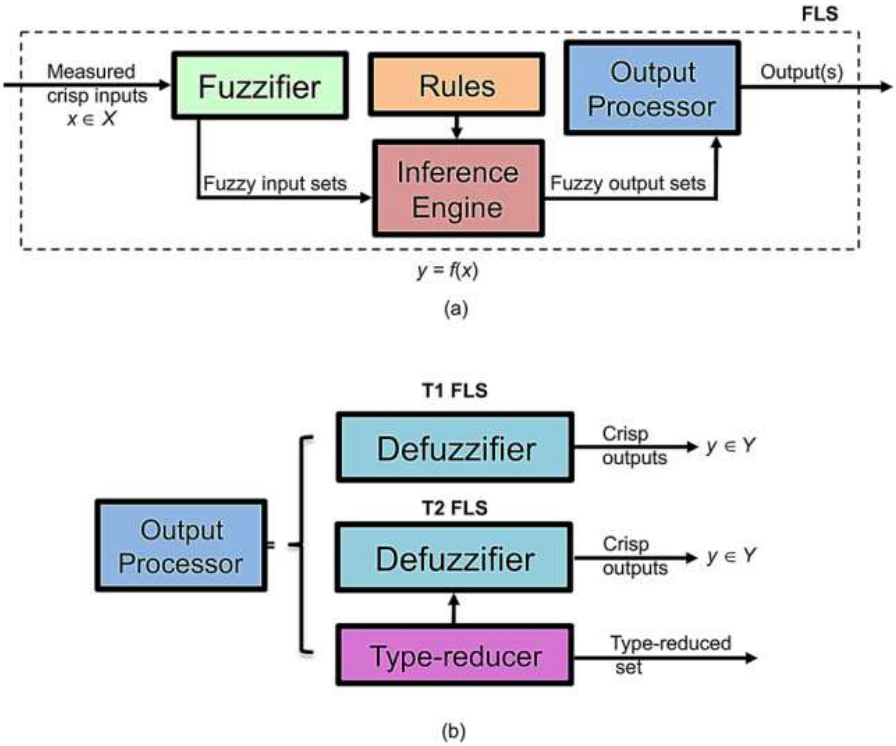

Figure 4 – (a) Components of a FLS, and (b) nature of the output processor for T1 and T2 FLSs (AGUIAR et al., 2020).

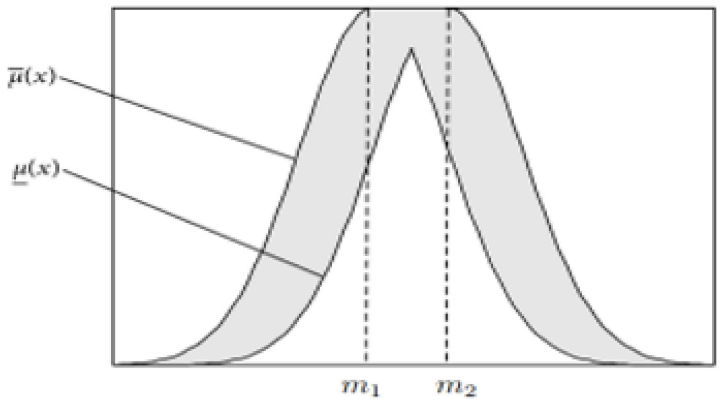

Figure 5 – Type 2 membership function with uncertainty in the mean (ABIYEV, 2014).

$$\overline{\mu}_{F_k^l}(x_k) = \begin{cases} exp\left[\frac{1}{2}\left(\frac{x-m_1}{\sigma}\right)^2\right], & x < m_1 \\ 1, & m_1 \leq x \leq m_2 \\ exp\left[\frac{1}{2}\left(\frac{x-m_2}{\sigma}\right)^2\right], & x > m_2 \end{cases} \tag{3.1}$$

and

$$\underline{\mu}_{F_k^l}(x_k) = \begin{cases} exp\left[\frac{1}{2}\left(\frac{x-m_2}{\sigma}\right)^2\right], & x \leq \frac{m_1+m_2}{2} \\ \\ exp\left[\frac{1}{2}\left(\frac{x-m_1}{\sigma}\right)^2\right], & x > \frac{m_1+m_2}{2} \end{cases}. \tag{3.2}$$

The ULST2-FLS presents the type reducer using the following structure:

$$f_{s2}(\mathbf{x}) = [Y_{Up}(\mathbf{x}), Y_{Low}(\mathbf{x})] \tag{3.3}$$

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

$$Y_{Up}(x_k) = \frac{\sum_{l=1}^{M} \theta_l \prod_{k=1}^{P} \overline{\mu}_{F_k^l}(x_k)}{\sum_{l=1}^{M} \prod_{k=1}^{P} \overline{\mu}_{F_k^l}(x_k)} \tag{3.4}$$

and

$$Y_{Low}(x_k) = \frac{\sum_{l=1}^{M} \theta_l \prod_{k=1}^{P} \underline{\mu}_{F_k^l}(x_k)}{\sum_{l=1}^{M} \prod_{k=1}^{P} \underline{\mu}_{F_k^l}(x_k)} \tag{3.5}$$

where θ_l is the weight associated with the lth rule, $l = 1, .., M$, and x_k is the kth input, $k = 1, ..., P$.

The defuzzification of ULST2-FLS is given by:

$$f_{s2}(\mathbf{x}) = \frac{Y_{Up}(\mathbf{x}) + Y_{Low}(\mathbf{x})}{2}. \tag{3.6}$$

For a set of input–output pairs $(\mathbf{x}^{(q)} : y^{(q)})$, where q indicates the qth iteration used for training the ULST2-FLS. This classifier follows with the problem of minimizing the cost function below:

$$\mathbf{J}\left(\mathbf{w}^{(q)}\right) = \frac{1}{2}\left[f_{s2}\left(\mathbf{x}^{(q)}\right) - y^{(q)}\right]^2, \tag{3.7}$$

in which $\mathbf{w}^{(q)}$ represents the vector of parameters for ULST2-FLS, defined by:

$$\begin{aligned} \mathbf{w}^{(q)} = \Big[& m_{1F_1^1}(q), ..., m_{1F_P^1}(q), ..., m_{1F_1^M}(q), ..., m_{1F_P^M}(q), \\ & m_{2F_1^1}(q), ..., m_{2F_P^1}(q), ..., m_{2F_1^M}(q), ..., m_{2F_P^M}(q), \\ & \sigma_{F_1^1}(q), ..., \sigma_{F_P^1}(q), ..., \sigma_{F_1^M}(q), ..., \sigma_{F_P^M}(q), \\ & \theta_1(q), ..., \theta_M(q)\Big] \end{aligned} \tag{3.8}$$

and the gradient $\nabla\mathbf{J}(\mathbf{w}^{(q)})$ of eq. (3.7) is expressed by:

$$\begin{aligned}\nabla \mathbf{J}\left(\mathbf{w}^{(q)}\right) = \Bigg[& \frac{\partial \mathbf{J}(\mathbf{w}^{(q)})}{\partial m_{1F_1^1}(q)}, ..., \frac{\partial \mathbf{J}(\mathbf{w}^{(q)})}{\partial m_{1F_P^1}(q)}, ..., \frac{\partial \mathbf{J}(\mathbf{w}^{(q)})}{\partial m_{1F_1^M}(q)}, ..., \frac{\partial \mathbf{J}(\mathbf{w}^{(q)})}{\partial m_{1F_P^M}(q)}, \\ & \frac{\partial \mathbf{J}(\mathbf{w}^{(q)})}{\partial m_{2F_1^1}(q)}, ..., \frac{\partial \mathbf{J}(\mathbf{w}^{(q)})}{\partial m_{2F_P^1}(q)}, ..., \frac{\partial \mathbf{J}(\mathbf{w}^{(q)})}{\partial m_{2F_1^M}(q)}, ..., \frac{\partial \mathbf{J}(\mathbf{w}^{(q)})}{\partial m_{2F_P^M}(q)}, \\ & \frac{\partial \mathbf{J}(\mathbf{w}^{(q)})}{\partial \sigma_{F_1^1}(q)}, ..., \frac{\partial \mathbf{J}(\mathbf{w}^{(q)})}{\partial \sigma_{F_P^1}(q)}, ..., \frac{\partial \mathbf{J}(\mathbf{w}^{(q)})}{\partial \sigma_{F_1^M}(q)}, ..., \frac{\partial \mathbf{J}(\mathbf{w}^{(q)})}{\partial \sigma_{F_P^M}(q)}, \\ & \frac{\partial \mathbf{J}(\mathbf{w}^{(q)})}{\partial \theta_1(q)}, ..., \frac{\partial \mathbf{J}(\mathbf{w}^{(q)})}{\partial \theta_M(q)} \Bigg]\end{aligned} \tag{3.9}$$

Accordingly, the parameters are updated by the following equations:

$$m_{1F_k^l}(q+1) = m_{1F_k^l}(q) - \alpha_s \frac{\partial \mathbf{J}(\mathbf{w}^{(q)})}{\partial m_{1F_k^l}(q)} \tag{3.10}$$

$$m_{2F_k^l}(q+1) = m_{2F_k^l}(q) - \alpha_s \frac{\partial \mathbf{J}(\mathbf{w}^{(q)})}{\partial m_{2F_k^l}(q)} \tag{3.11}$$

$$\sigma_{F_k^l}(q+1) = \sigma_{F_k^l}(q) - \alpha_s \frac{\partial \mathbf{J}(\mathbf{w}^{(q)})}{\partial \sigma_{F_k^l}(q)} \tag{3.12}$$

$$\theta_l(q+1) = \theta_l(q) - \alpha_s \frac{\partial \mathbf{J}(\mathbf{w}^{(q)})}{\partial \theta_l(q)} \tag{3.13}$$

in which α_s is the step size, and the first-order derivative of $\partial \mathbf{J}(\mathbf{w}^{(q)})$ with respect to the parameters $m_{1F_k^l}(q)$, $m_{2F_k^l}(q)$, $\sigma_{F_k^l}(q)$ and $\theta_l(q)$ are described in (AGUIAR et al., 2018).

3.2 Adam and SM combined with upper and lower singleton type-2 fuzzy logic system

The difference in the training phase between the traditional ULST2-FLS and the method Adam ULST2-FLS is that the latter is based on estimates of first and second moment of the first-order gradients. So the following calculations are based on the work presented by (KINGMA; BA, 2014). They start calculating $\mathbf{u}$ and $\mathbf{v}$ that are vectors composed respectively by:

$$\begin{aligned}\mathbf{u}^{(q)} = \Big[& u_{m_{1F_1^1}}(q), ..., u_{m_{1F_P^1}}(q), ..., u_{m_{1F_1^M}}(q), ..., u_{m_{1F_P^M}}(q), \\ & u_{m_{2F_1^1}}(q), ..., u_{m_{2F_P^1}}(q), ..., u_{m_{2F_1^M}}(q), ..., u_{m_{2F_P^M}}(q), \\ & u_{\sigma_{F_1^1}}(q), ..., u_{\sigma_{F_P^1}}(q), ..., u_{\sigma_{F_1^M}}(q), ..., u_{\sigma_{F_P^M}}(q), \\ & u_{\theta_1}(q), ..., u_{\theta_M}(q) \Big]\end{aligned} \tag{3.14}$$

and

$$\begin{aligned}\mathbf{v}^{(q)} =&\Big[v_{m_{1F_1^1}}(q),...,v_{m_{1F_P^1}}(q),...,v_{m_{1F_1^M}}(q),...,v_{m_{1F_P^M}}(q),\\ &v_{m_{2F_1^1}}(q),...,v_{m_{2F_P^1}}(q),...,v_{m_{2F_1^M}}(q),...,v_{m_{2F_P^M}}(q),\\ &v_{\sigma_{F_1^1}}(q),...,v_{\sigma_{F_P^1}}(q),...,v_{\sigma_{F_1^M}}(q),...,v_{\sigma_{F_P^M}}(q),\\ &v_{\theta_1}(q),...,v_{\theta_M}(q)\Big]\end{aligned} \tag{3.15}$$

These vectors are initiated as vectors of zeros and then are calculated respectively by:

$$\mathbf{u}^{(q+1)} = \beta_1 \cdot \mathbf{u}^{(q)} + (1-\beta_1)\cdot \nabla \mathbf{J}(\mathbf{w}^{(q)}) \tag{3.16}$$

and

$$\mathbf{v}^{(q+1)} = \beta_2 \cdot \mathbf{v}^{(q)} + (1-\beta_2)\cdot \left[\nabla \mathbf{J}(\mathbf{w}^{(q)})\right]^2 \tag{3.17}$$

where β_1 and β_2 are values that control the exponential decay rates for the moment estimates and β_1, $\beta_2 \in [0, 1)$.

From the estimates of first and second moment, it is possible to compute bias-corrected estimates of first and second moments, $\widehat{\mathbf{u}}$ and $\widehat{\mathbf{v}}$, expressed by:

$$\begin{aligned}\widehat{\mathbf{u}}^{(q)} =&\Big[\widehat{u}_{m_{1F_1^1}}(q),...,\widehat{u}_{m_{1F_P^1}}(q),...,\widehat{u}_{m_{1F_1^M}}(q),...,\widehat{u}_{m_{1F_P^M}}(q),\\ &\widehat{u}_{m_{2F_1^1}}(q),...,\widehat{u}_{m_{2F_P^1}}(q),...,\widehat{u}_{m_{2F_1^M}}(q),...,\widehat{u}_{m_{2F_P^M}}(q),\\ &\widehat{u}_{\sigma_{F_1^1}}(q),...,\widehat{u}_{\sigma_{F_P^1}}(q),...,\widehat{u}_{\sigma_{F_1^M}}(q),...,\widehat{u}_{\sigma_{F_P^M}}(q),\\ &\widehat{u}_{\theta_1}(q),...,\widehat{u}_{\theta_M}(q)\Big]\end{aligned} \tag{3.18}$$

and

$$\begin{aligned}\widehat{\mathbf{v}}^{(q)} =&\Big[\widehat{v}_{m_{1F_1^1}}(q),...,\widehat{v}_{m_{1F_P^1}}(q),...,\widehat{v}_{m_{1F_1^M}}(q),...,\widehat{v}_{m_{1F_P^M}}(q),\\ &\widehat{v}_{m_{2F_1^1}}(q),...,\widehat{v}_{m_{2F_P^1}}(q),...,\widehat{v}_{m_{2F_1^M}}(q),...,\widehat{v}_{m_{2F_P^M}}(q),\\ &\widehat{v}_{\sigma_{F_1^1}}(q),...,\widehat{v}_{\sigma_{F_P^1}}(q),...,\widehat{v}_{\sigma_{F_1^M}}(q),...,\widehat{v}_{\sigma_{F_P^M}}(q),\\ &\widehat{v}_{\theta_1}(q),...,\widehat{v}_{\theta_M}(q)\Big]\end{aligned} \tag{3.19}$$

These vectors are obtained by the following equations:

$$\widehat{\mathbf{u}}^{(q+1)} = \frac{\mathbf{u}^{(q)}}{1-(\beta_1)^{q+1}} \tag{3.20}$$

and

$$\widehat{\mathbf{v}}^{(q+1)} = \frac{\mathbf{v}^{(q)}}{1-(\beta_2)^{q+1}} \tag{3.21}$$

At last, the equations used to update the parameters in Adam ULST2-FLS are (FONSECA; AGUIAR, 2022):

$$m_{1F_k^l}(q+1) = m_{1F_k^l}(q) - \alpha_s \frac{\widehat{u}_{m_{1F_k^l}}(q+1)}{\sqrt{\widehat{v}_{m_{1F_k^l}}(q+1)} + \epsilon} \tag{3.22}$$

$$m_{2F_k^l}(q+1) = m_{2F_k^l}(q) - \alpha_s \frac{\widehat{u}_{m_{2F_k^l}}(q+1)}{\sqrt{\widehat{v}_{m_{2F_k^l}}(q+1)} + \epsilon} \tag{3.23}$$

$$\sigma_{F_k^l}(q+1) = \sigma_{F_k^l}(q) - \alpha_s \frac{\widehat{u}_{\sigma_{F_k^l}}(q+1)}{\sqrt{\widehat{v}_{\sigma_{F_k^l}}(q+1)} + \epsilon} \tag{3.24}$$

$$\theta_l(q+1) = \theta_l(q) - \alpha_s \frac{\widehat{u}_{\theta_l}(q+1)}{\sqrt{\widehat{v}_{\theta_l}(q+1)} + \epsilon} \tag{3.25}$$

in which α_s is the stepsize and ϵ is a small number to prevent the denominator from being zero.

In addition to the new parameters update method, the concept of set-membership was added, creating the SM Adam ULST2-FLS. This technique was proposed based on the sucess obtain in the use of the ser-membership concept that achieved great results with fast convergence speed during training phase with fewer use of hardware resources (AGUIAR et al., 2017; CLARKE; LAMARE, 2011; SALMENTO et al., 2017; SALMENTO; AGUIAR; RIBEIRO, 2015; WANG; DELAMARE, 2013; DINIZ, 2002; ABADI; ESKANDARI, 2013). Thus, the values β_1 and β_2 are replaced respectively by variable exponential decay rates $\mu_1(q)$ and $\mu_2(q)$, defined by:

$$\mu_1(q+1) = \begin{cases} \mu_1(q) + \dfrac{error(q)}{10^3 \cdot \overline{\gamma}}, & if\ error > \overline{\gamma} \\ \\ \mu_1(q) - \dfrac{error(q)}{10^3 \cdot \overline{\gamma}}, & otherwise \end{cases} \tag{3.26}$$

and

$$\mu_2(q+1) = \begin{cases} \mu_2(q) + \dfrac{error(q)}{10^3 \cdot \overline{\gamma}}, & if\ error > \overline{\gamma} \\ \\ \mu_2(q) - \dfrac{error(q)}{10^3 \cdot \overline{\gamma}}, & otherwise \end{cases} \tag{3.27}$$

where $error(q) = |f_{s2}\left(\mathbf{x}^{(q)}\right) - y^{(q)}|$ and $\overline{\gamma}$ is the upper bound constraint.

Finally the last modification in the original classifier ULST2-FLS was using the mean square error instead of the error modulus to calculate the variable exponential decay rates, resulting in MSE Adam ULST2-FLS. The utilization of mean square error has

consistently enhanced both accuracy and convergence speed in prior studies (AGUIAR et al., 2020), thus prompting its proposal in this context. So, $\mu_1(q)$ and $\mu_2(q)$ are calculated by the Eqs. (3.26) and (3.27) replacing $error(q)$ with $MSE(q) = \frac{1}{2}\left[f_{s2}\left(\mathbf{x}^{(q)}\right) - y^{(q)}\right]^2$.

3.3 GWO ULST2-FLS

A new method to update the parameters is proposed in this work, using the SI algorithm GWO (MIRJALILI; MIRJALILI; LEWIS, 2014). It is an efficient method based on the leadership hierarchy and hunting mechanism of grey wolves. This model is simple to implement, has little memory requirement and presents computationally efficiency. These characteristics, along with the fact that it doesn't require the computation of the gradient, make GWO a great alternative for the above mentioned methods.

The key features of the algorithm are saving the best solutions obtained so far over the course of iteration, using random parameters for exploration which avoids local stagnation, locating quickly the probable best solutions, smoothly transitioning between exploration and exploitation, and presents minimal parameters to adjust.

Grey Wolf Optimizer is inspired by the social hierarchy and hunting behavior of grey wolves. Grey wolves live in packs with a strict dominant hierarchy consisting of alpha, beta, delta, and omega wolves, as shown in Figure 6 (MIRJALILI; MIRJALILI; LEWIS, 2014). The alpha, despite not necessarily being the strongest, leads the pack, while betas support decision-making and discipline. Deltas take other roles of some importance and include scouts, sentinels, elders, hunters, and caretakers. At last the omegas, although they are at the lowest hierarchical level, play a crucial role as scapegoats and babysitters, maintaining pack harmony.

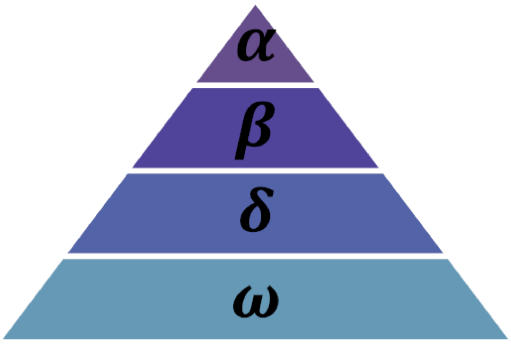

Figure 6 – Hierarchy of grey wolfs in which dominance decreases from top to down (MIRJALILI; MIRJALILI; LEWIS, 2014).

The GWO algorithm mathematically models this social hierarchy and hunting behavior to perform optimization. It divides candidate solutions into alphas, betas, deltas,

and omegas, with the search guided by their positions. The fittest solution is considered the alpha (α). The beta (β) and delta (δ) are the second and third best solutions respectively. And the rest of candidate solutions are assumed omegas (ω). The optimization process is guided by α, β and δ) solutions. Meanwhile the ω solutions follow the other three.

Aside from the social structure of wolves, collective hunting is another fascinating characteristics among grey wolves that are replicated by GWO. The hunting typically progresses through the following phases: tracking, chasing, and approaching the prey; followed by a phase of pursuing, encircling, and harassing the prey until it stops moving; culminating in a direct attack on the prey. These phases can be seen in Figure 7 (MIRJALILI; MIRJALILI; LEWIS, 2014).

Figure 7 – Hunting behavior of grey wolves: (A) chasing, approaching, and tracking prey; (B–D) pursuing, harassing, and encircling; (E) stationary situation and attack (MIRJALILI; MIRJALILI; LEWIS, 2014).

Encircling prey is simulated by adjusting positions based on the location of the prey. The proposed equations to mathematically model this encircling behavior are (MIRJALILI; MIRJALILI; LEWIS, 2014):

$$\mathbf{D} = |\mathbf{C} \cdot \mathbf{X}_p(t) - \mathbf{X}(t)| \tag{3.28}$$

$$\mathbf{X}(t+1) = \mathbf{X}_p(t) - \mathbf{A} \cdot \mathbf{D} \tag{3.29}$$

where t indicates the current iteration, $\mathbf{A}$ and $\mathbf{C}$ are coefficient vectors, $\mathbf{X}_p(t)$ is the position vector of the prey, and $\mathbf{X}(t)$ indicates the position vector of a grey wolf.

The vectors $\mathbf{A}$ and $\mathbf{C}$ are calculated as follows (MIRJALILI; MIRJALILI; LEWIS, 2014):

$$\mathbf{A} = 2 \cdot \mathbf{r}_1 - \mathbf{a} \tag{3.30}$$

$$\mathbf{C} = 2 \cdot \mathbf{r}_2 \tag{3.31}$$

where components of the vector $\mathbf{a}$ are linearly decreased from 2 to 0 over the course of iterations and $\mathbf{r}_1$ and $\mathbf{r}_2$ are random vectors in $[0, 1]$.

Grey wolves possess the ability to recognize the location of the prey and encircle them. Typically, the alpha leads the hunt, although the beta and delta may also participate in hunting occasionally. However, in an abstract search space, there is no prior knowledge of the optimal prey location. Then, in order to mathematically replicate grey wolf hunting behavior, it is assumed that the alpha, beta, and delta have superior knowledge about potential prey locations. Consequently, the three best solutions obtained thus far represents wolves α, β and δ and they are saved for the next iteration. Meanwhile, the other search agents, the ω wolves, have their positions updated based on the positions of the leading search agents. This behavior is mathematically proposed by the following formulas (MIRJALILI; MIRJALILI; LEWIS, 2014):

$$\mathbf{D}_\alpha = |\mathbf{C}_1 \cdot \mathbf{X}_\alpha - \mathbf{X}|,\ \mathbf{D}_\beta = |\mathbf{C}_2 \cdot \mathbf{X}_\beta - \mathbf{X}|,\ \mathbf{D}_\delta = |\mathbf{C}_3 \cdot \mathbf{X}_\delta - \mathbf{X}| \tag{3.32}$$

$$\mathbf{X}_1 = \mathbf{X}_\alpha - \mathbf{A}_1 \cdot (\mathbf{D}_\alpha),\ \mathbf{X}_2 = \mathbf{X}_\beta - \mathbf{A}_2 \cdot (\mathbf{D}_\beta),\ \mathbf{X}_3 = \mathbf{X}_\delta - \mathbf{A}_3 \cdot (\mathbf{D}_\delta) \tag{3.33}$$

$$\mathbf{X}(t+1) = \frac{\mathbf{X}_1 + \mathbf{X}_2 + \mathbf{X}_3}{3}. \tag{3.34}$$

As previously stated, grey wolves conclude their hunt by attacking the prey when it stops moving. The mathematical approach of this phenomenon is the decrease of the value $\mathbf{a}$. It is important to note that the variability of A is also decreasing by $\mathbf{a}$. Moreover $\mathbf{A}$ is a component that favors exploration by mathematically modeling the divergence between the wolves α, β, and δ to search for the prey. Finally the coefficient $\mathbf{C}$ also helps the exploration factor, showing a more random behavior throughout optimization, which avoids local optima.

In the context of ULST2-FLS, each vector of parameters $\mathbf{w}_n$ in a population of n vectors of parameters will be sorted into alpha, beta, delta or omegas. Therefore the

vectors which provided the three best solutions are the α, β and δ, which continue in the search for the next iteration. On the other hand, the rest of the vectors of parameters will be updated guided by the other three. The following equations are adaptations from Equations (3.32) and (3.33) that represent the case of ULST2-FLS:

$$\mathbf{D}_\alpha = |\mathbf{C}_1 \cdot \mathbf{w}_\alpha - \mathbf{w}|,\ \mathbf{D}_\beta = |\mathbf{C}_2 \cdot \mathbf{w}_\beta - \mathbf{w}|,\ \mathbf{D}_\delta = |\mathbf{C}_3 \cdot \mathbf{w}_\delta - \mathbf{w}| \tag{3.35}$$

$$\mathbf{w}_1 = \mathbf{w}_\alpha - \mathbf{A}_1 \cdot (\mathbf{D}_\alpha),\ \mathbf{w}_2 = \mathbf{w}_\beta - \mathbf{A}_2 \cdot (\mathbf{D}_\beta),\ \mathbf{w}_3 = \mathbf{w}_\delta - \mathbf{A}_3 \cdot (\mathbf{D}_\delta) \tag{3.36}$$

Lastly, the equation used to update the parameters in GWO ULST2-FLS is:

$$\mathbf{w}(t+1) = \frac{\mathbf{w}_1 + \mathbf{w}_2 + \mathbf{w}_3}{3}. \tag{3.37}$$

Algorithm 1 reproduces the GWO ULST2-FLS training process.

Algorithm 1 Training algorithm for GWO ULST2-FLS

Inputs:
N: Number of samples;
t_{max}: Maximum number of iterations;
n: Number of wolves
$(\mathbf{x}^{(q)} : y^{(q)})$: Set of input-output pairs;

Output:
accuracy: Accuracy of GWO ULST2-FLS;

procedure GWO ULST2-FLS TRAINING
    Initialize the grey wolf population $\mathbf{w}_i$ $(i = 1,\ 2,\ \ldots,\ n)$
    for $t = 1, \ldots, t_{max}$ **do**
        for $i = 1. \ldots, n$ **do**
            for $q = 1, \ldots, N$ **do**
                Evaluate $Y_{Up}(\mathbf{x})$ and $Y_{Low}(\mathbf{x})$, according to Eqs. (3.4) and (3.5);
                Calculate Eq. (3.6);
                if $f_{s2}(\mathbf{x}^{(q)}) > 0$ **then**
                    $f_{s2}(\mathbf{x}^{(q)}) = 1$;
                else
                    $f_{s2}(\mathbf{x}^{(q)}) = -1$;
                end if
            end for
            $wolf.accuracy(i) = (number\ of\ correct\ classifications)\ /N$
        end for
        Calculate $\mathbf{a}$, $\mathbf{r}_1$, $\mathbf{r}_2$, $\mathbf{A}$ and $\mathbf{C}$
        Update wolves α, β and δ:
        $\mathbf{w}_\alpha = the\ greater\ accuracy\ wolf$
        $\mathbf{w}_\beta = the\ second\ greater\ accuracy\ wolf$
        $\mathbf{w}_\delta = the\ third\ greater\ accuracy\ wolf$
        $\mathbf{D}_\alpha = |\mathbf{C}_1 \cdot \mathbf{w}_\alpha - \mathbf{w}|,\ \mathbf{D}_\beta = |\mathbf{C}_2 \cdot \mathbf{w}_\beta - \mathbf{w}|,\ \mathbf{D}_\delta = |\mathbf{C}_3 \cdot \mathbf{w}_\delta - \mathbf{w}|$
        $\mathbf{w}_1 = \mathbf{w}_\alpha - \mathbf{A}_1 \cdot (\mathbf{D}_\alpha),\ \mathbf{w}_2 = \mathbf{w}_\beta - \mathbf{A}_2 \cdot (\mathbf{D}_\beta),\ \mathbf{w}_3 = \mathbf{w}_\delta - \mathbf{A}_3 \cdot (\mathbf{D}_\delta)$
        Update parameters:
        $\mathbf{w}(t+1) = (\mathbf{w}_1 + \mathbf{w}_2 + \mathbf{w}_3)/3$
    end for
    $accuracy = accuracy\ of\ the\ wolf\ \alpha$
end procedure

4 Experimental Results

In order to analyze the performance of the proposed classifier, this section is divided in two different subsections. The first approaches the classifier effectiveness in three different benchmarks. Meanwhile, the second include the performance analyses in the hot box and hot wheels problem.

4.1 Benchmarks

The data sets chosen for the performance analyses were Appendicitis, Haberman, and Sonar provided by Knowledge Extraction based on Evolutionary Learning (KEEL) Repository (ALCALá-FDEZ et al., 2010). Table 1 presents information about the number of samples and the input features of them. All data sets must be associated with binary classification problems and they were chosen to represent a variety of number of samples and input features.

A prior study established that utilizing four rules optimally balances performance evaluation (AGUIAR et al., 2018). Utilizing more than four rules does not enhance performance but rather incurs higher computational costs. Consequently the classifier presented in this paper is composed of four rules, two rules for the class that has the presence of the event and other two rules for the class that does not have the presence of the event.

These rules helps the initialization process. The first rule of each class is heuristically created using the uncertain mean and the standard deviation from the training data set inputs of the respective class. In order to obtain the uncertain mean, two variations of the mean are generated heuristically by:

$$m_{1F} = {m_F}^1 - U{m_F}^1 \tag{4.1}$$

and

$$m_{2F} = {m_F}^1 + U{m_F}^1 \tag{4.2}$$

in which the value $U \in R \mid 0 \leq U \leq 1$.

Table 1 – Details of data sets

Data sets	Number of Samples	Input Features
Appendicitis[1]	106	7
Haberman[1]	306	3
Sonar[1]	208	60

[1]Data sets provided by Knowledge Extraction based on Evolutionary Learning (KEEL) Repository (ALCALá-FDEZ et al., 2010).

Summarizing, the following equation demonstrates the process of creation of the first rule.

$$\begin{aligned} rule_t^1 = \{If\ x_1\ is\ F_1^1(m_{1F_1^1}, m_{2F_1^1}, \sigma_{F_1^1})\ and...\ and \\ x_p\ is\ F_p^1(m_{1F_p^1}, m_{2F_p^1}, \sigma_{F_p^1}),\ then\ Y_t^1\} \end{aligned} \tag{4.3}$$

In (4.3), $rule_t^1$ presents the parameters of the membership functions associated with the first rule from tth class.

As for the creation of the second rule, a modified version of the first rule is generated. Essentially, it was heuristically adopted that the parameter uncertain mean (mean-1 and mean-2) are scaled by a value $A_m \in R \mid -1 \leq A_m \leq 1$ and the parameter standard deviation is scaled by a value $A_\sigma \in R$. Therefore,

$$\begin{aligned} &rule_t^2 \\ &= \{If\ x_1\ is\ F_1^2(A_m m_{1F_1^2}, -A_m m_{2F_1^2}, A_\sigma \sigma_{F_1^2})\ and... \\ &\quad and\ x_p\ is\ F_p^2(A_m m_{1F_p^2}, -A_m m_{2F_p^2}, A_\sigma \sigma_{F_p^2}), \\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad then\ Y_t^2\} \end{aligned} \tag{4.4}$$

describes the second rule. For this work, it was adopted $Y_l^1 = 1$, $Y_l^2 = -1$, $U = 0.3$, $A_m = 0.3$ and $A_\sigma = 1.5$. Figure 8 illustrates the procedure for creating and initializing rules in order to clarify this issue.

All numerical simulations have been carried out in Python 3.8.8 environment running in an Intel Core i7-5500U 2.40 GHz CPU. The operation system was a Windows 10 64 bits. All data sets have been presented for each classifier 33 times and, each time, seventy percent of the data set was randomly separated for the training phase and the thirty percent left was used for the test phase.

The maximum value of iterations assumed for the training phase was 100, because the improvement in accuracy was low for more iterations, and it did not compensate for the computational load that a greater number of epochs required. Therefore, there was no need of adopting a stopping criterion based on error. After evaluating performance, the number of wolves were heuristically adopted as ten as a greater number did not improved the classification efficiency. The performance metrics used for the numerical results analyses were accuracy, mean squared error (MSE), Cohen's kappa coefficient and F-score.

Although this work focuses on improvement of the training phase in the type-2 FLS, for an investigative purpose, it is important to present performance comparison with others non-fuzzy classifiers. For this reason, the datasets were also applied in binary classifiers based on different machine learning techniques, such as a gradient boosting:

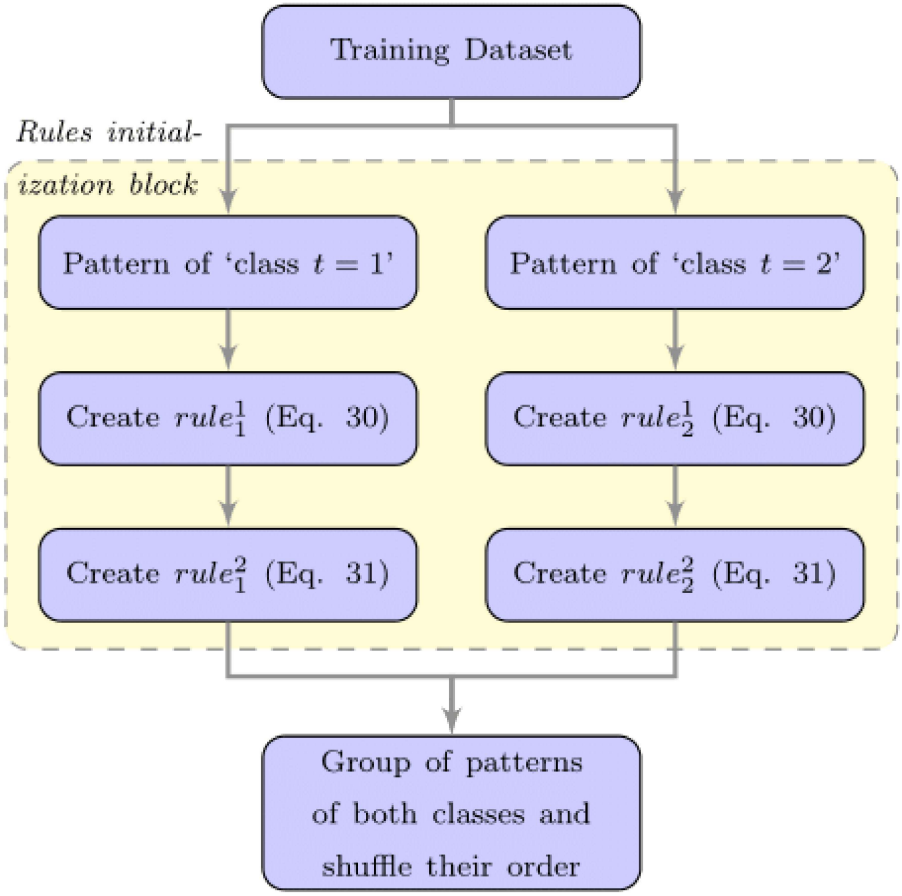

Figure 8 – Procedure of rule initialization applied for training T2FLS.

the Scalable Tree Boosting System (XGBoost) (CHEN; GUESTRIN, 2016); and two deep learning methods: Convolutional Neural Network (CNN) (KIM, 2017) and Long Short-term Memory (LSTM) Network (PALUSZEK; THOMAS, 2020). This decision enhaces performance analysis because the eficiency of the proposal is compared not only with its predecessors but also with classifiers from diverse research lines.

4.1.1 Initialization analysis

Table 2 presents the initial and the optimized values of MBF parameters, that is, the values before and after the training phase using the GWO ULST2-FLS for all the rules and for the Appendicitis data set. The comparison between these values reveal that the initial values undergo a lot of changes during the training phase, but it is also noticeable that the chosen procedure to initialize the parameters can result in a acceptable initialization.

Table 2 – Values of the mean and standard deviation for the initial and optimized MBFs for Appendicitis data set using GWO ULST2-FLS.

	Initial MBF				Optimized MBF			
	$m_{1F_1^l}$	$m_{2F_1^l}$	$\sigma_{F_1^l}$	Y_t^l	$m_{1F_1^l}$	$m_{2F_1^l}$	$\sigma_{F_1^l}$	Y_t^l
$rule_1^1$	0.1169	0.2171	0.4416	1	-0.0507	0.2472	0.9368	1
$rule_1^2$	0.3036	0.5639	1.2583	1	0.6065	0.8807	0.9133	1
$rule_2^1$	0.4169	0.5171	0.6624	-1	-0.1416	-0.0061	3.1916	-1
$rule_2^2$	0.0036	0.2639	1.8874	-1	-0.1047	0.6290	1.9105	-1

4.1.2 Classification rate analysis

Results obtained by the proposal, the four original fuzzy classifiers and the two deep learning methods are presented in Table 3.

Starting with the analysis of the fuzzy classifiers, it is possible to notice that the algorithm with GWO yielded competitive results in the trainig phase for all the metrics. Meanwhile, in the test phase, the proposal presented greater values of accuracy and kappa coefficient for almost all data sets.

In relation to the comparison of the fuzzy and non-fuzzy classifiers, by analysing the results, it is noticeable that the proposed classifier obtained greater mean accuracy in both training and testing phases. Additionally, they achieved lower values of MSE and better results for Kappa and F-score for most of the data-set when compared to the deep learning classifiers.

4.1.3 Convergence speed analysis

Figs. 9-11 show the convergence speed from the five different fuzzy classifiers for all the data sets. It is possible to notice that the proposed GWO ULST2-FLS classifier offers compatible results to the previous techniques in terms of accuracy. A limitation of the presented technique is that the mean square error during the training phase does not significantly decrease with the epochs. This might occur because, in each iteration, the best candidate solution changes significantly. Although this increases accuracy in each iteration, it does not reduce the MSE.

4.1.4 Statistical analysis

Statistical analysis was performed as in (AMARAL; RIBEIRO; AGUIAR, 2019). Therefore the two-sample t-test was applied for the test accuracy metric in order to

Table 3 – Performance comparison in terms of the mean and standard deviation.

Data Set	Method	Training accuracy %	Test accuracy %	Training MSE	Test MSE	Training Kappa	Test Kappa	Training Fscore	Test Fscore
Appendicitis	ULST2-FLS	90.25 (±2.21)	83.14 (±5.06)	0.31 (±0.07)	0.54 (±0.17)	0.65 (±0.08)	0.41 (±0.17)	0.94 (±0.01)	0.90 (±0.03)
	Adam ULST2-FLS	93.20 (±2.47)	83.14 (±4.20)	0.22 (±0.07)	0.58 (±0.16)	0.77 (±0.09)	0.43 (±0.16)	0.96 (±0.02)	0.90 (±0.03)
	SM Adam ULST2-FLS	**94.27 (±2.44)**	84.66 (±5.81)	0.20 (±0.08)	0.58 (±0.21)	**0.80 (±0.08)**	0.47 (±0.18)	**0.97 (±0.01)**	**0.91 (±0.04)**
	MSE Adam ULST2-FLS	93.86 (±2.37)	85.32 (±4.97)	0.22 (±0.06)	0.51 (±0.19)	0.78 (±0.09)	0.47 (±0.18)	0.96 (±0.01)	**0.91 (±0.03)**
	GWO ULST2-FLS	90.75 (±2.37)	**86.27 (±4.62)**	0.81 (±0.10)	0.82 (±0.12)	0.56 (±0.24)	**0.49 (±0.18)**	0.61 (±0.25)	0.55 (±0.17)
	XGBoost	91.56 (±1.70)	84.47 (±6.08)	**0.08 (±0.02)**	**0.16 (±0.06)**	0.69 (±0.07)	0.48 (±0.17)	0.74 (±0.06)	0.57 (±0.16)
	CNN	79.81 (±2.19)	81.06 (±5.06)	0.81 (±0.09)	0.76 (±0.20)	0.00 (±0.00)	0.00 (±0.00)	0.89 (±0.01)	0.89 (±0.03)
	LSTM	80.38 (±3.06)	79.73 (±7.08)	0.78 (±0.12)	0.81 (±0.28)	0.00 (±0.00)	0.00 (±0.00)	0.89 (±0.02)	0.89 (±0.04)
Haberman	ULST2-FLS	74.92 (±1.93)	73.52 (±4.66)	0.71 (±0.04)	0.75 (±0.10)	0.17 (±0.09)	0.10 (±0.08)	0.85 (±0.01)	0.84 (±0.02)
	Adam ULST2-FLS	76.05 (±1.61)	74.44 (±3.45)	0.67 (±0.04)	0.76 (±0.10)	0.26 (±0.07)	0.17 (±0.07)	0.85 (±0.01)	0.83 (±0.03)
	SM Adam ULST2-FLS	76.28 (±1.73)	73.39 (±4.33)	0.67 (±0.04)	0.76 (±0.10)	0.26 (±0.07)	0.17 (±0.07)	0.85 (±0.01)	0.83 (±0.03)
	MSE Adam ULST2-FLS	75.55 (±1.63)	**75.49 (±3.53)**	0.69 (±0.03)	0.71 (±0.08)	0.24 (±0.06)	0.21 (±0.08)	0.85 (±0.01)	**0.85 (±0.02)**
	GWO ULST2-FLS	75.79 (±2.43)	74.37 (±3.88)	0.91 (±0.08)	0.97 (±0.02)	0.14 (±0.14)	0.14 (±0.12)	0.23 (±0.20)	0.23 (±0.17)
	XGBoost	78.42 (±1.66)	73.06 (±4.23)	**0.22 (±0.02)**	**0.27 (±0.04)**	**0.37 (±0.09)**	**0.23 (±0.11)**	0.50 (±0.11)	0.39 (±0.12)
	CNN	73.29 (±1.72)	73.62 (±4.57)	1.07 (±0.07)	1.06 (±0.18)	0.01 (±0.02)	0.01 (±0.03)	0.85 (±0.01)	0.85 (±0.03)
	LSTM	**78.70 (±1.65)**	73.22 (±3.40)	0.85 (±0.07)	1.07 (±0.14)	0.33 (±0.06)	0.20 (±0.01)	**0.87 (±0.01)**	0.83 (±0.02)
Sonar	ULST2-FLS	86.16 (±14.88)	70.19 (±9.45)	0.36 (±0.30)	0.86 (±0.15)	0.72 (±0.29)	0.40 (±0.17)	0.87 (±0.14)	0.74 (±0.11)
	Adam ULST2-FLS	98.30 (±3.44)	73.51 (±6.48)	0.05 (±0.10)	0.93 (±0.24)	0.97 (±0.07)	0.46 (±0.14)	0.98 (±0.04)	0.78 (±0.05)
	SM Adam ULST2-FLS	97.26 (±8.40)	74.54 (±10.54)	0.08 (±0.19)	0.86 (±0.25)	0.94 (±0.18)	0.49 (±0.17)	**0.99 (±0.03)**	**0.79 (±0.07)**
	MSE Adam ULST2-FLS	**98.90 (±1.04)**	73.87 (±4.48)	**0.04 (±0.04)**	0.97 (±0.16)	**0.98 (±0.02)**	0.47 (±0.09)	**0.99 (±0.01)**	0.76 (±0.05)
	GWO ULST2-FLS	98.46 (±1.09)	**76.67 (±3.38)**	0.95 (±0.12)	0.89 (±0.04)	0.42 (±0.30)	**0.52 (±0.07)**	0.60 (±0.32)	0.72 (±0.06)
	XGBoost	88.65 (±2.09)	73.35 (±6.65)	0.11 (±0.02)	**0.27 (±0.07)**	0.77 (±0.04)	0.46 (±0.13)	0.88 (±0.02)	0.71 (±0.06)
	CNN	78.52 (±3.63)	74.29 (±4.05)	0.86 (±0.15)	1.03 (±0.16)	0.56 (±0.08)	0.48 (±0.09)	0.81 (±0.03)	0.77 (±0.04)
	LSTM	80.84 (±2.61)	73.22 (±5.74)	0.77 (±0.10)	1.07 (±0.23)	0.61 (±0.05)	0.46 (±0.11)	0.83 (±0.02)	0.76 (±0.05)

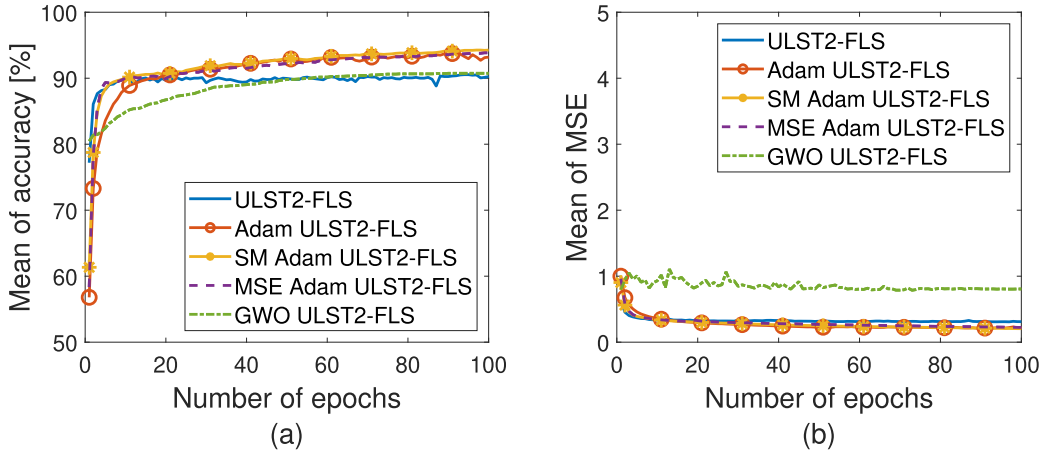

Figure 9 – Convergence speed for Appendicitis data set. (a) Accuracy. (b) MSE.

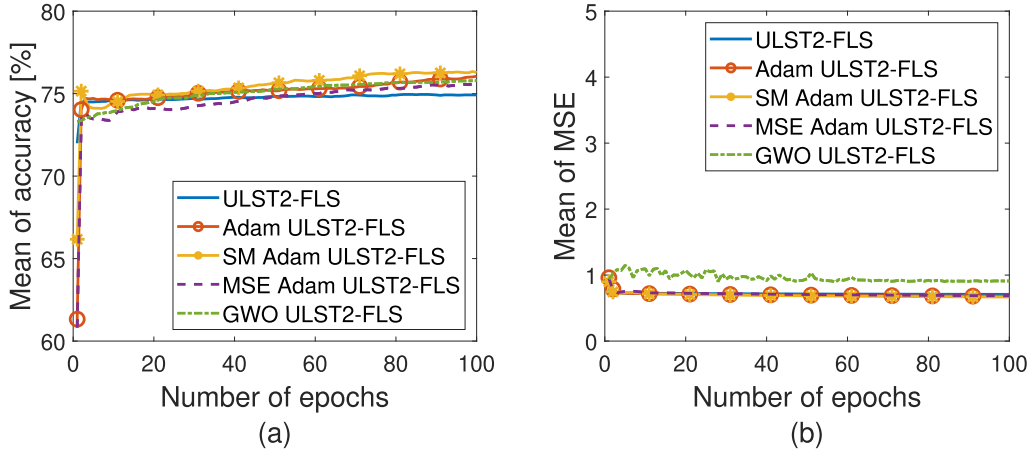

Figure 10 – Convergence speed for Haberman data set. (a) Accuracy. (b) MSE.

evaluate the results in Table 3. Considering two sets of samples $\mathcal{G}_1$ and $\mathcal{G}_2$, the two-sample t-test allows us to infer assumptions from two independent data samples and to verify their statistical validity. This statistical test is expressed as:

$$\tau = \frac{\overline{\mathcal{G}_1} - \overline{\mathcal{G}_2}}{\sqrt{\frac{\sigma^2_{\mathcal{G}_1}}{L_{\mathcal{G}_1}} + \frac{\sigma^2_{\mathcal{G}_2}}{L_{\mathcal{G}_2}}}} \tag{4.5}$$

where $\overline{\mathcal{G}_1}$, $\overline{\mathcal{G}_2}$, $\sigma^2_{\mathcal{G}_1}$ and $\sigma^2_{\mathcal{G}_2}$ are the means and standard deviations values of the samples belonging to $\mathcal{G}_1$ and $\mathcal{G}_2$, respectively. Also, $L_{\mathcal{G}_1} = \#\{\mathcal{G}_1\}$ and $L_{\mathcal{G}_2} = \#\{\mathcal{G}_2\}$, in which $\#$ denotes the Cardinality operator. The degree of freedom is defined as $L_{\mathcal{G}_1} + L_{\mathcal{G}_1} - 2$. In addition to the determination of τ, it becomes important to infer the hypothesis, which are given by:

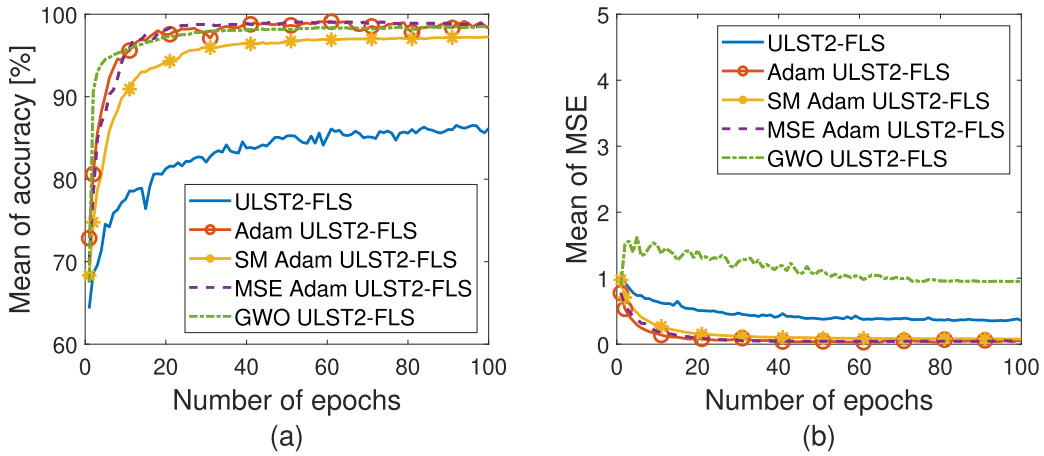

Figure 11 – Convergence speed for Sonar data set. (a) Accuracy. (b) MSE.

$$\begin{cases} \mathcal{H}_0 : \overline{\mathcal{G}_1} = \overline{\mathcal{G}_2} \\ \mathcal{H}_1 : \overline{\mathcal{G}_1} \neq \overline{\mathcal{G}_2} \end{cases} \tag{4.6}$$

Given a significance level α, usually around 0.05, the p-value is calculated from τ and represents the lowest value of α to reject the null hypothesis ($\mathcal{H}_0$). Thus, values of the p-value below α means that the null hypothesis is not true (MOORE; KIRKLAND, 2007).

The sets of samples $\mathcal{G}_1$ and $\mathcal{G}_2$ of Eq. (4.5) are the test accuracy metric obtained for the 33 execution times, where $\mathcal{G}_1$ refers to GWO ULST2-FLS and $\mathcal{G}_2$ can denote ULST2-FLS, Adam ULST2-FLS, SM Adam ULST2-FLS, MSE Adam ULST2-FLS, and non-fuzzy classifiers XGBoost, CNN and LSTM, as listed in Table 4. The degree of freedom presented in these statistical tests is 64, which is relatively high, so there is no need to verify the normality of error distributions (MOORE; KIRKLAND, 2007).

Table 4 presents the results of the two-sample t-test performed for the test accuracy metric in Table 3. The rejection of the null hypothesis are indicated by the letters 'W' and 'L' representing respectively the wins and the losses of the method tested. Meanwhile the acceptance of the null hypothesis is described by 'E' which means equality of the tested methods.

The results show that GWO ULST2-FLS certainly presented greater results than the other methods in most data sets. Specifically, the proposed method demonstrates statistically superior performance when compared to ULST2-FLS and Adam ULST2-FLS on the Appendicitis and Sonar datasets. In Haberman data set, the performances of all classifiers were statistically similar. Analyzing the comparison to the non-fuzzy methods, it can be noted that the proposal offered greater test accuracy results in most cases.

Table 4 – Statistical analyses performed by two-sample t-test for the test accuracy metric.

Data set	$\mathcal{G}_1$	$\mathcal{G}_2$	p-value	Lower boundary	Upper boundary	$\mathcal{H}_0$
Appendicitis	GWO ULST2-FLS	ULST2-FLS	0.01	0.01	0.06	W
	GWO ULST2-FLS	Adam ULST2-FLS	0.01	0.01	0.05	W
	GWO ULST2-FLS	SM Adam ULST2-FLS	0.22	-0.01	0.04	E
	GWO ULST2-FLS	MSE Adam ULST2-FLS	0.43	-0.01	0.03	E
	GWO ULST2-FLS	XGBoost	0.18	-0.04	0.01	E
	GWO ULST2-FLS	CNN	0.00	0.03	0.08	W
	GWO ULST2-FLS	LSTM	0.00	0.04	0.09	W
Haberman	GWO ULST2-FLS	ULST2-FLS	0.42	-0.01	0.03	E
	GWO ULST2-FLS	Adam ULST2-FLS	0.94	-0.02	0.02	E
	GWO ULST2-FLS	SM Adam ULST2-FLS	0.33	-0.01	0.03	E
	GWO ULST2-FLS	MSE Adam ULST2-FLS	0.22	-0.03	0.01	E
	GWO ULST2-FLS	XGBoost	0.20	-0.01	0.03	E
	GWO ULST2-FLS	CNN	0.47	-0.01	0.03	E
	GWO ULST2-FLS	LSTM	0.20	-0.01	0.03	E
Sonar	GWO ULST2-FLS	ULST2-FLS	0.00	0.03	0.10	W
	GWO ULST2-FLS	Adam ULST2-FLS	0.02	0.01	0.06	W
	GWO ULST2-FLS	SM Adam ULST2-FLS	0.27	-0.02	0.06	E
	GWO ULST2-FLS	MSE Adam ULST2-FLS	0.11	-0.01	0.06	E
	GWO ULST2-FLS	XGBoost	0.01	0.01	0.06	W
	GWO ULST2-FLS	CNN	0.01	0.01	0.04	W
	GWO ULST2-FLS	LSTM	0.00	0.01	0.06	W

4.2 Hot box and hot wheels

In this subsection, the hot box and hot wheels problem will be addressed using the classifiers to differentiate the proper warnings from the improper ones. Proper warnings are the ones that really indicate hot box or hot wheels problems on trains. Whereas improper warnings can be given when an external factor, such as sensor defect, solar incidence or longitudinal misalignment, affects the temperature measurement and do not indicate real hot box and hot wheels problems.

The data set used in this classification was provided by MRS Logística S.A. It is diverse enough to represent the complexity of the problem, which means that the continuous update of the proposed models is not required.

This data set consists of a list of temperature measurements on the right and left

wheels and bearings of the train when a warning have been triggered. They were collected by various hot box and hot wheels detectors controlled by MRS Logística S.A. and are labeled as proper or improper warnings by a specialist from the company. There are data for 209 different triggered warning, which leads to a diversity of cases.

These data were treated by the extraction of some characteristics used for the classification, which were the mean, the median, the standard deviation, the minimum value and the maximum value of these measured temperatures. As a result, the final data set is composed of 209 number of samples, each one with 20 input features.

Likewise the classifiers are composed of four rules, two rules for the each class. The first rule for each class were heuristically created in the same way described in Eq. 4.3 and the second rule for each class was generated by a modification in the first one as it is expressed by Eq. (4.4). For this problem, the same parameter were adopted.

In the same way, the data set have been presented for each classifier 33 times, each time, seventy percent of the data set was randomly divided for the training phase and thirty percent for the test phase. As previously adopted, the maximum number of iterations set for the training phase was 100, as the improvement in accuracy beyond this point was minimal and did not justify the increased computational load required by additional epochs. Therefore, it was unnecessary to adopt a stopping criterion based on error.

The performance metrics used for the numerical results analyses were accuracy, mean squared error (MSE), Cohen's kappa coefficient and F-score as well.

The dataset were also applied in binary classifiers based on different machine learning techniques, such as a gradient boosting: the Scalable Tree Boosting System (XGBoost) (CHEN; GUESTRIN, 2016); and two deep learning methods: Convolutional Neural Network (CNN) (KIM, 2017) and Long Short-term Memory (LSTM) Network (PALUSZEK; THOMAS, 2020).

4.2.1 Classification rate analysis

Table 5 presents the results of the classification rate obtained by the proposed classifier GWO ULST2-FLS and by the original classifiers ULST2-FLS, Adam ULST2-FLS, SM Adam ULST2-FLS, MSE Adam ULST2-FLS and the deep learning classifiers CNN and LSTM. Observing this table, it is possible to see that the proposal achieved greater values of accuracy during the test phases compared to the other fuzzy and non-fuzzy methods. Additionally, the proposed method provided better results for Kappa coefficient during test phase and F-score during training and test phases.

Table 5 – Performance comparison in terms of the mean and standard deviation for HB and HW problem.

Method	Training accuracy %	Test accuracy %	Training MSE	Test MSE	Training Kappa	Test Kappa	Training Fscore	Test Fscore
ULST2-FLS	86.63 (±1.95)	86.49 (±3.26)	0.35 (±0.04)	0.38 (±0.06)	0.10 (±0.16)	0.08 (±0.14)	0.34 (±0.13)	0.33 (±0.10)
Adam ULST2-FLS	92.14 (±3.41)	88.30 (±4.58)	0.22 (±0.06)	0.34 (±0.10)	0.60 (±0.24)	0.39 (±0.27)	0.69 (±0.14)	0.50 (±0.18)
SM Adam ULST2-FLS	91.41 (±3.96)	88.30 (±4.58)	0.23 (±0.06)	0.36 (±0.10)	0.51 (±0.30)	0.37 (±0.26)	0.66 (±0.19)	0.52 (±0.19)
MSE Adam ULST2-FLS	91.25 (±3.37)	89.02 (±3.41)	0.23 (±0.06)	0.34 (±0.09)	0.53 (±0.26)	0.38 (±0.24)	0.60 (±0.22)	0.49 (±0.18)
GWO ULST2-FLS	91.45 (±1.77)	**92.06 (±3.48)**	0.85 (±0.16)	0.81 (±0.18)	0.28 (±0.23)	**0.54 (±0.18)**	0.72 (±0.39)	**0.96 (±0.02)**
XGBoost	**96.28 (±1.23)**	90.67 (±3.62)	**0.04 (±0.01)**	**0.09 (±0.04)**	**0.84 (±0.05)**	**0.54 (±0.17)**	**0.98 (±0.01)**	0.95 (±0.02)
CNN	87.13 (±2.49)	86.45 (±3.51)	0.51 (±0.10)	0.54 (±0.14)	0.24 (±0.14)	0.17 (±0.16)	0.31 (±0.14)	0.29 (±0.14)
LSTM	89.54 (±2.77)	88.01 (±3.93)	0.42 (±0.11)	0.48 (±0.16)	0.39 (±0.23)	0.27 (±0.22)	0.48 (±0.19)	0.36 (±0.19)

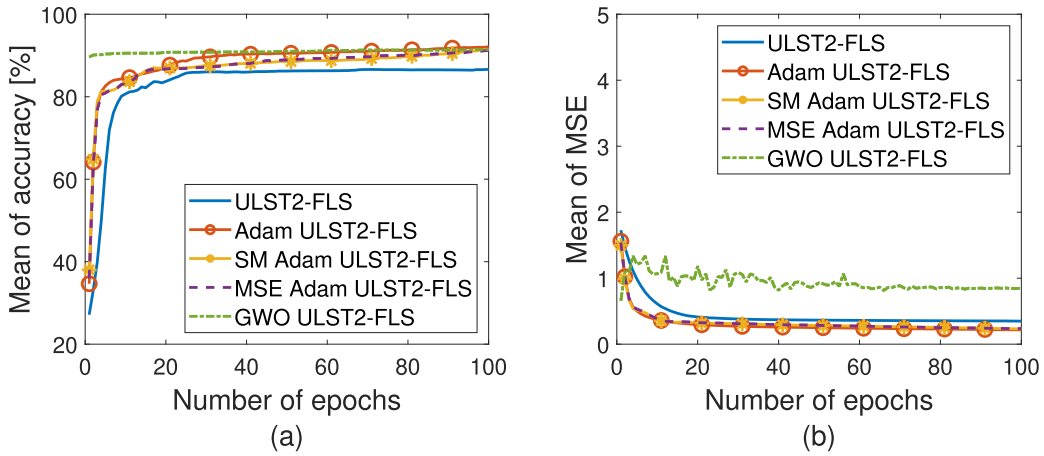

Figure 12 – Convergence speed for HB and HW data set. (a) Accuracy. (b) MSE.

4.2.2 Convergence speed analysis

Fig. 12 shows the convergence speed of the classifiers. Analysing the mean of accuracy for each epoch, it can be noted that the proposed classifier presented great convergence speed when compared to the other classifiers. This outcome is very important for industrial application, which requires real-time application and a low cost method. Therefore, the proposals, which offered a classification with higher accuracy and convergence speed, are better for the examined application compared to the original classifier. However, it presents the same limitation in terms of mean square error (MSE) during the training phase.

4.2.3 Statistical analysis

The statistical test were performed in the same way in described in 4.1.4. The sets of samples $\mathcal{G}_1$ and $\mathcal{G}_2$ of Eq. (4.5) are the test accuracy metric obtained for the 33 execution times, where $\mathcal{G}_1$ refers to GWO ULST2-FLS and $\mathcal{G}_2$ can denote ULST2-FLS, Adam ULST2-FLS, SM Adam ULST2-FLS, MSE Adam ULST2-FLS, XGBoost, CNN and LSTM as listed in Table 6. Also, the degree of freedom presented in these statistical tests is 64, which is relatively high, so there is no need to verify the normality of error distributions (MOORE; KIRKLAND, 2007).

Table 6 presents the results of the two-sample t-test performed for the test accuracy metric in Table 5. Likewise, the rejection of the null hypothesis are indicated by the letters 'W' and 'L' representing respectively the wins and the losses of the method tested. Meanwhile the acceptance of the null hypothesis is described by 'E' which means equality of the tested methods.

The results show that GWO ULST2-FLS presented superior performance than all

Table 6 – Statistical analyses performed by two-sample t-test for the test accuracy metric for HB and HW problem.

$\mathcal{G}_1$	$\mathcal{G}_2$	p-value	Lower boundary	Upper boundary	$\mathcal{H}_0$
GWO ULST2-FLS	ULST2-FLS	0.00	0.04	0.07	W
GWO ULST2-FLS	Adam ULST2-FLS	0.00	0.02	0.06	W
GWO ULST2-FLS	SM Adam ULST2-FLS	0.00	0.02	0.06	W
GWO ULST2-FLS	MSE Adam ULST2-FLS	0.00	0.01	0.05	W
GWO ULST2-FLS	XGBoost	0.12	0.00	0.03	E
GWO ULST2-FLS	CNN	0.00	0.04	0.07	W
GWO ULST2-FLS	LSTM	0.00	0.02	0.06	W

the other fuzzy classifier and superior than most non-fuzzy classifiers. This conclusion proves that the proposal can lead to better results than most other classifiers that are based on different research approaches, demonstrating compatibility for handling the hot box and hot wheels problem.

5 CONCLUSIONS

This work has introduced a Type-2 Fuzzy Logic System trained by the Swarm Intelligence algorithm Grey Wolf Optimizer to enhace the method and obtain greater accuracy in classification problems and convergence speed during the training phase.

As regards the comparison of the fuzzy classifiers, the GWO ULST2-FLS offered competitive training results for all data sets presented in the work. Moreover, the proposal have achieved greater classification rate during testing phase when compared to the original classifier. In addition, the proposed method offered competitive values of Cohen's Kappa and F-Score metrics. Combining all of these factors, it is possible to notice that GWO ULST2-FLS is a more attractive method for training ULST2-FLS and is more likely to obtain higher accuracy for most binary classification problems. In relation to the comparison between the proposal and the deep learning methods, the proposed method presented better values of all the studied metrics in most applications, indicating that the proposal is also a great alternative even out of the fuzzy area.

For these reasons, the presented model proved to be an appropriate option in all analyzed situations and feasible for industrial application, especially in the question of warning classification in hot box and hot wheel problems, which requires high accuracy for almost real-time application, as it is demanded for most industrial application.

Future works will be focused on extending the concepts of GWO ULST2-FLS to upper and lower non-singleton type-2 FLS for handling the presence of uncertainty in the measured data set.

REFERENCES

AAR. **Nationwide Wayside Detector System**. 2015. <https://docplayer.net/17829351-Nationwide-wayside-detector-system.html>.

ABADI, M. S. E.; ESKANDARI, H. Performance study of set-membership nlms algorithm over an adaptive incremental network. In: **3th International eConference on Computer and Knowledge Engineering**. [S.l.: s.n.], 2013. p. 257–261.

ABIYEV, R. Credit rating using type-2 fuzzy neural networks. **Mathematical Problems in Engineering**, v. 2014, p. 1–8, 03 2014.

AGUIAR, E. P. d.; NOGUEIRA, F. M. de A.; VELLASCO, M. M. B. R.; RIBEIRO, M. V. Set-membership type-1 fuzzy logic system applied to fault classification in a switch machine. **IEEE Transactions on Intelligent Transportation Systems**, v. 18, n. 10, p. 2703–2712, 2017.

AGUIAR, E. P. de; AMARAL, R. P. F.; VELLASCO, M. M. B. R.; RIBEIRO, M. V. An enhanced singleton type-2 fuzzy logic system for fault classification in a railroad switch machine. **Electric Power Systems Research**, Elsevier, v. 158, p. 195–206, 2018.

AGUIAR, E. P. de; FERNANDES, T. E.; FERNANDO, M. d. A.; SILVEIRA, D. D.; VELLASCO, M. M.; RIBEIRO, M. V. A new model to distinguish railhead defects based on set-membership type-2 fuzzy logic system. **International Journal of Fuzzy Systems**, Springer, p. 1–13, 2020.

AGUIAR, E. P. de; NOGUEIRA, F.; AMARAL, R.; FABRI, D.; ROSSIGNOLI, S.; FERREIRA, J.; VELLASCO, M.; TANSCHEIT, R.; VELLASCO, P.; RIBEIRO, M. Classification of events in switch machines using bayes, fuzzy logic system and neural network. In: . [S.l.: s.n.], 2014. v. 459.

AGUIAR, E. P. de; NOGUEIRA, F. M.; AMARAL, R.; FABRI, D. F.; ROSSIGNOLI, S. C. d.; FERREIRA, J. G.; VELLASCO, M.; TANSCHEIT, R.; VELLASCO, P.; RIBEIRO, M. Eann 2014: a fuzzy logic system trained by conjugate gradient methods for fault classification in a switch machine. **Neural Computing and Applications**, p. 1–15, 04 2015.

AGUIAR, E. P. de; NOGUEIRA, F. M. de A.; VELLASCO, M. M. B. R.; RIBEIRO, M. V. Set-membership type-1 fuzzy logic system applied to fault classification in a switch machine. **IEEE Transactions on Intelligent Transportation Systems**, v. 18, n. 10, p. 2703–2712, 2017.

AGUIAR, E. Pestana de; SILVA, L.; MOREIRA, A.; GOLIATT, L.; NOGUEIRA, F.; VELLASCO, M.; RIBEIRO, M. Type-1 fuzzy logic system applied to classification of rail head defects. In: . [S.l.: s.n.], 2016.

ALCALá-FDEZ, J.; FERNáNDEZ, A.; LUENGO, J.; DERRAC, J.; GARCíA, S.; SANCHEZ, L.; HERRERA, F. Keel data-mining software tool: Data set repository, integration of algorithms and experimental analysis framework. **Journal of Multiple-Valued Logic and Soft Computing**, v. 17, p. 255–287, 01 2010.

ALMARAASHI, M.; JOHN, R.; HOPGOOD, A.; AHMADI, S. Learning of interval and general type-2 fuzzy logic systems using simulated annealing: Theory and practice. **Information Sciences**, Elsevier, v. 360, p. 21–42, 2016.

AMARAL, R. P. F.; RIBEIRO, M. V.; AGUIAR, E. P. de. Type-1 and singleton fuzzy logic system trained by a fast scaled conjugate gradient methods for dealing with binary classification problems. **Neurocomputing**, Elsevier, v. 355, p. 57–70, 2019.

ANTF. **General Information**. <https://www.antf.org.br/informacoes-gerais/>.

BOJAN-DRAGOS, C.-A.; PRECUP, R.-E.; PREITL, S.; ROMAN, R.-C.; HEDREA, E.-L.; SZEDLAK-STINEAN, A.-I. Gwo-based optimal tuning of type-1 and type-2 fuzzy controllers for electromagnetic actuated clutch systems. **IFAC-PapersOnLine**, Elsevier, v. 54, p. 189–194, 01 2021.

BRAREN, H.; KENNELLY, M.; EIDE, E. Wayside Detection: Component Interactions and Composite Rules. In: . [S.l.: s.n.], 2009. (Rail Transportation Division Conference, ASME 2009 Rail Transportation Division Fall Technical Conference), p. 111–117.

CHEN, T.; GUESTRIN, C. Xgboost: A scalable tree boosting system. In: . [S.l.: s.n.], 2016. p. 785–794.

CHONG, S. Y.; SHIN, H. A review of health and operation monitoring technologies for trains. **Smart Structures and Systems**, v. 6, p. 1079–1105, 12 2010.

CLARKE, P.; LAMARE, R. C. de. Low-complexity reduced-rank linear interference suppression based on set-membership joint iterative optimization for ds-cdma systems. **IEEE Transactions on Vehicular Technology**, v. 60, n. 9, p. 4324–4337, 2011.

DAVIRAN, M.; GHEZELBASH, R.; MAGHSOUDI, A. Gwokm: A novel hybrid optimization algorithm for geochemical anomaly detection using grey wolf optimizer and k-means clustering model. **Geochemistry**, Elsevier, p. 126036, 10 2023.

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

FONSECA, L.; AGUIAR, E. Pestana de. Stochastic optimization combined with type-2 fuzzy logic system for the classification of trends in hot boxes and hot wheels. **International Journal of Fuzzy Systems**, v. 24, 08 2022.

FRA. **Effects of Temperature on Wheel Shelling**. <https://railroads.dot.gov/rolling-stock/current-projects/effects-temperature-wheel-shelling>.

_____. **An Implementation Guide for Wayside Detector Systems**. <https://railroads.dot.gov/sites/fra.dot.gov/files/fra_net/18667/Wayside%%20Implementation%20Guide.pdf>.

_____. **Train Accidents and Rates, Office of Safiety Analysis**. <https://railroads.dot.gov/rolling-stock/current-projects/effects-temperature-wheel-shelling>.

HAGRAS, H. A hierarchical type-2 fuzzy logic control architecture for autonomous mobile robots. **IEEE Transactions on Fuzzy Systems**, v. 12, n. 4, p. 524–539, 2004.

HAIDARI, A.; TEHRANI, P. H. Thermal load effects on fatigue life of a cracked railway wheel. **Latin American Journal of Solids and Structures**, v. 12, p. 1144–1157, 06 2015.

HAYKIN, S. **Adaptive Filter Theory**. [S.l.]: Prentice-Hall Inc, Upper Saddle River, NJ, USA, 1996.

JOHN, R.; COUPLAND, S. Extensions to type-1 fuzzy: Type-2 fuzzy logic and uncertainty. In: _____. [S.l.: s.n.], 2006. p. 89–102.

KIM, P. Convolutional neural network. In: **MATLAB Deep Learning: With Machine Learning, Neural Networks and Artificial Intelligence**. Berkeley, CA: Apress, 2017. cap. 6, p. 121–147.

KINGMA, D. P.; BA, J. Adam: A method for stochastic optimization. **arXiv e-prints**, dec 2014. Disponível em: <https://ui.adsabs.harvard.edu/abs/2014arXiv1412.6980K>.

LIANG, Q.; KARNIK, N.; MENDEL, J. Connection admission control in atm networks using survey-based type-2 fuzzy logic systems. **IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)**, v. 30, n. 3, p. 329–339, 2000.

LIANG, Q.; MENDEL, J. M. Equalization of nonlinear time-varying channels using type-2 fuzzy adaptive filters. **IEEE Transactions on Fuzzy Systems**, v. 8, n. 5, p. 551–563, 2000.

_____. Interval type-2 fuzzy logic systems: theory and design. **IEEE Transactions on Fuzzy Systems**, v. 8, n. 5, p. 535–550, 2000.

_____. Overcoming time-varying co-channel interference using type-2 fuzzy adaptive filter. **IEEE Trans. Circuits Syst.**, v. 9, n. 6, p. 1419–1428, 2000.

_____. Mpeg vbr video traffic modeling and classification using fuzzy techniques. **IEEE Transactions on Fuzzy Systems**, v. 9, n. 1, p. 183–193, 2001.

MELIN, P.; CASTILLO, O. A new method for adaptive control of nonlinear plants using type-2 fuzzy logic and neural networks. **International Journal of General Systems**, v. 33, p. 289–304, 2004.

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

Mendel, J. M. Type-2 fuzzy sets and systems: an overview. **IEEE Computational Intelligence Magazine**, v. 2, n. 1, p. 20–29, 2007.

MENDEL, J. M.; JOHN, R. I.; LIU, F. Interval type-2 fuzzy logic systems made simple. **IEEE Transactions on Fuzzy Systems**, v. 14, n. 6, p. 808–821, 2006.

MIRJALILI, S.; MIRJALILI, S.; LEWIS, A. Grey wolf optimizer. **Advances in Engineering Software**, Elsevier, v. 69, p. 46–61, 03 2014.

MOORE, D.; KIRKLAND, S. **The Basic Practice of Statistics**. [S.l.]: WH Freeman, 2007.

OZEN, T.; GARIBALDI, J. Investigating adaptation in type-2 fuzzy logic systems applied to umbilical acid-base assessment. **European Symposium on Intelligent Technologies**, p. 289–294, 07 2003.

PALO, M.; GALAR, D.; NORDMARK, T.; ASPLUND, M.; LARSSON, D. Condition monitoring at the wheel/rail interface for decision-making support. **Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit**, v. 228, p. 705–715, 07 2014.

PALUSZEK, M.; THOMAS, S. **Practical MATLAB Deep Learning**. Berkeley, CA: Apress, 2020.

SALMENTO, M.; AGUIAR, E. d.; RIBEIRO, M. Reduced rank adaptive filter with variable step size for impulsive uwb plc systems. In: **XXXIII Brazilian Telecommunications Symposium**. [S.l.: s.n.], 2015. p. 1–5.

SALMENTO, M. L. G.; Pestana de Aguiar, E.; CAMPONOGARA Ândrei; RIBEIRO, M. V. An enhanced receiver for an impulsive uwb-based plc system for low-bit rate applications. **Digital Signal Processing**, v. 70, p. 145–154, 2017.

TARAWNEH, C.; ARANDA, J.; HERNANDEZ, V.; RAMIREZ, C. An analysis of the efficacy of wayside hot-box detector data. In: . [S.l.: s.n.], 2018. p. V001T02A012.

TARAWNEH, C.; ARANDA, J.; HERNANDEZ, V.; CROWN, S.; MONTALVO, J. An investigation into wayside hot-box detector efficacy and optimization. **International Journal of Rail Transportation**, v. 8, p. 1–21, 06 2019.

TARAWNEH, C.; FUENTES, A.; WILSON, B.; COLE, K.; NAVARRO, L. Thermal analysis of railroad bearings: Effect of wheel heating. In: . [S.l.: s.n.], 2009.

TSB. **Transportation Safety Board of Canada, Railway Investigation Report R13T0122**. 2013. <https://www.tsb.gc.ca/eng/rapports-reports/rail/2013/r13t0122/r13t0122.html>.

VOESTALPINE Railway Systems, PHOENIX MDS HBD/HWD Hot Box Detection Hot Wheel Detection. <https://www.voestalpine.com/railway-systems/en/products/phoenix-mds-hbd-hwd-hot-box-detection-and-hot-wheel-detection/>.

WAGNER, C.; HAGRAS, H. Uncertainty and type-2 fuzzy sets and systems. In: . [S.l.: s.n.], 2010. p. 1 – 5.

WANG, L.; DELAMARE, R. C. Low-complexity constrained adaptive reduced-rank beamforming algorithms. **IEEE Transactions on Aerospace and Electronic Systems**, v. 49, n. 4, p. 2114–2128, 2013.

WU, D.; TAN, W. A type-2 fuzzy logic controller for the liquid-level process. In: **2004 IEEE International Conference on Fuzzy Systems**. [S.l.: s.n.], 2004. v. 2, p. 953–958.

WU, H.; MENDEL, J. M. Classifier designs for binary classifications of ground vehicles. In: CARAPEZZA, E. M. (Ed.). **Unattended Ground Sensor Technologies and Applications V**. [S.l.]: SPIE, 2003. v. 5090, p. 122–133.